GAUSSIAN SEMI PARAMETRIC ESTIMATION OF FRACTIONAL COINTEGRATION

By Carlos Velasco

Universidad Carlos III de Madrid

First Version received March 2000

Abstract. We analyse consistent estimation of the memory parameters of a nonstationary fractionally cointegrated vector time series. Assuming that the cointegrating relationship has substantially less memory than the observed series, we show that a multivariate Gaussian semi-parametric estimate, based on initial consistent estimates and possibly tapered observations, is asymptotically normal. The estimates of the memory parameters can rely either on original (for stationary errors) or on differenced residuals (for nonstationary errors) assuming only a convergence rate for a preliminary slope estimate. If this rate is fast enough, semi-parametric memory estimates are not affected by the use of residuals and retain the same asymptotic distribution as if the true cointegrating relationship were known. Only local conditions on the spectral densities around zero frequency for linear processes are assumed. We concentrate on a bivariate system but discuss multi-variate generalizations and show the performance of the estimates with simulated and real data.

Keywords. Residual-based estimation; nonstationary time series; multiple time series; long memory; long range dependence.

1. INTRODUCTION

Since the introduction of the concept of cointegration by Granger (1981) and Engle and Granger (1987), a vast literature has developed for the analysis of dynamic relationships among nonstationary time series. For many years, the theoretical and empirical focus was on the analysis of (integer) integrated time series but, as proposed initially, the idea of cointegration fits naturally in the broader field of fractionally integrated processes.

At the same time that the analysis of stationary long range dependent and fractional processes has progressed in many directions, nonstationary extensions have arisen a growing interest. Following Hurvich and Ray (1995), we generalize the terminology of I(d) processes with memory parameter $d < \frac{3}{2}$ for time series z_t with covariance stationary increments Δz_t (with $\Delta = 1-L$, where L is the lag operator). Thus we say that z_t is a stationary I(d_z) process if it has spectral density $f_z(\lambda)$ satisfying, for $0 < G_z < \infty$

$$f_z(\lambda) \sim G_z \lambda^{-2d_z} \qquad \text{as } \lambda \to 0^+ \qquad \text{for } d_z < \frac{1}{2}$$

If z_t is nonstationary, but has zero mean stationary increments Δz_t which are $I(d_z - 1)$, $d_z \in [\frac{1}{2}, \frac{3}{2})$, then we also say that z_t is $I(d_z)$. We focus on the behaviour of the relevant spectral density at the origin, parameterized by the memory d_z, and avoid any restriction at other frequencies apart from integrability for covariance stationarity. Note that this definition covers the traditional I(1) and I(0) paradigm. However, this is not the only alternative for definition of nonstationary fractional processes, and, for example, Robinson and Marinucci (2001) and Tanaka (1999) use truncated fractional difference filters, which result in nonstationary processes for any value of d.

A time series vector is cointegrated if a non null linear combination of the components has lower memory than any of them and the order of cointegration, or reduction of the memory, indicates the strength of the long run relationship linking the components of the vector. It is often implicit the assumption that the original series are nonstationary while some linear combinations can be in a stationary equilibrium, or at least are closer to stationarity. The classical assumption is that the original series have a unit root, $d = 1$, and that a linear combination is weakly dependent with memory $\delta = 0$, typically denoted as CI(1, 0) cointegration. Many inference tools have been designed for the estimation of the cointegration relationship, once the I(1) hypothesis was not rejected by standard unit root tests. These include ordinary least squares (OLS), two step, fully modified, spectral and nonlinear regressions, and maximum likelihood and canonical correlation procedures. Dolado and Marmol (1998) try to extend some efficient methods to cover fractional environments.

When no assumption is made about the memory of the series, additional inference problems come to forth. In this spirit, Robinson and Marinucci (2001) design narrow band spectral estimates, based on the proposal of Robinson (1994a) for stationary long memory series. These estimates are consistent for a wide range of memory values, including stationary ones, and may achieve better convergence rates than simpler OLS alternatives. However, no definitive answer is available for inference in a general set up because the asymptotic distributions depend on unknown parameters.

The main extra difficulty is the determination of the cointegration order. This entails estimation of the memory d of the original series (and testing that all series have the same memory) and estimation of the memory δ of the cointegrating relationship. The first problem can be tackled with standard multi variate parametric and semi parametric methods, taking into account the zero frequency spectral singularity implied by cointegration. The same procedures can be implemented without modification using residuals of a fitted regression model as proposed by Robinson and Marinucci (2001) and Hassler *et al.* (2002) using Geweke and Porter Hudak's (1983) log periodogram regression or Robinson's (1995b) local Whittle estimate. Additionally, narrow band estimates based on nonparametric coherence estimation at low frequencies can be employed as in Velasco (2002) and coherence evidence on possibly cointegrating vectors can be obtained using Lobato and Velasco's (2000) methods.

In this paper, we explore univariate residual inference for the memory of the cointegrating relationship based on the Gaussian semi parametric estimate optimizing Künsch (1987) and Robinson's (1995b) local Whittle likelihood. An alternative approach is the joint estimation of the memory parameters of the vector series and cointegrating residuals, possibly together with the cointegrating relationship. This second alternative would fit in the framework of Smith and Chen's (1997) semi parametric regression model or in the bivariate ARFIMA cointegrated system of Dueker and Startz (1998). Then, we pursue this possibility using the semi parametric multi variate approach of Lobato (1999). The two step procedure of Lobato (1999) requires initial estimates of the parameter vector with the semi parametric convergence rate depending on the number of Fourier frequencies used. This can be achieved by log periodogram regressions for Gaussian series, as suggested in Hassler *et al.* (2002), or more efficiently by Robinson's (1995b) Gaussian semi parametric estimates for both a component of the original vector and the regression residuals as we propose here. These initial memory estimates and the two step procedure require initial slope estimates with convergence rates as established in Assumption 1.

The residual based estimates are studied under different possibilites, distinguishing consistency and asymptotic normality, with original and differenced residuals. To deal with nonstationary vector series, we always consider differenced regressors; however, we consider both differenced and original cointegrating residuals, depending on possible knowledge about their nonstationarity, although we show that original residuals provide consistent estimates for most cases of practical interest. We may also taper the data as in Velasco (1999a, b) to avoid unexpected problems due to nonstationary or noninvertible series. Then, the two step estimates are asymptotically normal distributed if the cointegration degree $d - \delta$ is larger than $\frac{1}{2}$, which implies superconsistency of slope estimates see Assumption 1 and this asymptotic distribution is the same as if the cointegrating errors were actually observed.

The rest of paper is organized as follows. Section 2 reviews the properties of the residual periodogram and sets the main assumptions for residual based inference. This is analysed in Section 3 for differenced residuals and, in Section 4, for original residuals. In Section 5, the joint estimation procedure is proposed. Then we report in Section 6 the results of a small simulation exercise and the analysis of a real data example in Section 7 and conclude. All proofs are postponed to Sections 9 and 10.

2. RESIDUAL PERIODOGRAM

Let the observable nonstationary series y_t be generated by the equation

$$y_t = \beta x_t + u_t \qquad t = 1, \ldots, T \tag{1}$$

where $x_t \sim \mathrm{I}(d), \frac{1}{2} < d < \frac{3}{2}, u_t \sim \mathrm{I}(\delta), 0 \leqslant \delta < d$, and let $\tilde{\beta}$ be a consistent estimate of β based on T observations of (y_t, x_t), so the observed residuals are $\tilde{u}_t = y_t \quad \tilde{\beta} x_t$.

We base our frequency domain inference for the memory of u_t on these residuals or their differences, obtained through estimates $\tilde{\beta}$ whose rate of convergence satisfy Assumption 1:

Assumption 1. *Let* $x_t \sim \mathrm{I}(d),\ d \in (\frac{1}{2}, \frac{3}{2})$*;* $u_t \sim \mathbb{I}(\delta),\ \delta \in [0, d)$, *as* $T \to \infty$,

Case I: $\tilde{\beta} \quad \beta = O_p(T^{\delta - d}) \qquad \text{if } \delta + d > 1 \text{ or } \delta = 0,\ d = 1.$
Case II: $\tilde{\beta} \quad \beta = O_p(T^{\delta - d} \log T) \qquad \text{if } \delta + d = 1,\ \delta > 0.$
Case III: $\tilde{\beta} \quad \beta = O_p(T^{1-2d}) \qquad \text{if } \delta + d < 1.$

This assumption holds under regularity conditions when $\tilde{\beta}$ are the OLS estimates; see, for example, Davidson and de Jong (2000) or Robinson and Marinucci (2001) with a different definition of nonstationary processes. The narrow band frequency domain LS estimates of Robinson and Marinucci (2001) also satisfy this assumption when the bandwidth is chosen appropriately; though, in this case, we consider a further case in Assumption 1:

Case IV: $\tilde{\beta} \quad \beta = O_p(n^{1-d-\delta} T^{\delta - d}) \qquad \text{if } \delta + d < 1$
and n satisfies $n^{-1} + nT^{-1} \to 0.$

The bandwidth number, n, defines the band of frequencies where the local regression is performed, and, since it is allowed to increase arbitrarily slowly with T, the assumed rate for $\tilde{\beta}$ is faster than that of Case III and of similar form to Cases I and II, apart from slow increasing factors. Note that the slower the increase n, the better.

No substantial difference arises in the following analysis if more than one regressor is present, all with the same memory d and with regression coefficient estimates of the same convergence rate.

We set now the regularity conditions for the analysis of the asymptotic properties of the residual periodogram and semi parametric inference on the memory parameter of the cointegrating errors. We assume that they hold for the regressor component x_t of the observable cointegrated vector (y_t, x_t) or for the unobservable error series u_t, so $z \in \{x, u\}$, with memory parameter $d_z \in \{d, \delta\}$, and with the obvious implications on y_t. We follow mainly Hassler *et al.* (2002). but we do not assume Gaussianity anywhere.

Assumption 2. *When* $d_z \in (\ \frac{1}{2}, \frac{1}{2})$, *the spectral density* $f_z(\lambda)$ *satisfies for* $0 < \gamma \leq 2, 0 < G_z < \infty$,

$$f_z(\lambda) = G_z \lambda^{-2d_z}(1 + O(\lambda^{\gamma})) \qquad \text{as } \lambda \to 0^+$$

and when $d_z \in [\frac{1}{2}, \frac{3}{2})$, Δz_t *is zero mean and its spectral density* $f_{\Delta z}(\lambda)$ *satisfies*

$$f_{\Delta z}(\lambda) = G_z \lambda^{-2(d_z - 1)}(1 + O(\lambda^{\gamma})) \qquad \text{as } \lambda \to 0^+.$$

Assumption 2 strengthens the definition of I(d) process imposing a rate on the approximation of $f_z(\lambda)$ by $G_z\lambda^{-2d_z}$, while the next assumption imposes the differentiability of the spectral density, both conditions focusing only on low frequencies. We exclude nonstationary processes with drift, though certain tapering schemes allow the consistent estimation of memory parameters independently of this (Velasco, 1999a, b).

Assumption 3. *For* $v_t = z_t$, $0 \leqslant d_z < \frac{1}{2}$, *or for* $v_t = \Delta z_t$, $\frac{1}{2} \leqslant d_z < \frac{3}{2}$, $f_v(\lambda)$ *is differentiable in a neighbourhood* $(0, \varepsilon)$ *of the origin, and*

$$\left|\frac{\mathrm{d}}{\mathrm{d}\lambda} f_v(\lambda)\right| = O(f_v(\lambda)\lambda^{-1}) \qquad \text{as } \lambda \to 0^+.$$

This assumption holds for standard parametric models, including possibly nonstationary ($d_z \geqslant 0.5$) fractional ARIMAs. For simplicity, we assume that the parameter γ in Assumption 2 is common to x_t and u_t, while for y_t this parameter is then $\min\{\gamma, d - \delta\}$, where $2(d - \delta)$ will be implicitly assumed to be larger than 1 for the asymptotic normality of our semi parametric estimates; see Theorems 2 and 4 6.

2.1. *Periodogram of differenced residuals*

Denote the possibly tapered cross periodogram of two sequences a_t, b_t, $t = 1,\ldots,T$, as

$$I_{ab}(\lambda) = w_a(\lambda)\overline{w_b}(\lambda)$$

$$w_a(\lambda) = \left(2\pi\sum_{t=1}^{T} h_t^2\right)^{-1/2} \sum_{t=1}^{T} h_t a_t \exp(\mathrm{i}t\lambda)$$

where the overline indicates complex conjugation and h_t is a taper sequence, $0 \leqslant h_t \leqslant 1, t = 1,\ldots,T$. The periodogram is the basic static for our memory estimates. In this sub section, we only consider the non tapered case, $h_t \equiv 1$, but discuss related results in next sections when appropriate tapering is applied.

