CONSISTENT TESTING OF COINTEGRATING RELATIONSHIPS

BY FRANCESC MARMOL AND CARLOS VELASCO[1]

In this paper we investigate methods for testing the existence of a cointegration relationship among the components of a nonstationary fractionally integrated (NFI) vector time series. Our framework generalizes previous studies restricted to unit root integrated processes and permits simultaneous analysis of spurious and cointegrated NFI vectors. We propose a modified F-statistic, based on a particular studentization, which converges weakly under both hypotheses, despite the fact that OLS estimates are only consistent under cointegration. This statistic leads to a Wald-type test of cointegration when combined with a narrow band GLS-type estimate. Our semiparametric methodology allows consistent testing of the spurious regression hypothesis against the alternative of fractional cointegration without prior knowledge on the memory of the original series, their short run properties, the cointegrating vector, or the degree of cointegration. This semiparametric aspect of the modelization does not lead to an asymptotic loss of power, permitting the Wald statistic to diverge faster under the alternative of cointegration than when testing for a hypothesized cointegration vector. In our simulations we show that the method has comparable power to customary procedures under the unit root cointegration setup, and maintains good properties in a general framework where other methods may fail. We illustrate our method testing the cointegration hypothesis of nominal GNP and simple-sum ($M1$, $M2$, $M3$) monetary aggregates.

KEYWORDS: Cointegration, spurious regression, long memory, fractional processes, narrow-band frequency analysis, Wald test, semiparametric inference.

1. INTRODUCTION

IT IS A WELL-KNOWN EMPIRICAL FACT that many economic time series are typically nonstationary. This nonstationarity can induce (seemingly) significant correlations between the levels of these time series, despite the fact that, from a theoretical viewpoint, there is no justification for any relationship between them, giving rise to what is known in the econometric literature as the *spurious* problem. Using Monte Carlo simulations, Granger and Newbold's (1974) classic study showed that this phenomenon occurs when independent random walks are regressed on one another. In 1986, Phillips developed an asymptotic theory for regressions between independent integrated of order one I(1) processes showing the invalidity of standard ordinary least squares (OLS) inference. In particular, the customary F-statistic has divergent asymptotic behavior in such regressions so that there are no asymptotically correct critical

[1]The first author's research was supported by DGICYT Grant SEC2001-0890. The second author's research was supported by Spanish Dirección General de Enseñanza Superior, Ref. No. BEC2001-1270. We are grateful to P. M. Robinson, D. Marinucci, and J. Gonzalo and the participants at the Workshop on Nonparametric, Semiparametric and Resampling Methods, Universidad Carlos III de Madrid, 2000, ESEM 2002, Workshop on Econometric Time Series Analysis, Johannes Kepler Universität, 2003, seminars at the Universities of Navarra and Santiago, and the coeditor and three referees for valuable comments.

values for this conventional significance test. Extending Phillips' (1986) approach, it is now well understood that the spurious regression phenomenon also occurs for a wider class of time series, including I(2) processes (Haldrup (1994)), stochastic unit root processes (Granger and Swanson (1997)), nonstationary fractionally integrated processes (Marmol (1998)), and some particular types of stationary processes (Tsay and Chung (2000), Granger, Hyung, and Jeon (2001)).

Standard OLS inference is not useful in spurious systems because regression estimates are not consistent for any fixed projection vector and the usual standardization does not take full account of the residual autocorrelation. Therefore, spuriously significant test statistics appear, despite the fact that the degree of serial dependence of the OLS residuals is similar to that of the series levels.

In this paper we explore alternative forms of studentization in the frequency domain of least squares estimates in the presence of nonstationary fractionally integrated (in short NFI) processes. We say that z_t is a $(p \times 1)$ vector of NFI processes with filter matrix $\Delta(L) = \mathrm{diag}\{(1-L)^{-d_1}, \dots, (1-L)^{-d_p}\}$, $d_i > \frac{1}{2}$, $i = 1, \dots, p$, if $\Delta(L)^{-1} z_t$ is a short-memory I(0) process, where L is the lag operator. This model has been used in applied work where flexible characterization of low frequency dynamics is important, since fractionally integrated processes accommodate unit root-type persistence as well as long range dependence and mean reversion. We propose an alternative to the usual F-statistic with an autocorrelation-robust frequency domain studentization. This provides an adequate automatic normalization such that the new Wald or adjusted F-statistic has a well-defined (nonstandard) limiting distribution, which in the case of common memory d