We first explore basic quantities for the analysis of estimates based on $\Delta\tilde{u}_t$, appropriate when u_t is nonstationary, $\delta > 0.5$. Set the Fourier frequencies $\lambda_j = 2\pi j/T$. Frequency domain inference on δ based on increments of residuals depends on the differenced residual periodogram,

$$I_{\Delta\tilde{u}\Delta\tilde{u}}(\lambda_j) = I_{\Delta u\Delta u}(\lambda_j) - (\tilde{\beta} - \beta)\{I_{\Delta u\Delta x}(\lambda_j) + I_{\Delta x\Delta u}(\lambda_j)\} + (\tilde{\beta} - \beta)^2 I_{\Delta x\Delta x}(\lambda_j)$$

and on the normalized difference with respect to the true errors periodogram, defined as

$$r_j^* := \frac{I_{\Delta\tilde{u}\Delta\tilde{u}}(\lambda_j) - I_{\Delta u \Delta u}(\lambda_j)}{g_{\Delta u}(\lambda_j)} = -(\tilde{\beta} - \beta)\frac{I_{\Delta u \Delta x}(\lambda_j) + I_{\Delta x \Delta u}(\lambda_j)}{g_{\Delta u}(\lambda_j)} + (\tilde{\beta} - \beta)^2 \frac{I_{\Delta x \Delta x}(\lambda_j)}{g_{\Delta u}(\lambda_j)}$$

with

$$g_z(\lambda) := G_z \lambda^{-2d_z} \sim f_z(\lambda) \qquad \text{as } \lambda \to 0^+.$$

We can bound this error using

$$|r_j^*| \leqslant 2|\tilde{\beta} - \beta| \frac{|I_{\Delta u \Delta x}(\lambda_j)|}{g_{\Delta u}(\lambda_j)} + (\tilde{\beta} - \beta)^2 \frac{I_{\Delta x \Delta x}(\lambda_j)}{g_{\Delta u}(\lambda_j)} \tag{2}$$

together with Assumption 1 about $\tilde{\beta} - \beta$ and the next lemma on the properties of the periodogram, taken from Robinson (1995a).

Lemma 1. *If Assumptions* 2 *and* 3 *hold and* $d_z \in (-\frac{1}{2}, \frac{1}{2})$, *for* $j = 1, \ldots, m$, $m/T \to 0$ *as* $T \to \infty$,

$$\mathrm{E}[I_{zz}(\lambda_j)] = f_z(\lambda_j)\{1 + O[j^{-1}\log(j+1)]\} = O(\lambda_j^{-2d_z}).$$

Note that we can apply this lemma to the increments of nonstationary series with $d_z \in (0.5, 1.5)$, which are the series typically involved when we use $\Delta\tilde{u}_t$. Also, we can use Cauchy inequality to bound the contribution of cross periodograms obtaining, for example, that

$$\mathrm{E}|I_{\Delta u \Delta x}(\lambda_j)| \leqslant \mathrm{E}[I_{\Delta u \Delta u}(\lambda_j)]^{1/2} \mathrm{E}[I_{\Delta x \Delta x}(\lambda_j)]^{1/2} = O(\lambda_j^{2-d-\delta})$$

when $\delta \in (0.5, d)$.

2.2. *Periodogram of original residuals*

When using original residuals $\tilde{u}_t$, we are led to the analysis of

$$I_{\tilde{u}\tilde{u}}(\lambda_j) = I_{uu}(\lambda_j) - (\tilde{\beta} - \beta)[I_{ux}(\lambda_j) + I_{xu}(\lambda_j)] + (\tilde{\beta} - \beta)^2 I_{xx}(\lambda_j)$$

and defining

$$r_j := \frac{I_{\tilde{u}\tilde{u}}(\lambda_j) - I_{uu}(\lambda_j)}{g_u(\lambda_j)}$$

we can bound the effect of slope estimation on the residual periodogram with

$$|r_j| \leqslant 2|\tilde{\beta} - \beta| \frac{|I_{ux}(\lambda_j)|}{g_u(\lambda_j)} + (\tilde{\beta} - \beta)^2 \frac{I_{xx}(\lambda_j)}{g_u(\lambda_j)}. \tag{3}$$

This effect depends directly on the convergence rate of $\tilde{\beta} - \beta$ so, for original residuals, it depends on the values of d and δ, and also on whether $d + \delta < 1$ or not, and on whether $d < 1$ or not (Assumption 1). Now, similar results to Lemma 1 hold for the original residuals periodogram, even in some nonstationary situations. From Velasco (1999a) and Hurvich and Ray (1995, Theorem 1), we obtain Lemma 2:

LEMMA 2. *If Assumptions* 2 *and* 3 *hold and* $d_z \in (\ \frac{1}{2},\frac{3}{2})$, *for* $j = 1,\ldots,m$, $m/T \to 0$ *as* $T \to \infty$,

$$\begin{aligned} \mathrm{E}[I_{zz}(\lambda_j)] &= f_z\{(\lambda_j)(1+O[j^{-1}+j^{2(d_z-1)}]\log(j+1)\} = O(\lambda_j^{-2d_z}) \qquad && 0.5 < d_z < 1 \\ &= O(\lambda_j^{-2d_z} j^{2(d_z-1)}) && 1 \leqslant d_z < 1.5. \end{aligned}$$

This lemma shows that the periodogram is asymptotically unbiased for f_z at frequencies close to the origin as long as $d_z < 1$ but, when $d_z \geqslant 1$ and $d_z \leqslant -0.5$, this property no longer holds and we may require tapering to control this bias and the magnitude of r_j or r_j^*.

3. GAUSSIAN SEMI-PARAMETRIC ESTIMATE WITH DIFFERENCED RESIDUALS

We analyse inference on δ, based on the differenced residuals $\Delta\tilde{u}_t$, which we may expect to have memory close to $\delta - 1$, so only nonstationary errors u_t ($\delta > 0.5$) could be considered in principle to avoid non invertibility problems. Consider the semi parametric estimate of Robinson (1995b), proposed initially by Künsch (1987), based on a local Gaussian Whittle likelihood on the frequency domain,

$$Q^*(\delta, G_u) = \frac{1}{m}\sum_{j\ 1}^{m}\left[\log G_u\lambda_j^{-2(\delta-1)} + \frac{I_{\Delta\tilde{u}\Delta\tilde{u}}(\lambda_j)}{G_u\lambda_j^{-2(\delta-1)}}\right]$$

with $1/m + m/T \to 0$ as $T \to \infty$, where the constant G_u can be concentrated out. For $m = T$, this is approximately minus twice the log likelihood, up to constants, of the data under Gaussianity, but this assumption is not required for our analysis. We calculate the periodogram $I_{\Delta\tilde{u}\Delta\tilde{u}}(\lambda_j)$ using the differenced residuals $\Delta\tilde{u}_t$. Then, following the discussion in Robinson (1995b), we define

$$\tilde{\delta}^* = \widetilde{\delta\quad 1} + 1$$

where

$$\widetilde{\delta\quad 1} = \arg\min_{\delta\in\Theta} R^*_{\Delta\tilde{u}}(\delta)$$

and

$$R^*_{\Delta\tilde{u}}(\delta) = \log \tilde{G}^*_{\Delta\tilde{u}}(\delta) \quad 2(\delta \quad 1)\frac{1}{m}\sum_{1}^{m}\log\lambda_j$$

$$\tilde{G}^*_{\Delta\tilde{u}}(\delta) = \frac{1}{m}\sum_{1}^{m}\lambda_j^{2(\delta-1)} I_{\Delta\tilde{u}\Delta\tilde{u}}(\lambda_j)$$

and the true value $\delta_o \in \Theta = [\nabla_1, \nabla_2]$, $0.5 < \nabla_1, \nabla_2 < d < 1.5$. Denote by δ any admissible value of the parameter. For the analysis of the asymptotic properties of $\tilde{\delta}^*$, we take the following assumptions from Robinson (1995b):

Assumption 4. *For $v_t = u_t$, $0 \leqslant \delta < \frac{1}{2}$, or for $v_t = \Delta u_t$, $\frac{1}{2} \leqslant \delta < \frac{3}{2}$,*

$$u_t = \sum_{j\ 0}^{\infty} \alpha_j \epsilon_{t-j} \qquad \sum_{j\ 0}^{\infty} \alpha_j^2 < \infty$$

where

$$\mathrm{E}[\epsilon_t | \mathcal{F}_{t-1}] = 0 \qquad \mathrm{E}[\epsilon_t^2 | \mathcal{F}_{t-1}] = 1 \qquad a.s., t = 0, \pm 1, \ldots$$

in which $\mathcal{F}_t$ is the σ field of events generated by ϵ_t, $s \leqslant t$, and there exists a random variable ϵ, such that $\mathrm{E}\epsilon^2 < \infty$ and for all $\eta > 0$ and some $C > 0$, $P(|\epsilon_t| > \eta) \leqslant CP(|\epsilon| > \eta)$.

Assumption 5. *In a neighbourhood $(0, \varepsilon)$ of the origin,*

$$\alpha(\lambda) = \sum_{j\ 0}^{\infty} \mathrm{e}^{\mathrm{i}j\lambda} \alpha_j$$

is differentiable and

$$\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} \alpha(\lambda) \right| = O\left(\frac{|\alpha(\lambda)|}{\lambda} \right) \qquad as\ \lambda \to 0^+$$

Assumption 5 implies Assumption 3, since $f_v(\lambda) = |\alpha(\lambda)|^2/(2\pi)$. This assumption is standard in the parametric literature, as well as the linear process Assumption 4 (Giraitis and Surgailis, 1990; Hosoya, 1997); it is only used for the asymptotic distribution, but is not necessary for consistency of estimates. For the derivation of the asymptotic distribution of estimates, we also need finite fourth moments, while the restriction of constant conditional innovations variances could be relaxed by assuming boundedness of the eighth moment as in Robinson and Henry (1999):

Assumption 6. *Assumption 4 holds and also*

$$\mathrm{E}[\epsilon_t^3 | \mathcal{F}_{t-1}] = \mu_3 \qquad a.s. \quad [\epsilon_t^4 | \mathcal{F}_{t-1}] = \mu_4 \qquad a.s. \qquad t = 0, \pm 1, \ldots$$

for finite constants μ_3 and μ_4.

We now present our first result, indicating the implications for the particular choice of m as a power of T, which is usually considered in practical applications because the choice of m does not depend in this case on conditions established in terms of the unknown d and δ_o. We stress log T consistency, because estimates of δ may be used for studentization of other estimates, like those of β, whose convergence rate depends of δ (Robinson, 1994a).

Theorem 1. *Under Assumptions* 1 4, *with* $\delta_o \in (\frac{1}{2}, d)$, *and as* $T \to \infty$

$$m^{\delta_o - d} \log^2 T + mT^{-1} \to 0 \tag{4}$$

then

$$\log T(\tilde{\delta}^* \quad \delta_o) \to_p 0.$$

If $m \sim CT^a$, as $T \to \infty$, then (4) holds for any $a \in (0, 1)$. Note that, under the conditions of Theorem 1, $d + \delta_o > 1$ holds, so only Case I of Assumption 1 applies.

We now give sufficient conditions for the asymptotic normality of $\tilde{\delta}^*$ when the cointegrating residuals have noticeably less memory than the observed series, $d \quad \delta_o > \frac{1}{2}$, so the estimate $\tilde{\beta}$ used for residual calculation has a faster convergence rate than $T^{1/2}$. Now $\delta_o \in \Theta = [\nabla_1, \nabla_2], \frac{1}{2} < \nabla_1 < \nabla_2 < d \quad \frac{1}{2} < 1$.

Theorem 2. *Under Assumptions* 1 6 *with* $\delta_o \in (\frac{1}{2}, d \quad \frac{1}{2})$, *and as* $T \to \infty$

$$m^{-1} + m^{1+2\gamma} T^{-2\gamma} \log^2 m \to 0 \tag{5}$$

then

$$m^{1/2}(\tilde{\delta}^* \quad \delta_o) \xrightarrow{d} \mathrm{N}(0, \tfrac{1}{4}).$$

Thus the asymptotic distribution of $\tilde{\delta}^*$ remains the same as if Δu_t where actually observed when residuals satisfying Assumption 1 are used instead. If $m \sim CT^a$, then (5) holds if $0 < a < 2\gamma/(1 + 2\gamma)$.

Tapering. Using tapered differenced residuals in the periodogram calculation, we conjecture that it should be possible to estimate δ consistently when Δu_t is non invertible, $0 \leqslant \delta_o \leqslant 0.5$, though possibly restricting the allowed set of values of d and δ_o when $d + \delta_o \leqslant 1$ as in Theorems 3 and 4 (because of Cases II and III in Assumption 1). Then, we could complete the consistency analysis of $\tilde{\delta}^*$ for $0 \leqslant \delta_o < d < 1.5$ and the asymptotic distribution for $0 \leqslant \delta_o < d - 0.5 < 1$ using differenced residuals and the full cosine bell taper

$$h_t = \frac{1}{2}\left(1 \quad \cos\frac{2\pi t}{T}\right)$$

if we strengthen Assumption 3 for u_t with Assumption 7:

Assumption 7. *Assumption* 3 *holds and the derivative* $f_v'(\lambda)$, *satisfies for some* $\alpha > 0$ *and* $|\mu| < \lambda/2$,

$$|f_v'(\lambda + \mu) \quad f_v'(\lambda)| = O(|f_v'(\lambda)| \lambda^{-\alpha} |\mu|^{\alpha}) \qquad as \quad \lambda \to 0^+.$$

This condition generalizes 'higher order bias' assumptions used by Robinson (1994b) and Giraitis *et al.* (1997) for semi parametric estimation of the memory of long memory stationary series, and by Velasco (1999a, b), for nonstationary series. As a result of tapering, the asymptotic variance of the memory estimate is multiplied by the factor $\Phi = 35/18$ for the cosine taper, and other tapers would lead to similar results (Hurvich and Chen, 2000).

4. GAUSSIAN SEMI-PARAMETRIC ESTIMATE WITH ORIGINAL RESIDUALS

We now consider the semi parametric estimate of δ based on the minimization of the low frequency local likelihood

$$Q(\delta, G_u) = \frac{1}{m}\sum_{j \quad 1}^{m}\left\{\log G_u \lambda_j^{-2\delta} + \frac{I_{\Delta\tilde{u}\Delta\tilde{u}}(\lambda_j)}{G_u \lambda_j^{-2\delta}}\right\}$$

for the original residuals, $\tilde{u}_t$. *A priori*, this alternative approach is more appropriate than the one based on differenced residuals if the cointegrating errors u_t have memory $\delta_o \leqslant 0.5$, when differencing may lead to non invertibility. However, we show now that it can also be used consistently for moderate nonstationary errors.

Let for $\delta_o \in \Theta = [\nabla_1, \nabla_2]$, $-0.5 < \nabla_1 < \nabla_2 < \min\{1, d\}$,

$$\tilde{\delta} = \arg\min_{\delta \in \Theta} R_{\tilde{u}}(\delta)$$

where

$$R_{\tilde{u}}(\delta) = \log \tilde{G}_{\tilde{u}}(\delta) \quad 2\delta \frac{1}{m}\sum_{1}^{m} \log \lambda_j$$

$$\tilde{G}_{\tilde{u}}(\delta) = \frac{1}{m}\sum_{1}^{m} \lambda_j^{2\delta} I_{\tilde{u}\tilde{u}}(\lambda_j).$$

The next theorem gives sufficient conditions for the consistency of $\tilde{\delta}$

THEOREM 3. *Under Assumptions* 1 4 *with* $\delta_o \in [0,1)$, $\delta_o < d$, *and for some* $\epsilon > 0$, *as* $T \to \infty$,

$$\left[\left\{m^{\delta_o - d}\log m + m^{\epsilon - 1/(4e)}\right\}\xi_T + m^{\delta_o - d} + m^{\delta_o - 1} + m^{\epsilon - 1}\right]\log^2 T + mT^{-1} \to 0, \quad (6)$$

where

$$\xi_T = \begin{cases} \log^2 T & \text{when } \delta_o + d = 1, \delta_o > 0 \\ T^{1-\delta_o - d} & \text{when } \delta_o + d < 1 \\ 0 & \text{otherwise} \end{cases}$$

then $\log T(\tilde{\delta} \quad \delta_o) \to_p 0$.

If $m \sim CT^a$, then (6) holds if

$$\max\left\{0, \frac{1 \quad \delta_o \quad d}{d \quad \delta_o}, \ 4e(1 \quad \delta_o \quad d)\right\} < a < 1.$$

Notice that, when $d+\delta_o \geqslant 1$, we can choose a arbitrarily close to 0 (including the CI(1, 0) case) but that, otherwise, it could be not possible to choose $a < 1$ (or m such that $mT^{-1} \to 0$) for some combination of values of d and δ_o, since (6) requires

$$d + \delta_o > 1 \quad \frac{1}{4\mathrm{e}} \approx 0.91$$

Thus our results do not cover cointegrating regressions with mildly nonstationary data (d close to 0.5) and nearly weakly dependent residuals (δ_o close to 0), which are globally stationary, $\delta_o + d < 1$, such as CI(0, d), for any $0.5 < d \leqslant 0.82$. If narrow band LS estimates are employed, Corollary 1 covers all posible situations for all $a \in (0,1)$:

Corollary 1. *Using a sequence n growing arbitrarily slowly with T in Case IV of Assumption* 1, *the conclusions of Theorem* 3 *hold replacing* (6) *by* (4).

For the asymptotic normality of $\widetilde{\delta}$, we need to assume the same conditions on $d - \delta_o$ as when using differenced residuals, to ensure a fast enough convergence rate for $\widetilde{\beta}$, but we are not able to cover nonstationary errors with $0.5 \leqslant \delta_o < 0.75$ as is possible for observed data (Velasco, 1999b). Set $\delta_o \in \Theta = [\nabla_1, \nabla_2]$, $-0.5 < \nabla_1 < \nabla_2 < 0.5$.

Theorem 4. *Under Assumptions* 1 6, *with* $\delta_o \in [0, \frac{1}{2}), d \quad \delta_o > \frac{1}{2}$, *and for some* $\epsilon > 0$, *as* $T \to \infty$,

$$\left\{ m^{\delta_o - d + 1/2} \log m + m^{\epsilon - 1/(4e)} \right\} \xi_T + m^{-1} + m^{1+2\gamma} T^{-2\gamma} \log^2 m \to 0 \tag{7}$$

then

$$m^{1/2}(\tilde{\delta} \quad \delta_o) \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{4}\right).$$

If $m \sim CT^a$ then (7) holds if

$$\max\left\{0, \frac{1 \quad \delta_o \quad d}{d \quad \delta_o \quad 0.5}, 4\mathrm{e}(1 \quad \delta_o \quad d)\right\} < a < \frac{2\gamma}{1 + 2\gamma}.$$

When $d+\delta_o \geqslant 1$, we can choose a arbitrarily close to 0 again and therefore some $a > 0$ satisfying (7) exist for any value of γ but, when $d + \delta_o < 1$, required choice of a (or m in general) may not be possible. Thus, for $\gamma = 2$, it is always possible to choose $a < \frac{4}{5}$ so m satisfies (7) when

$$9d + \delta_o > 7 \qquad \text{and} \qquad d + \delta_o > 1 \quad \frac{1}{5\mathrm{e}}$$

and both hold when, for example, $d > 1 - 1/5\mathrm{e} \approx 0.93$. In this case, the use of narrow band LS instead of OLS estimates, allows asymptotic inference based on $\tilde{\delta}$ when $d + \delta_o < 1$ for all cases, avoiding further restrictions beyond stationarity of errors, $\delta_o \in [0, \frac{1}{2})$, and superconsistency, $d \quad \delta_o > \frac{1}{2}$:

Corollary 2. *Using a sequence n growing arbitrarily slowly with T in Case IV of Assumption* 1, *the conclusions of Theorem* 4 *hold replacing* (7) *by* (5).

Tapering. To deal with nonstationary but undifferenced residuals, it can be shown that, with tapered residuals, $\tilde{\delta}$ is consistent up to $\delta_o < d < 1.5$, and also asymptotically normal for any $\delta_o < d - 0.5 < 1$, under the same conditions of Theorems 3 and 4 (together with Assumption 7), just adjusting the asymptotic variance as for observed data (Velasco, 1999b, Thms 5 and 6).

Corollary 3. *For tapered residuals, if additionally Assumption* 7 *is satisfied, Theorem* 3 *holds for* $\delta_o \in [0, \frac{3}{2})$ *and Theorem* 4 *holds when* $\delta_o \in [0, 1), d \quad \delta_o > \frac{1}{2}$, *increasing the asymptotic variance by a factor* $\Phi = 35/18$.

Comparing with the semi parametric alternative of running a residual based log periodogram regression (Hassler *et al.*, 2002), there are certain similarities and differences. On the one hand, log periodogram regressions also have the upper limit $\delta_o < d - 0.5 < 1$ for the asymptotic normality of estimates of δ, because the convergence rate of $\tilde{\beta}$ in Assumption 1. This limitation is independent of the specific method employed, or whether this is carried out with or without tapering, with original or differenced residuals. Similarly, the conditions

$$\frac{1 \quad d \quad \delta_o}{d \quad \delta_o} < 1$$

for consistency and

$$\frac{1 \quad d \quad \delta_o}{d \quad \delta_o \quad 0.5} < \frac{2\gamma}{2\gamma + 1}$$

for asymptotic normality are required when original residuals are used in the log periodogram regression.

On the other hand, a consistent residual log periodogram regression further requires $d - \delta_o > 0.5$ (while Gaussian estimates only need $d - \delta_o > 0$), though this restriction can be avoided, at least in part, by use of pooled log periodograms. Furthermore, the analysis of residual log periodogram regressions is highly complicated for non Gaussian series and requires a trimming of the very first periodogram ordinates.

5. JOINT ESTIMATION OF MEMORY PARAMETERS

To test joint hypothesis on the parameter vector $\theta = (\delta, d)'$ or to increase the efficiency of our previous estimates, of interest is the joint estimation of the parameter vector θ, building on initial slope estimates satisfying Assumption 1. A particular interesting hypothesis to be tested is that of $d \quad \delta > \frac{1}{2}$, an assumption in

Theorems 2 and 4. In this section, we analyse the estimation of (δ, d) using the procedure proposed by Lobato (1999) for multi variate observed time series, extending Robinson's (1995b) univariate method. This is a two step procedure based on a local multiple Gaussian likelihood and requires initial estimates of the parameters with the same semi parametric rate of convergence, e.g. those of Theorems 2 or 4.

Another possibility not studied here is a two step Gaussian semiparametric estimate of the full parameter vector $\theta = (\delta, d, \beta)'$, based on the same local likelihood around zero frequency. This approach is similar to the Smith and Chen (1997) semi parametric Gaussian estimation of the memory parameter and regression coefficients for deterministic regressors observed with long memory errors. Alternatively, Dueker and Startz (1998) proposed a full parametric vector ARFIMA model, where the series acting as the dependent variable is replaced by the unobserved cointegrated errors, and, if nonstationary, the original series can be replaced by its differences. However, none of these references provide justification for valid inference.

Different inputs (i.e. combinations of original and differenced series) can be used in the estimation. We first concentrate on original residuals and assume now that Assumptions 2 and 3 hold also for the cross spectral density of u_t, $\delta < 0.5$ (Δu_t, $0.5 \leqslant \delta < 1.5$), and Δx_t for the same $\gamma > 0$ so, $0 < \gamma \leqslant 2$,

$$\mathbf{f}(\lambda) = \begin{pmatrix} f_{uu}(\lambda) & f_{u\Delta x}(\lambda) \\ f_{\Delta xu}(\lambda) & f_{\Delta x\Delta x}(\lambda) \end{pmatrix} = \Lambda^{-1}\mathbf{G}\Lambda^{-1}(1 + O(\lambda^{\gamma}))$$

as $\lambda \to 0^+$, where $\Lambda = \Lambda(\theta) = \mathrm{diag}\{\lambda^{\delta}, \lambda^{d-1}\}$ and

$$G = \begin{pmatrix} G_u & G_{ux} \\ G_{xu} & G_x \end{pmatrix}$$

is real, nonsingular and symmetric (Lobato, 1997). We work directly with Δx_t since $d > 0.5$. We could consider estimation for $\delta \geqslant 0.5$ working with original residuals u_t and applying tapering. If it is known that $\delta > 0.5$, we can alternatively substitute u_t by Δu_t and δ by $\delta - 1$ in Λ as we analyse in Theorem 6 below.

Define the periodogram matrix of $(\tilde{u}_t, \Delta x_t)$ as

$$\mathbf{I}(\lambda_j) = \mathbf{I}(\lambda_j, \beta) = \mathbf{w}(\lambda_j)\mathbf{w}^*(\lambda_j)$$

where the * superscript indicates simultaneous transposition and complex con jugation and

$$\mathbf{w}(\lambda_j) = (2\pi T)^{-1}\sum_{t\ 1}^{T}\begin{pmatrix} y_t & \tilde{\beta}x_t \\ & \Delta x_t \end{pmatrix}\exp(\mathrm{i}\lambda_j t) = \begin{pmatrix} w_y(\lambda_j) & \tilde{\beta}w_x(\lambda_j) \\ & w_{\Delta x}(\lambda_j) \end{pmatrix}.$$

We take the following assumptions from Lobato (1999), where $v_t = (u_t, \Delta x_t)'$, $\delta < 0.5$, which are vector extensions of Assumptions 4 6, and imply Assumption 3.

Assumption 8.

$$v_t = \mu + \sum_{j=0}^{\infty} A_j \varepsilon_{t-j} \qquad \sum_{j=0}^{\infty} \|A_j\|^2 < \infty$$

where $\| \, \|$ *denotes the supremum norm and* ε_t *satisfies*

$$\mathrm{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$$

$$\mathrm{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = I_2$$

$$\mathrm{E}(\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t) | \mathcal{F}_{t-1}) = \mu_{abc} \qquad |\mu_{abc}| < \infty \text{ for } a,b,c = 1,2$$

$$\mathrm{E}(\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t)\varepsilon_d(t) | \mathcal{F}_{t-1}) = \mu_{abcd} \qquad |\mu_{abcd}| < \infty \text{ for } a,b,c,d = 1,2$$

where $\mathcal{F}_{t-1}$ *is the* σ *field of events generated by* $\{\varepsilon_s, \; s \leqslant t-1\}$.

Assumption 9. *As* $\lambda \to 0^+$

$$\frac{\mathrm{d}A_a(\lambda)}{\mathrm{d}\lambda} = O(\lambda^{-1} ||A_a(\lambda)||)$$

for $a = 1, 2$, *where* $A_a(\lambda)$ *is the a th row of*

$$A(\lambda) = \sum_{j=0}^{\infty} A_j \exp(\mathrm{i}j\lambda).$$

We consider the local multi variate Whittle likelihood

$$Q(\mathbf{G}, \delta, d) = \frac{1}{m} \sum_{j=1}^{m} \{\log |\Lambda_j^{-1} \mathbf{G} \Lambda_j^{-1}| + \mathrm{tr}[(\Lambda_j^{-1} \mathbf{G} \Lambda_j^{-1})^{-1} \mathrm{Re}[\mathbf{I}(\lambda_j)]]\}$$

with $1/m + m/T \to 0$ as $T \to \infty$, where $\Lambda_j = \Lambda_j(\delta, d) = \mathrm{diag}\{\lambda_j^{\delta}, \lambda_j^{d-1}\}$. We can concentrate out the matrix $\mathbf{G}$ in Q, setting

$$\hat{\mathbf{G}}(\delta, d) = \frac{1}{m} \sum_{j=1}^{m} \Lambda_j \mathrm{Re}[\mathbf{I}(\lambda_j)] \Lambda_j$$

while the concentrating likelihood is

$$\mathcal{L}(\delta, d) = -2(\delta + d - 1) \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j + \log |\hat{\mathbf{G}}(\delta, d)|.$$

We suppose, for simplicity, that $m^{1/2}$ consistent estimates $\tilde{\delta}$ and $\tilde{d}$ of δ_o and d_o, and a $T^{d_o - \delta_o}$ consistent estimate $\tilde{\beta}$ of β_o are available, so we suppose (if we wish to use the non tapered univariate estimates of previous sections) that only Cases I and II of Assumption 1 can hold, though some combinations $\delta_o + d_o < 1$ may allow consistent estimation under further assumptions (see Theorem 4) or with improved regression estimates (see Corollary 2). The two step estimate of the parameter vector θ is

$$\hat{\theta} = \tilde{\theta} - \left(\mathcal{L}''_{\theta\theta}(\tilde{\theta})\right)^{-1} \mathcal{L}'_{\theta}(\tilde{\theta})$$

where $\tilde{\theta} = (\tilde{\delta}, \tilde{d})'$ and

$$\mathcal{L}''_{\theta\theta}(\theta) = \frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta\partial\theta'}$$

$$\mathcal{L}'_{\theta}(\theta) = \frac{\partial \mathcal{L}(\theta)}{\partial\theta}.$$

Setting θ_o and $\mathbf{G}_o$ for the true values, we obtain the following result using the previous arguments and Lobato (1999).

Theorem 5. *Under Assumptions* 1 (*Cases* I *and* II), 8, 9, $0 \leqslant \delta_o < \min\{\frac{1}{2}, d_o - \frac{1}{2}\}, \delta_o + d_o > 1$, *and for* $\eta = \min\{\gamma, d_o - \delta_o\}$,

$$m^{-1} + m^{1+2\eta} T^{-2\eta} \log^2 m \to 0 \qquad \text{as } T \to \infty \tag{8}$$

then

$$\sqrt{m}(\hat{\theta} - \theta_o) \xrightarrow{d} \mathcal{N}(0, E^{-1}),$$

where $E = 2(I + \mathbf{G}_o \circ \mathbf{G}_o^{-1})$ *and* $\circ$ *is the Hadamard product.*

The unknown constants in E can be estimated consistently as, for example, in Lobato and Velasco (2000). Again, the estimates of the memory parameters (δ, d) retain the same convergence rate and asymptotic distribution as if β were known. The condition (8) provides a symmetric set up generalizing (5) to account for the regularity parameter of y_t in terms of γ and the cointegration order, $d_o - \delta_o$, since in applications we cannot distinguish y_t from x_t in model (1).

Alternatively, if it is known that $\delta_o > 0.5$, we could calculate the (tapered) periodogram matrix

$$\mathbf{I}^{\Delta}(\lambda_j) = \mathbf{w}^{\Delta}(\lambda_j)\mathbf{w}^{\Delta}(\lambda_j)^*$$

of the differenced residuals and regressors, $(\Delta\tilde{u}_t, \Delta x_t)'$, using

$$\mathbf{w}^{\Delta}(\lambda_j) = (2\pi T)^{-1} \sum_{t=1}^{T} h_t \begin{pmatrix} \Delta(y_t - \tilde{\beta} x_t) \\ \Delta x_t \end{pmatrix} \exp(\mathrm{i}\lambda_j t) = \begin{pmatrix} w_{\Delta y}(\lambda_j) - \tilde{\beta} w_{\Delta x}(\lambda_j) \\ w_{\Delta x}(\lambda_j) \end{pmatrix}$$

In this case, we set in Assumptions 8 and 9 that $v_t = (\Delta u_t, \Delta x_t)'$ and substitute Λ by

$$\Lambda^{\Delta} = \Lambda^{\Delta}(\theta) = \operatorname{diag}\{\lambda^{\delta-1}, \lambda^{d-1}\}.$$

The concentrating likelihood is now

$$\mathcal{D}(\delta, d, \beta) = \quad 2(\delta + d \quad 2)\frac{1}{m}\sum_{j\ \ 1}^{m}\log\lambda_j + \log|\hat{\mathbf{G}}^{\Delta}(\delta, d, \beta)|$$

with

$$\hat{\mathbf{G}}^{\Delta}(\delta, d, \beta) = \frac{1}{m}\sum_{j\ \ 1}^{m}\Lambda_j^{\Delta}\mathrm{Re}[\mathbf{I}^{\Delta}(\lambda_j)]\Lambda_j^{\Delta}$$

and $\Lambda_j^{\Delta} = \mathrm{diag}\{\lambda_j^{\delta-1}, \lambda_j^{d-1}\}$, and the properties of the two step estimate of the parameter vector θ,

$$\widehat{\theta^{\Delta}} = \tilde{\theta} \quad \left(\mathcal{D}''_{\theta\theta}(\tilde{\theta})\right)^{-1}\mathcal{D}'_{\theta}(\tilde{\theta})$$

are given in Theorem 6.

Theorem 6. *Under Assumptions* 1, 8, 9, $\frac{1}{2} \leqslant \delta_o < d_o \quad \frac{1}{2} < 1$, *and* (8),

$$\sqrt{m}(\widehat{\theta^{\Delta}} \quad \theta_o) \xrightarrow{d} \mathrm{N}(0, E^{-1}).$$

For checking the condition $d \quad \delta > \frac{1}{2}$, required for all our asymptotic distribution analysis, we can use the previous results by means of the t statistic based on the estimates of the memory parameters,

$$\tau_m(r) = \frac{\sqrt{m}\left(\hat{d} \quad \hat{\delta} \quad r\right)}{((\ \ 1, 1)\hat{E}^{-1}(\ \ 1, 1)')^{1/2}}$$

where $\hat{E}$ is an appropriate consistent estimate of E. Thus, under the assumptions of Theorem 5 or 6 and the null hypothesis that $d \quad \delta = r_o$, $\tau_m(r_o) \xrightarrow{d} \mathrm{N}(0, 1)$ for any $r_o > \frac{1}{2}$. Therefore, if in applications $\tau_m(0.5) > -z_{\alpha}$, where $\Pr(\mathrm{N}(0,1) > z_{\alpha}) = \alpha$, we can hope that our sufficient conditions for Gaussian semi parametric inference on (d, δ) hold as we can not reject the null hypothesis at the α significance level for some r_o small enough. However, no power analysis or feasible testing for the null of no cointegration ($d - \delta = 0$) can be deduced from our results, since for $0 \leqslant d - \delta = r \leqslant 0.5$ the asymptotic distribution of $\tau_m(r)$ depends on that of $\tilde{\beta}$.

Tapering could permit to cross the lower boundary for δ_o, but always keeping $d - \delta_o > 0.5$ for root m consistency and asymptotic normality. These procedures could cover more general situations, e.g. models with several regressors x_{kt}, $k = 1, \ldots, N$, of the same memory $d = d_k$ as y_t, defining a local likelihood $Q(\mathbf{G}, \delta, d)$ in terms of the vector $(u_t, \Delta x_{1t}, \ldots, \Delta x_{Nt})$, where

$$u_t = y_t \quad \beta_1 x_{1t} \quad \cdots \quad \beta_N x_{Nt}$$

is replaced by residuals. If, on the other hand, different memory levels d_k are allowed, further complications may arise, since at least two of the original $N + {}^{-}1$ variables should share a maximal memory parameter and $\delta < \min d_k$. In any case, the choice of the dependent variable y_t among the available series can be guided by maximizing the value of the local likelihood Q or Q^{Δ}. We illustrate some of these issues in Section 7 with a real data example.

6. SIMULATION WORK

We perform a limited Monte Carlo analysis of some of the procedures suggested in this paper. We simulate I(1) cointegrated bivariate time series, with I(0) errors and estimate the two memory parameters with the following methods, residuals based on initial OLS estimates of β:

1 Gaussian semi parametric estimate $\tilde{d}$, based on Δx_t.
2 Residual Gaussian semi parametric estimate $\tilde{\delta}$, based on original OLS residuals $\tilde{u}_t$.
3 Gaussian joint two step semi parametric estimate $(\hat{\delta}, \hat{d})$, based on $(\tilde{u}_t,\ \Delta x_t)$.

We simulate time series with two sample sizes, $T = 192$, 384, and use bandwidth numbers $m = 25$, 50 and $m = 40$, 80 respectively. We try non tapered and tapered series with the cosine bell and Zhurbenko taper of order 2, which is given by the triangular window, so we use sample sizes that are multiples of 6 for simplicity of computations.

We use two data generating mechanisms, $\beta = 1$,

Model 1: $\Delta x_t \sim \text{NID}(0, 1)$ and $u_t \sim \text{NID}(0, 2)$, independent.
Model 2: $\Delta x_t \sim \text{NID}(0, 1)$ and $(1-0.3L)u_t \sim \text{NID}(0, 2)$, with correlation equal to 0.3.

The second model incorporates short term effects in the error series and endogeneity with the regressors. Note that $d_o - \delta_o = 1$ and $d_o + \delta_o = 1$ but we are in Case I of Assumption 1, and Theorems 4 and 6 apply.

The results are summarized in Tables I (Model 1) and II (Model 2). We do not report simulation results for the Zhurbenko taper, as they are similar to those with the cosine taper, but with small variability increments. We also tried to skip the first $\ell \geqslant 1$ frequencies in the estimates, but in no case this trimming improved the results, by contrast to the residual log periodogram estimate (Hassler *et al.*, 2002). We calculate the bias, standard error and mean square error across 1000 simulations, and give in parenthesis the asymptotic standard error from Theorem 5 for joint estimation of memory parameters (which in Table I coincides with the values for the individual estimation). These are the main conclusions we can draw from the simulations:

- Joint estimation of $(\delta,\ d)$ is always recommended compared to individual estimation of δ and d (although no differences arise in Table I because x_t and u_t are independent).

TABLE I

Simulation Results for Model 1

	T	m	δ, d	Univariate		Multi-variate	
				$\tilde{\delta}$	$\tilde{d}$	$(\hat{\delta}$	$\hat{d})$
No taper	192	25	bias	0.0482	0.0136	0.0466	0.0164
			sd (0.100)	0.1430	0.1338	0.1388	0.1303
			mse	0.0228	0.0181	0.0214	0.0172
		50	bias	0.0254	0.0090	0.0247	0.0084
			sd (0.071)	0.0919	0.0840	0.0915	0.0837
			mse	0.0091	0.0071	0.0090	0.0071
	384	40	bias	0.0245	0.0057	0.0248	0.0063
			sd (0.079)	0.1006	0.0964	0.1000	0.0952
			mse	0.0107	0.0093	0.0106	0.0091
		80	bias	0.0124	0.0033	0.0128	0.0030
			sd (0.056)	0.0638	0.0630	0.0634	0.0627
			mse	0.0043	0.0040	0.0042	0.0039
Cosine taper	192	25	bias	0.0279	0.0233	0.0275	0.0259
			sd (0.139)	0.2271	0.2098	0.2225	0.2038
			mse	0.0524	0.0446	0.0503	0.0422
		50	bias	0.0155	0.0155	0.0153	0.0148
			sd (0.110)	0.1412	0.1306	0.1396	0.1294
			mse	0.0202	0.0173	0.0197	0.0170
	384	40	bias	0.0107	0.0175	0.0105	0.0185
			sd (0.099)	0.1506	0.1466	0.1487	0.1442
			mse	0.0228	0.0218	0.0222	0.0211
		80	bias	0.0123	0.0039	0.0126	0.0039
			sd (0.078)	0.0640	0.0640	0.0636	0.0637
			mse	0.0042	0.0041	0.0042	0.0041

- Estimation of δ is more difficult than that of d, as expected by the use of residuals, especially in terms of bias for the second model. In this model, the bias of $\tilde{\delta}$ and $\hat{\delta}$ seems very sensitive to bandwidth choice.
- The variability is often larger than the asymptotic variance predicted by the central limit theorems, although, for the first model, the approximation is reasonable and improves with sample size.
- Tapering always led to more variability as expected but, in this case, gave no extra bias protection.

7. EMPIRICAL EXAMPLE

We apply in this section the above procedures to three US monetary aggregates for the years 1978 1999 (264 monthly observations). The data are taken from the St Louis Federal Reserve Bank for three series, LTD (large denomination time deposits), M2 and M3. We work with the increment rates of the original series, $m2_t$, $m3_t$ and ltd_t, e.g. $m2_t = \Delta \log M2_t$, see Figure 1. In fact, M2 and LTD are two of the components of the larger aggregate M3, so we are trying to analyse

TABLE II

SIMULATION RESULTS FOR MODEL 2

	T	m	δ, d	Univariate		Multi-variate	
				$\tilde{\delta}$	$\tilde{d}$	$(\hat{\delta}$	$\hat{d})$
No taper	192	25	bias	0.0112	0.0136	0.0087	0.0228
			sd (0.098)	0.1400	0.1338	0.1353	0.1287
			mse	0.0197	0.0181	0.0184	0.0171
		50	bias	0.1163	0.0090	0.1105	0.0168
			sd (0.069)	0.0905	0.0840	0.0891	0.0830
			mse	0.0217	0.0072	0.0201	0.0072
	384	40	bias	0.0125	0.0057	0.0101	0.0102
			sd (0.077)	0.1007	0.0964	0.0976	0.0932
			mse	0.0103	0.0093	0.0096	0.0088
		80	bias	0.0858	0.0039	0.0783	0.0119
			sd (0.05)	0.0641	0.0640	0.0624	0.0630
			mse	0.0115	0.0041	0.0100	0.0041
Cosine taper	192	25	bias	0.0349	0.0233	0.0319	0.0329
			sd (0.136)	0.2281	0.2098	0.2200	0.2034
			mse	0.0532	0.0446	0.0494	0.0424
		50	bias	0.1416	0.0155	0.1352	0.0242
			sd (0.096)	0.1417	0.1306	0.1386	0.1285
			mse	0.0401	0.0173	0.0375	0.0171
	384	40	bias	0.0305	0.0175	0.0270	0.0236
			sd (0.108)	0.1504	0.1466	0.1470	0.1420
			mse	0.0235	0.0218	0.0223	0.0207
		80	bias	0.1039	0.0113	0.0957	0.0204
			sd (0.076)	0.0987	0.0981	0.0967	0.0965
			mse	0.0205	0.0098	0.0185	0.0097

indirectly the contribution of the remaining components of M3 (eurodollar deposits, institutional money market mutual funds, etc.) to the persistence properties of its growth rate, $m3_t$. For all semi parametric estimates, we take $m = 20$ and employ non tapered estimates and non differenced residuals. This bandwidth choice is not optimally estimated and only made so as to avoid seasonal frequencies, which may affect the results if included in the objective functions. Numbers in parenthesis in Tables III and IV are estimated standard errors for estimates and p values for test statistics.

A preliminary analysis of the memory of the series using a multi variate 2 step Gaussian estimate (Lobato, 1999) on ΔX_t, where $X_t = (\mathrm{ltd}_t\ \mathrm{m2}_t,\ \mathrm{m3}_t)$, procedures $\hat{d}_{\mathrm{ltd}} = 0.501$ (0.08), $\hat{d}_{\mathrm{m2}} = 0.553$ (0.07) and $\hat{d}_{\mathrm{m3}} = 0.548$ (0.07), while a Wald test of equal memory based on these estimates cannot reject the null hypothesis of common memory $d_X = d_{\mathrm{ltd}} = d_{\mathrm{m2}} = d_{\mathrm{m3}}$ with p value 0.81. In this case, the common memory estimate is $\hat{d}_X = 0.539$ (0.06). However, if the series are to be cointegrated, the spectral density matrix at zero frequency of X_t is singular and we cannot justify inference based on standard procedures. The generalized squared coherence at zero frequency for the three series see Lobato and Velasco (2000) for the bivariate case is estimated close to 0.95 for both unrestricted or restricted (common d_X) models, calling for some caution in the interpretation of these results.

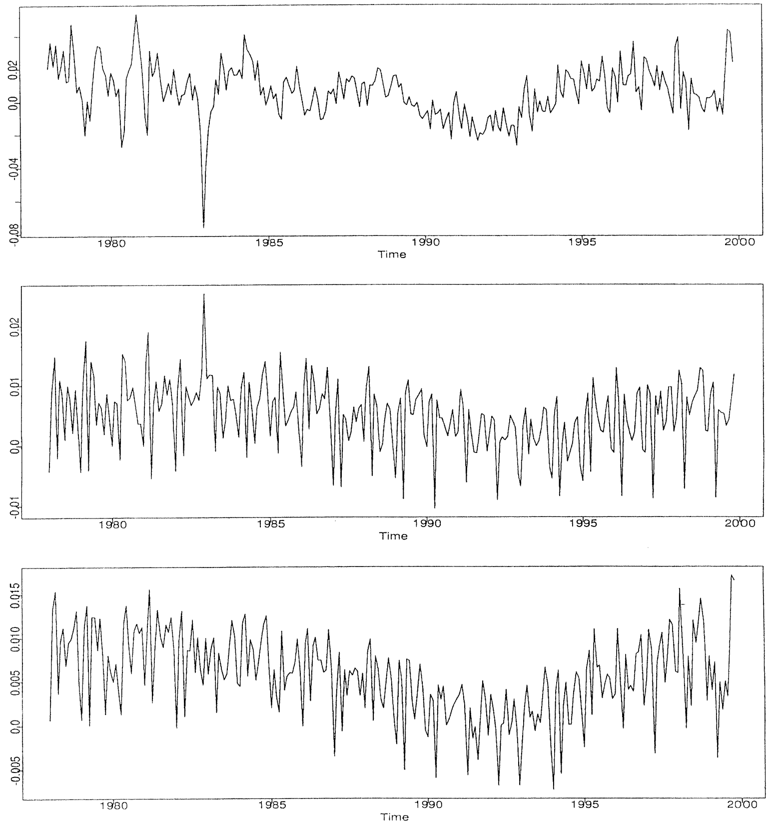

FIGURE 1. ltd_t, $\mathrm{m2}_t$ and $\mathrm{m3}_t$ (increment rates computed as the differences of the logarithm of the original data: LTD, M2 and M3).

In Table III, we report the results of a pairwise investigation of the memory properties of our series (without model), excluding a series ('dependent series') at each time. The main conclusion about the regressors memory is that perhaps ltd_t is stationary, having slightly less memory than $\mathrm{m2}_t$ and $\mathrm{m3}_t$ (the Wald tests have relatively low p values for the null hypotheses $d_{\mathrm{ltd}} = d_{\mathrm{m2}}$ or $d_{\mathrm{ltd}} = d_{\mathrm{m3}}$), but $\mathrm{m2}_t$ and $\mathrm{m3}_t$ are nonstationary and may share the same memory. Therefore, Models 2 and 3 are balanced memory regressions, while Model 1 is unbalanced with more persistent (and possibly cointegrated) regressors. These conclusions agree with the zero frequency squared coherence estimates, which are very low for the two pairs with ltd_t included, but give. 0.77 for ($\mathrm{m2}_t$, $\mathrm{m3}_t$), so $\mathrm{m2}_t$ seems to be the main long term contribution to $\mathrm{m3}_t$ (but perhaps not the only one). Note that, if $\mathrm{m2}_t$ and $\mathrm{m3}_t$

TABLE III

MEMORY ESTIMATES OF LEVEL SERIES

Model	Dependent var.	Regressors memory $\hat{d}_X$	$\hat{d}_{\mathrm{ltd}}$	$\hat{d}_{\mathrm{m2}}$	$\hat{d}_{\mathrm{m3}}$	H_0: common d_X Wald test
1	LTD	0.6072		0.5826	0.6352	0.4760
		(0.079)		(0.088)	(0.088)	[0.490]
2	M2	0.6252	0.4126		0.6901	3.721
		(0.079)	(0.109)		(0.109)	[0.054]
3	M3	0.6261	0.4870	0.6789		1.644
		(0.079)	(0.112)	(0.112)		[0.200]

TABLE IV

MEMORY ESTIMATES OF COINTEGRATED RELATIONSHIPS

		Restricted regression		Unrestricted regression			
Model	Dependent variable	$\hat{d}_u$	$\hat{d}_X$	$\hat{d}_u$	$\hat{d}_{\mathrm{ltd}}$	$\hat{d}_{\mathrm{m2}}$	$\hat{d}_{\mathrm{m3}}$
1	LTD	0.2603	0.6042	0.2785		0.5836	0.6352
		(0.112)	(0.079)	(0.111)		(0.087)	(0.087)
2	M2	0.2423	0.6152	0.2425	0.4126		0.6991
		(0.112)	(0.079)	(0.111)	(0.108)		(0.108)
3	M3	0.3090	0.6161	0.3090	0.4770	0.6789	
		(0.111)	(0.079)	(0.111)	(0.112)	(0.112)	

were *exactly* cointegrated then standard inference is not valid because of the singularity of their spectral density matrix at zero frequency.

For joint estimation, we may choose restricted (common d_X) or unrestricted (different d) Gaussian estimation. If ltd_t is chosen as a regressor, the second option seems more reasonable, while the first can be more sensible if ltd_t is the dependent variable though, for our data, both lead to similar conclusions in this last case (Model 1). For all models and the two procedures, we find in Table IV that the estimated memory of the residuals, $\hat{d}_u$, is always very close to 0.3 with standard error about 0.11, even for the less interpretable Model 1, probably because, on the right hand side of the regression model, some components are cancelling out. Common regressors memory is estimated for all models very close to $\hat{d}_X = 0.61$ (0.08), which is similar to the values of $\hat{d}_{\mathrm{m2}}$ and $\hat{d}_{\mathrm{m3}}$ for Model 1. However for Models 2 and 3, unrestricted estimation shows that ltd_t may have much lower memory than $\mathrm{m2}_t$ or $\mathrm{m3}_t$, while these two series have very similar memory levels. The value of the maximized likelihood indicates that Model 3 is to be preferred, specially for unrestricted estimation, as expected.

The previous evidence leads to the idea that $\mathrm{m2}_t$ and $\mathrm{m3}_t$ might be cointegrated themselves and that their cointegration relationship might be then cointegrated with ltd_t, which would make only a second order long term contribution to $\mathrm{m3}_t$. However, bivariate modelling of ($\mathrm{m2}_t$, $\mathrm{m3}_t$) is not conclusive about this issue since, for $\mathrm{m3}_t$ as dependent variable, we obtain $\hat{d}_u = 0.566$ and $\hat{d}_{\mathrm{m2}} = 0.651$ (0.11), while, for dependent variable $m2_t$, the results are $\hat{d}_u = 0.169$ and $\hat{d}_{m3} = 0.701$ (0.11), with $\hat{d}_u$ in little aggrement with $\hat{d}_{ltd}$.

8. DISCUSSION

We have shown that consistent estimation of the cointegration degree can be based on residuals obtained by, for example, OLS estimation, covering a wide range of models and memory levels, and avoiding short memory missspecifications. If the cointegrating relationship reduces in sufficiently amount the memory of the observed vector, standard semi parametric asymptotics are valid. The Monte Carlo experiment performed has shown that, if the sample size is moderate, the asymptotics approximate reasonably well the finite sample properties of the estimates. However, further empirical evidence with other sample sizes, bandwidth numbers and data generating processes is required for the choice of the preferred estimation strategy.

For general regressor vectors x_t, it is necessary to consider conditions on the minimal and maximal memory components of the vector (generalizing Assumption 1). The main difficulty is the possibility of several cointegration relationships among the regressors series which might cause singularities in the coherence matrix of x_t if several memory parameters d_x are to be estimated. This requires y_t to have the higher level of memory of the series analysed.

Though the methods proposed in this paper provide confidence intervals for the parameters of cointegrated systems, it is not straightforward to develop cointegration Wald tests under the null of no cointegration, since first step estimates are only available when $\beta \neq 0$ and $\delta < d$. Neither Lagrange multiplier tests are easy, since the semi parametric model is unindentified for β when $\delta = d$.

9. PROOFS OF SECTION 3

PROOF OF THEOREM 1. From Robinson (1995b), Theorem 1, we set

$$G(\delta) = G_o \frac{1}{m} \sum_{1}^{m} \lambda_j^{2(\delta - \delta_o)}$$

and $\Theta_1 = \{\delta : \nabla \leqslant \delta < \nabla_2\}$, where $\nabla = \nabla_1$ when $\delta_o < \frac{1}{2} + \nabla_1$ and $\delta_o \geqslant \nabla > \delta_o \quad \frac{1}{2}$ otherwise. Then the Theorem follows from Robinson (1997, Thm 3) if

$$\sum_{n \quad 1}^{m-1} \left(\frac{n}{m}\right)^{2(\nabla - \delta_0)+1} \frac{1}{n^2} \left| \sum_{j \quad 1}^{n} r_j^* \right| \xrightarrow{p} 0 \tag{9}$$

and additionally, for $\psi > 0$ arbitrarily small

$$\log^2 T \sum_{n\ \ 1}^{m-1} \left(\frac{n}{m}\right)^{1-2\psi} \frac{1}{n^2} \left| \sum_{j\ \ 1}^{n} r_j^* \right| \xrightarrow{\text{p}} 0 \tag{10}$$

and

$$\frac{\log^2 T}{m} \sum_{j\ \ 1}^{n} r_j^* \xrightarrow{\text{p}} 0. \tag{11}$$

When $\delta_o \geqslant \frac{1}{2} + \nabla_1$, we also need to show that

$$P(\inf_{\Theta_2} S_{\Delta\tilde{u}}(\delta) \leqslant 0) = o(1)$$

where $\Theta_2 = \{\delta{:}\nabla_1 \leqslant \delta < \nabla\}$ and $S_z^*(\delta) = R_z^*(\delta) \quad R_2^*(\delta_o)$. This follows if

$$\sup_{\Theta_2} |S^*_{\Delta\tilde{u}}(\delta) \quad S^*_{\Delta u}(\delta)| = \sup_{\Theta_2} |R^*_{\Delta\tilde{u}}(\delta) \quad R^*_{\Delta u}(\delta)| = o_p(1)$$

and, from Robinson (1997), it is sufficient to show that

$$\frac{1}{m} \sum_{j\ \ 1}^{m} (a_j \quad 1) r_j^* \xrightarrow{\text{p}} 0 \tag{12}$$

with

$$a_j = \begin{cases} \left(\frac{j}{h}\right)^{2(\nabla-\delta_o)} & 1 \leqslant j \leqslant h \\ \left(\frac{j}{h}\right)^{2(\nabla_1-\delta_o)} & h < j \leqslant m \end{cases}$$

for

$$h = \exp\left(m^{-1} \sum_{1}^{m} \log j \right).$$

Proof of (9). For a generic positive constant $C < \infty$, the left hand side of (9) is bounded by

$$C m^{2(\delta_o-\nabla)-1} \sum_{j\ \ 1}^{m} j^{2(\nabla-\delta_o)} |r_j^*| \qquad \text{for } \nabla < \delta_o \tag{13}$$

and by

$$C \frac{\log m}{m} \sum_{j\ \ 1}^{m} |r_j^*| \qquad \text{for } \nabla = \delta_o. \tag{14}$$

Using (2), the left hand side of (13) is bounded by

$$C m^{2(\delta_o-\nabla)-1} \left\{ 2|\tilde{\beta} \quad \beta| \sum_{j\ \ 1}^{m} j^{2(\nabla-\delta_o)} \frac{|I_{\Delta u \Delta x}(\lambda_j)|}{g_{\Delta u}(\lambda_j)} + (\tilde{\beta} \quad \beta)^2 \sum_{j\ \ 1}^{m} j^{2(\nabla-\delta_o)} \frac{I_{\Delta x \Delta x}(\lambda_j)}{g_{\Delta u}(\lambda_j)} \right\}$$

and taking expectations of the summands, this is

$$O_p\left(m^{2(\delta_o-\nabla)-1}\left\{T^{\delta_o-d}\sum_{j=1}^{m} j^{2(\nabla-\delta_o)}\lambda_j^{\delta_o-d}+T^{2(\delta_o-d)}\sum_{j=1}^{m} j^{2(\nabla-\delta_o)}\lambda_j^{2(\delta_o-d)}\right\}\right)$$
$$=O_p\left(m^{2(\delta_o-\nabla)-1}\left\{\sum_{j=1}^{m} j^{2\nabla-\delta_o-d}+\sum_{j=1}^{m} j^{2(\nabla-d)}\right\}\right)$$
$$=O_p\left(m^{(\delta_o-d)}+m^{2(\delta_o-\nabla)-1}\log m\right)$$
$$=o_p(1)$$

because $d+\delta_o>1$. Now (14) follows similarly.

Proof of (10). The left hand side of (10) is bounded by

$$C\,m^{2\psi-1}\log^2 T\sum_{j=1}^{m} j^{-2\psi|r_j^*|}$$
$$\leqslant Cm^{2\psi-1}\log^2 T\left\{2|\tilde\beta-\beta|\sum_{j=1}^{m} j^{-2\psi}\frac{|I_{\Delta u\Delta x}(\lambda_j)|}{g_{\Delta u}(\lambda_j)}+(\tilde\beta-\beta)^2\sum_{j=1}^{m} j^{-2\psi}\frac{I_{\Delta x\Delta x}(\lambda_j)}{g_{\Delta u}(\lambda_j)}\right\}$$

and this is, by (4), is

$$O_p\left(m^{2\psi-1}\log^2 T\left\{T^{\delta-d}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{\delta_o-d}+T^{2(\delta-d)}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{2(\delta_o-d)}\right\}\right)$$
$$=O_p\left(m^{2\psi-1}\log^2 T\left\{\sum_{j=1}^{m} j^{\delta_o-d-2\psi}+\sum_{j=1}^{m} j^{2(\delta_o-d)-2\psi}\right\}\right)$$
$$=O_p(m^{\delta_o-d}\log^2 T)$$
$$=o_p(1).$$

Proof of (11). As $d+\delta_o>1$, the left hand side of (11) is bounded by

$$\frac{\log^2 T}{m}\sum_{j=1}^{m}|r_j^*|\leqslant\frac{\log^2 T}{m}\left\{2|\tilde\beta-\beta|\sum_{j=1}^{m}\frac{|I_{\Delta u\Delta x}(\lambda_j)|}{g_{\Delta u}(\lambda_j)}+(\tilde\beta-\beta)^2\sum_{j=1}^{m}\frac{I_{\Delta x\Delta x}(\lambda_j)}{g_{\Delta u}(\lambda_j)}\right\}$$
$$=O_p\left(\frac{\log^2 T}{m}\left\{\sum_{j=1}^{m} j^{\delta_o-d}+\sum_{j=1}^{m} j^{2(\delta_o-d)}\right\}\right)$$
$$=O_p(m^{\delta_o-d}\log^2 T)$$
$$=o_p(1).$$

Proof of (12). Using (11), the left hand side of (12) is bounded by

$$\frac{1}{m}\sum_{j=1}^{m} a_j|r_j^*|+o_p(1)$$

and as in Robinson (1995b), we can use that $h \sim m/e$ as $T \to \infty$, and that $a_j = O(1)$, uniformly for $j > h$, so the first term on the right hand side is bounded by

$$Cm^{2(\delta_o-\nabla)-1}|\tilde{\beta} \quad \beta| \sum_{j \quad 1}^{m} j^{2(\nabla-\delta_o)} \lambda_j^{\delta_o-d} + Cm^{2(\delta_o-\nabla)-1}|\tilde{\beta} \quad \beta|^2 \sum_{j \quad 1}^{m} j^{2(\nabla-\delta_o)} \lambda_j^{\delta_o-d}$$

and, as $d+\delta_o > 1$ and $\delta_o < d$, using the same argument of the proof of (9), this is

$$O_p\left(m^{2(\delta_o-\nabla)-1} \sum_{j \quad 1}^{m} j^{2\nabla-\delta_o-d} + m^{2(\delta_o-\nabla)-1} \sum_{j \quad 1}^{m} j^{2(\nabla-d)} \right) = o_p(1).$$

Proof of Theorem 2. Since under the conditions of the Theorem, $\tilde{\delta}^*$ is consistent (compare with Theorem 1 without the $\log^2 T$ terms), from Robinson (1995b, Thm 2), we need to show in first place that

$$\sup_{\delta \in \Theta_1 \cap N_\psi} \left| \frac{\tilde{G}^*_{\Delta\tilde{u}}(\delta) \quad G(\delta)}{G(\delta)} \right| = o_p(\log^{-6} m)$$

where N_ψ is defined as in Robinson (1995b, p. 1634), and this is implied by,

$$\sup_{\delta \in \Theta_1 \cap N_\psi} \left| \frac{\tilde{G}^*_{\Delta\tilde{u}}(\delta) \quad \tilde{G}^*_{\Delta u}(\delta)}{G(\delta)} \right| = o_p(\log^{-6} m).$$

This, in turn, follows if

$$\log^6 m \sum_{n \quad 1}^{m-1} \left(\frac{n}{m}\right)^{1-2\psi} \frac{1}{n^2} \left| \sum_{j \quad 1}^{n} r_j^* \right| \xrightarrow{\text{p}} 0 \qquad (15)$$

from equation (4.7) in Robinson (1995b). Then, we also need that, for $k = 0, 1, 2,$

$$|\tilde{F}^*_{k,\Delta u}(\delta_o) \quad \tilde{F}^*_{k,\Delta\tilde{u}}(\delta_o)| \xrightarrow{\text{p}} 0 \qquad (16)$$

where

$$\tilde{F}^*_{k,z}(\delta) = \frac{1}{m} \sum_{j \quad 1}^{m} (\log j)^k \lambda_j^{(2\delta-1)} I_{zz}(\lambda_j)$$

and that

$$m^{1/2} \left| \frac{\mathrm{d}R^*_{\Delta u}(\delta_o)}{\mathrm{d}\delta} \quad \frac{\mathrm{d}R^*_{\Delta\tilde{u}}(\delta_o)}{\mathrm{d}\delta} \right| \xrightarrow{\text{p}} 0 \qquad (17)$$

where

$$\frac{\mathrm{d}R_z^*(\delta)}{\mathrm{d}\delta} = 2\frac{\tilde{H}_z^*(\delta)}{\tilde{G}_z^*(\delta)}$$

$$\tilde{H}_z^*(\delta) = \frac{1}{m}\sum_1^m v_j \lambda_j^{2(\delta-1)} I_{zz}(\lambda_j)$$

$$v_j = \log j \quad \frac{1}{m}\sum_1^m \log j = O(\log m).$$

Proof of (15). It holds by the proof of (10) and (11) ignoring the $\log^2 T$ terms.

Proof of (16). The left hand side of (16) is bounded by

$$\left|\frac{G_u}{m}\sum_{j\ 1}^m (\log j)^k r_j^*\right| \leqslant C\frac{\log^k m}{m}\sum_{j\ 1}^m |r_j^*|$$
$$= o_p(1)$$

by the proof of (11).

Proof of (17). The left hand side of (17) is bounded by

$$2m^{1/2}\left|\frac{\tilde{H}_{\Delta u}^*(\delta_o)}{\tilde{G}_{\Delta u}^*(\delta_o)} \quad \frac{\tilde{H}_{\Delta\tilde{u}}^*(\delta_o)}{\tilde{G}_{\Delta\tilde{u}}^*(\delta_o)}\right| \leqslant 2m^{1/2}\left|\tilde{G}_{\Delta u}^*(\delta_o)\right|^{-1}\left|\tilde{H}_{\Delta u}^*(\delta_o) \quad \tilde{H}_{\Delta\tilde{u}}^*(\delta_o)\right|$$
$$+ 2m^{1/2}\left|\frac{\tilde{H}_{\Delta\tilde{u}}^*(\delta_o)}{\tilde{G}_{\Delta u}^*(\delta_o)\tilde{G}_{\Delta\tilde{u}}^*(\delta_o)}\right|\left|\tilde{G}_{\Delta u}^*(\delta_o) \quad \tilde{G}_{\Delta\tilde{u}(\delta_o)}^*\right|.$$

First, by Robinson (1995b),

$$\tilde{G}_{\Delta u}^*(\delta_o) = G_o + o_p(1)$$

and, by (5),

$$m^{1/2}\left|\tilde{H}_{\Delta u}^*(\delta_o) \quad \tilde{H}_{\Delta\tilde{u}}^*(\delta_o)\right| \leqslant m^{-1/2}\sum_1^m |v_j||r_j^*| = O_p\left(m^{-1/2}\log m\sum_1^m j^{\delta_o - d}\right)$$
$$= o_p(1).$$

Similarly

$$m^{1/2}\left|\tilde{G}_{\Delta u}^*(\delta_o) \quad \tilde{G}_{\Delta\tilde{u}}^*(\delta_o)\right| = o_p(1)$$

by the same argument of the proof of (15), and

$$\tilde{H}_{\Delta\tilde{u}}^*(\delta_o) = O_p(1)$$

by Robinson (1995b, p. 1644) and the theorem follows. QED.

PROOF OF THEOREM 3. From Theorem 1, the result follows if, $\delta_o - 0.5 < \nabla < \delta_o$,

$$\sum_{n\ \ 1}^{m-1} \left(\frac{n}{m}\right)^{2(\nabla-\delta_o)+1} \frac{1}{n^2} \left| \sum_{j\ \ 1}^{n} r_j \right| \xrightarrow{p} 0 \tag{18}$$

for $\psi > 0$ arbitararily small,

$$\log^2 T \sum_{n\ \ 1}^{m-1} \left(\frac{n}{m}\right)^{1-2\psi} \frac{1}{n^2} \left| \sum_{j\ \ 1}^{n} r_j \right| \xrightarrow{p} 0 \tag{19}$$

and

$$\frac{\log^2 T}{m} \sum_{j\ \ 1}^{n} r_j \xrightarrow{p} 0 \tag{20}$$

and finally,

$$\frac{1}{m} \sum_{n\ \ 1}^{m} (a_j \quad 1) r_j \xrightarrow{p} 0. \tag{21}$$

PROOF OF (18). The left hand side of (18) is bounded by

$$Cm^{2(\delta_o-\nabla)-1} \sum_{j\ \ 1}^{m} j^{2(\nabla-\delta_o)} |r_j| \qquad \text{for } \nabla < \delta_o \tag{22}$$

and by

$$c \frac{\log m}{m} \sum_{j\ \ 1}^{m} |r_j| \qquad \text{for } \nabla = \delta_o. \tag{23}$$

Using (3), (22) is bounded by

$$Cm^{2(\delta_o-\nabla)-1} \left\{ 2|\tilde{\beta} \quad \beta| \sum_{j\ \ 1}^{m} j^{2(\nabla-\delta_o)} \frac{|I_{ux}(\lambda_j)|}{g_u(\lambda_j)} + (\tilde{\beta} \quad \beta)^2 \sum_{j\ \ 1}^{m} j^{2(\nabla-\delta_o)} \frac{I_{xx}(\lambda_j)}{g_u(\lambda_j)} \right\} \tag{24}$$

and, if $d + \delta_o > 1$ (Case I), (24) is

$$O_p\left(m^{2(\delta_o-\nabla)-1} \left\{ \sum_{j\ \ 1}^{m} j^{2\nabla-\delta_o-d} + \sum_{j\ \ 1}^{m} j^{2(\nabla-d)} \right\} \right) = O_p\left(m^{\delta_o-d} + m^{2(\delta_o-\nabla)-1} \log m \right)$$
$$= o_p(1)$$

when $\delta_o < d < 1$, and is

$$O_p\left(m^{2(\delta_o-\nabla)-1}\left\{\sum_{j\ 1}^{m} j^{2\nabla-\delta_o-1}+\sum_{j\ 1}^{m} j^{2(\nabla-d)}\right\}\right)=O_p\left(m^{\delta_o-1}+m^{2(\delta_o-\nabla)-1}\log m\right)$$
$$=o_p(1)$$

when $d \geqslant 1$, noting that $\delta_o < 1$.

If $d+\delta_o < 1$ (Case III), then $d < 1$, so (24) is

$$O_p\left(m^{2(\delta_o-\nabla)-1}\left\{T^{1-\delta_o-d}\sum_{j\ 1}^{m} j^{2\nabla-\delta_o-1}+T^{2(1-\delta_o-d)}\sum_{j\ 1}^{m} j^{2(\nabla-d)}\right\}\right)$$
$$=O_p\begin{pmatrix} T^{1-\delta_o-d}[m^{\delta_o-d}+m^{2(\delta_o-\nabla)-1}\log m\mathbb{I}\{2\nabla\quad d\quad \delta_o \leqslant \quad 1\}] \\ +T^{2(1-\delta_o-d)}[m^{2(\delta_o-d)}+m^{2(\delta_o-\nabla)-1}\log m\mathbb{I}\{2(\nabla\quad d)\leqslant\quad 1\}]\end{pmatrix}$$
$$=o_p(1)+O_p\left(T^{2(1-\delta_o-d)}+m^{2(\delta_o-\nabla)-1}\log m\mathbb{I}\{2(\nabla\quad d)\leqslant\quad 1\}\right)$$

because of (6). The choice of ∇ is restricted to satisfy $2(\delta_o-\nabla)-1 < -1/(2e)$, by the proof of Theorem 1 in Robinson (1995b). Then we can choose ∇ large enough such that for some $\epsilon > 0$, and using (6),

$$T^{2(1-\delta_o-d)}m^{2(\delta_o-\nabla)-1}\log m\mathbb{I}\{2(\nabla\quad d)\leqslant\quad 1\}=O(T^{2(1-\delta_o-d)}m^{\epsilon-1/(2e)})=o(1).$$

In case II, (24) is

$$O_p(m^{\delta_o-d}\log T+m^{2(\delta_o-\nabla)-1}\log m\log^2 T)=o_p(1)$$

with (6) and $d < 1$.

Now we can bound (23) by

$$Cm^{-1}\log m\left\{2|\tilde{\beta}\quad\beta|\sum_{j\ 1}^{m}\frac{|I_{ux}(\lambda_j)|}{g_u(\lambda_j)}+(\tilde{\beta}\quad\beta)^2\sum_{j\ 1}^{m}\frac{I_{xx}(\lambda_j)}{g_u(\lambda_j)}\right\} \tag{25}$$

and, if $d+\delta_o > 1$ (Case I), this is

$$O_p\left(m^{-1}\log m\left\{\sum_{j\ 1}^{m} j^{\delta_o-d}+\sum_{j\ 1}^{m} j^{2(\nabla-d)}\right\}\right)=O_p(m^{\delta_o-d}\log m)$$
$$=o_p(1)$$

when $\delta_o < d < 1$, and is

$$O_p\left(m^{-1}\log m\left\{\sum_{j\ 1}^{m} j^{\delta_o-1}+\sum_{j\ 1}^{m} j^{2(\delta_o-1)}\right\}\right)=O_p(m^{\delta_o-1}\log^2 m)$$
$$=o_p(1)$$

when $d \geqslant 1$ (with $\delta_o < 1$). In Case II, this is

$$O_p(m^{-1}\log m \log^2 T) = o_p(1)$$

If $d+\delta_o < 1$ (Case III), then $d<1$, so (25) is

$$\begin{aligned} &O_p\left(m^{-1}\log m\left\{T^{1-\delta_o-d}\sum_{j=1}^{m} j^{\delta_o-d} + T^{2(1-\delta_o-d)}\sum_{j=1}^{m} j^{2(\delta_0-d)}\right\}\right)\\ &\quad = O_p(T^{1-\delta_o-d}m^{\delta_o-d}\log m + T^{2(1-\delta_o-d)}[m^{-1}\log m + m^{2(\delta_o-d)}]\log m)\\ &\quad = o_p(1) \end{aligned}$$

because of (6) and because

$$T^{2(1-\delta_o-d)}m^{-1}\log m \to 0.$$

Proof of (19). The left hand side of (19) is bounded by

$$\begin{aligned} &Cm^{2\psi-1}\log^2 T\sum_{j=1}^{m} j^{-2\psi}|r_j|\\ &\leqslant m^{2\psi-1}\log^2 T\left\{2|\tilde\beta-\beta|\sum_{j=1}^{m} j^{-2\psi}\frac{|I_{ux}(\lambda_j)|}{g_u(\lambda_j)} + (\tilde\beta-\beta)^2\sum_{j=1}^{m} j^{-2\psi}\frac{I_{xx}(\lambda_j)}{g_u(\lambda_j)}\right\} \end{aligned}$$

and, using (6), if $d+\delta_o > 1$, $d<1$, this is

$$\begin{aligned} &O_p\left(m^{2\psi-1}\log^2 T\left\{T^{\delta_o-d}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{\delta_o-d} + T^{2(\delta_o-d)}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{2(\delta_o-d)}\right\}\right)\\ &\quad = O_p(\log^2 T[m^{\delta_o-d} + m^{2\psi-1}\log m\,\mathbb{I}_{\{2(\delta_o-d)-2\psi\leqslant -1\}}])\\ &\quad = o_p(1) \end{aligned}$$

and, for $d \geqslant 1$, this is

$$\begin{aligned} &O_p\left(m^{2\psi-1}\log^2 T\left\{T^{\delta_o-d}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{\delta_o-d}j^{d-1} + T^{2(\delta_o-d)}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{2(\delta_o-d)}j^{2(d-1)}\right\}\right)\\ &\quad = O_p(\log^2 T[m^{\delta_o-1} + m^{2\psi-1}\log m])\\ &\quad = o_p(1). \end{aligned}$$

In the Case II, this is

$$O_p(\log^4 T[m^{\delta_o-d} + m^{2\psi-1}\log m]) = o_p(1)$$

with (6). Finally, if $d+\delta_o < 1$ (Case III), this is

$$O_p\left(m^{2\psi-1}\log^2 T\left\{T^{1-2d}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{\delta_o-d}+T^{2-4d}\sum_{j=1}^{m} j^{-2\psi}\lambda_j^{2(\delta_o-d)}\right\}\right)$$

$$=O_p\left(\left\{\begin{array}{c} T^{1-\delta_o-d}m^{\delta_o-d}+T^{1-\delta_o-d}m^{2\psi-1}\log m\mathbb{I}_{\{\delta_o-d-2\psi\leqslant-1\}} \\ +T^{2(1-\delta_o-d)}m^{2(\delta_o-d)}+T^{2(1-\delta_o-d)}m^{2\psi-1}\log m\mathbb{I}_{\{2(\delta_o-d-\psi)\leqslant-1\}}\end{array}\right\}\log^2 T\right)$$

and this is $o_p(1)$ by (6) again.

Proof of (20). For $d+\delta_o>1$ and using (6), the left hand side of (20) is bounded by

$$\begin{aligned}\frac{\log^2 T}{m}\sum_{j=1}^{m}|r_j| &= O_p\left(\frac{\log^2 T}{m}\left\{\sum_{j=1}^{m} j^{\delta_o-d}+\sum_{j=1}^{m} j^{2(\delta_o-d)}\right\}\right)\\ &= O_p([m^{\delta_o-d}+m^{-1}\log m]\log^2 T)\\ &= o_p(1)\end{aligned}$$

when and if $d\geqslant 1$, a bound is

$$\begin{aligned}\frac{\log^2 T}{m}\sum_{j=1}^{m}|r_j| &= O_p\left(\frac{\log^2 T}{m}\left\{\sum_{j=1}^{m} j^{\delta_o-1}+\sum_{j=1}^{m} j^{2(\delta_o-1)}\right\}\right)\\ &= O_p([m^{\delta_o-1}+m^{-1}\log m]\log^2 T)\\ &= o_p(1).\end{aligned}$$

In Case II, (20) is

$$O_p([m^{\delta_o-d}\log T+m^{-1}\log m\log^2 T]\log^2 T)=o_p(1)$$

by (6), and when $d+\delta_o<1$, (20) is

$$O_p\left(\frac{\log^2 T}{m}\left\{T^{1-\delta_o-d}\sum_{j=1}^{m} j^{\delta_o-d}+T^{2(1-\delta_o-d)}\sum_{j=1}^{m} j^{2(\delta_o-d)}\right\}\right)$$

$$=O_p\left(\left\{T^{1-\delta_o-d}m^{\delta_o-d}+T^{2(1-\delta_o-d)}[m^{2(\delta_o-d)}+m^{-1}\log m\mathbb{I}_{\{2(\delta_o-d)\leqslant-1\}}]\right\}\log^2 T\right)$$

$$=o_p(1)$$

because of (6).

Proof of (21). Proceeding as before, the left hand side of (21) is bounded by

$$\frac{1}{m}\sum_{j=1}^{m} a_j|r_j|+o_p(1)=O_p\left(m^{2(\delta_o-\nabla)-1}\sum_{j=1}^{m} j^{2(\nabla-\delta_o)}|r_j|\right)+o_p(1)$$

and, if $d+\delta_o>1$, the first term on the right hand side is

$$O_p\left(m^{2(\delta_o-\nabla)-1}\sum_{j\;\;1}^{m}j^{2\nabla-\delta_0-d}+m^{2(\delta_o-\nabla)-1}\sum_{j\;\;1}^{m}j^{2(\nabla-d)}\right)=O_p(m^{\delta_o-d}+m^{2(\delta_o-\nabla)-1}\log m)$$
$$=o_p(1)$$

when $d < 1$, and if $d \geqslant 1$, is

$$O_p\left(m^{2(\delta_o-\nabla)-1}\sum_{j\;\;1}^{m}j^{2\nabla-\delta_o-1)}+m^{2(\delta_o-\nabla)-1}\sum_{j\;\;1}^{m}j^{2(\nabla-1)}\right)=O_p(m^{\delta_o-1}+m^{2(\delta_o-\nabla)-1}\log m)$$
$$=o_p(1).$$

A similar bound follows in Case II using (6) and if $d+\delta_o<1$, this is

$$O_p\left(m^{2(\delta_o-\nabla)-1}T^{1-\delta_o-d}\sum_{j\;\;1}^{m}j^{2\nabla-\delta_o-d}+m^{2(\delta_o-\nabla)-1}T^{2(1-\delta_o-d)}\sum_{j\;\;1}^{m}j^{2(\nabla-d)}\right)$$
$$=O_p\left(T^{1-\delta_o-d}m^{\delta_o-d}+T^{2(1-\delta_o-d)}[m^{2(\delta_o-d)}+m^{2(\delta_o-\nabla)-1}\log m\mathbb{I}_{\{2(\nabla-d)\leqslant-1\}}]\right)$$

which is $o_p(1)$ by (6) proceeding as in the proof of (18). QED.

PROOF OF THEOREM 4. As in the proof of Theorem 2, we need to show that

$$\log^6 m\sum_{n\;\;1}^{m-1}\left(\frac{n}{m}\right)^{1-2\psi}\frac{1}{n^2}\left|\sum_{j\;\;1}^{n}r_j\right|\overset{p}{\to}0. \tag{26}$$

Then we also need to show that, for $k=0,\ 1,\ 2$,

$$\left|\tilde{F}_{k,\Delta u}(\delta_o)\quad\tilde{F}_{k,\Delta\tilde{u}}(\delta_o)\right|\overset{p}{\to}0 \tag{27}$$

where

$$\tilde{F}_{k,z}(\delta)=\frac{1}{m}\sum_{j\;\;1}^{m}(\log j)^k\lambda_j^{2\delta}I_{zz}(\lambda_j)$$

and that

$$m^{1/2}\left|\frac{\mathrm{d}R_{\Delta u}(\delta_o)}{\mathrm{d}\delta}\quad\frac{\mathrm{d}R_{\Delta\tilde{u}}(\delta_o)}{\mathrm{d}\delta}\right|\overset{p}{\to}0 \tag{28}$$

where

$$\frac{\mathrm{d}R_z(\delta)}{\mathrm{d}\delta}=2\frac{\tilde{H}_z(\delta)}{\tilde{G}_z(\delta)}$$

and

$$\tilde{H}_z(\delta)=\frac{1}{m}\sum_{1}^{m}v_j\lambda_j^{2\delta}I_{zz}(\lambda_j)$$

Proof of (26) First, (26) holds by the proof of (19) and (20), with the same choice of m given by (7).

Proof of (27) The left hand side of (27) is bounded by, for $k = 0$, 1, 2,

$$\left|\frac{G}{m}\sum_{j\ \ 1}^{m}(\log j)^k r_j\right| \leqslant C\frac{\log^k m}{m}\sum_{j\ \ 1}^{m}|r_j|$$
$$= o_p(1)$$

by (7) and the proof of (20).

Proof of (28) The left hand side of (28) is bounded by

$$2m^{1/2}\left|\frac{\tilde{H}_u(\delta_o)}{\tilde{G}_u(\delta_o)} \quad \frac{\tilde{H}_{\tilde{u}}(\delta_o)}{\tilde{G}_{\tilde{u}}(\delta_o)}\right| \leqslant 2m^{1/2}|\tilde{G}_u(\delta_o)|^{-1}|\tilde{H}_u(\delta_o) \quad \tilde{H}_{\tilde{u}}(\delta_o)|$$
$$+ 2m^{1/2}\left|\frac{\tilde{H}_{\tilde{u}}(\delta_o)}{\tilde{G}_u(\delta_o)\tilde{G}_{\tilde{u}}(\delta_o)}\right||\tilde{G}_u(\delta_o) \quad \tilde{G}_{\tilde{u}}(\delta_o)|.$$

First,

$$\tilde{G}_u(\delta_o) = G + o_p(1)$$

from Robinson (1995b) and Velasco (1999b), and

$$m^{1/2}|\tilde{H}_u(\delta_o) \quad \tilde{H}_{\tilde{u}}(\delta_o)| \leqslant m^{-1/2}\sum_{1}^{m}|v_j||r_j|$$
$$= O_p\left(m^{-1/2}\sum_{1}^{m}|r_j|\log m\right).$$

Using (7) when $d + \delta_o \geqslant 1$, this is

$$O_p(m^{-1/2}[\log m + m^{\delta_o - d + 1}]\log m) = o_p(1) \qquad d < 1, d + \delta_o > 1$$
$$O_p(m^{-1/2}[\log m \log^2 T + m^{\delta_o - d + 1}\log T]\log m) = o_p(1) \qquad d < 1, d + \delta_o = 1$$
$$O_p(m^{-1/2}[\log m + m^{\delta_o}]\log m) = o_p(1) \qquad d \geqslant 1$$

while, if $d + \delta_o \leqslant 1$, is

$$O_p\left(m^{-1/2}\left[T^{1-\delta_o-d}m^{\delta_o-d+1} + T^{2(1-\delta_o-d)}\left\{m^{2(\delta_o-d)+1} + \log m\mathbb{I}_{\{2(\delta_o-d)\leqslant -1\}}\right\}\right]\log m\right)$$
$$= o_p(1)$$

and the theorem follows from

$$\tilde{H}_u(\delta_o) = O_p(1) \qquad \text{and} \qquad m^{1/2}|\tilde{G}_u(\delta_o) \quad \tilde{G}_{\tilde{u}}(\delta_o)| = o_p(1).$$

ACKNOWLEDGEMENTS

I am grateful to the Associate Editor and a referee for helpful comments and to U. Hassler, M. Salvador and participants at the Econometrics Society European Meeting, Santiago de Compostela, 1999, the Workshop on Nonparametric, Semiparametric and Resampling Methods, Universidad Carlos III de Madrid, 2000, the European meeting of statisticians, Funchal, 2001, and the X Seminario sobre Validación y Especificación de Modelos Econométricos, Zaragoza, 2001. Research funded by Spanish Dirección General de Enseñanza Superior, Ref. no. BEC2001 1270.

REFERENCES

Davidson, J. and de Jong, R. M. (2000) The functional central limit theorem and weak convergence to stochastic integrals II. *Econometric Theory* 16, 643 66

Dolado, J. and Marmol, P. (1998) Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes. Preprint Universidad Carlos III de Madrid.

Dueker, M. and Startz, R. (1998) Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. *Review of Economics and Statistics* 80, 420 6

Engle, R. and Granger, C. (1987) Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55, 251 76.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221 38.

Giraitis, L. and Surgailis, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Probability Theory and Related Fields* 86, 87 104.

———, Robinson, P. M. and Samarov, A (1997) Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *Journal of Time Series Analysis* 18, 49 60.

Granger, C. W. J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121 30.

Hassler, U., Marmol, F. and Velasco, C. (2002) Fractional cointegrating regressions and the residual periodogram. *Journal of Econometrics* (forth coming).

Hosoya, Y. (1997) A limit theory for long-range dependence and statistical inference of related model. *Annals of Statistics* 25, 105 37.

Hurvich, C. M. and Chen, W. W. (2000) An efficient taper for potentially overdifferenced long memory time series. *Journal of Time Series Analysis* 21, 155 80.

——— and Ray, B. K. (1995) Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *Journal of time Series Analysis* 16, 17 42.

Künsch, H. R. (1987) Statistical aspects of self-similar processes. *Proceedings 1st World Congress of the Bernoulli Society*, Utrecht: VNU Science Press, pp. 67 74.

Lobato, I. N. (1997) Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis* 18, 137 55.

——— (1999) A semiparametric two step estimator in a multivariate long memory model. *Journal of Econometrics* 90, 129 53.

——— and Velasco, C. (2000) Long memory in stock market trading volume. *Journal of Business and Economic Statistics* 18, 410 27.

Robinson, P. M. (1994a) Semiparametric analysis of long-memory time series. *Annals of Statistic* 22, 515 39.

——— (1994b) Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields* 99, 443 73.

——— (1995a) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048 72.

——— (1995b) Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23, 1630–1661
——— (1997) Large sample inference for nonparametric regression with dependent errors. *Annals of Statistics* 25, 2054–83.
——— and HENRY, M. (1999) Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels. *Econometric Theory* 15, 299–336.
——— and MARINUCCI, D. (2001) Narrow band analysis of nonstationary processes. *Annals of Statistics* 29, 947–86.
SMITH, R. L. and CHEN, F.-L. (1997) Regression in long-memory time series. In *Athens Conference on Applied Probability and Time Series* Vol. II (eds P. M. Robinson and M. Rosenblatt). New York: Springer, pp. 378–91.
TANAKA, K. (1999) The nonstationary fractional unit root. *Econometric Theory* 15, 549–82.
VELASCO, C. (1999a) Non-stationary log-periodogram regression. *Journal of Econometrics* 91, 325–71.
——— (1999b) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.
——— (2002) Nonparametric frequency domain analysis of nonstationary multivariate time series. *Journal of Statistical Planning and Inference* (forthcoming).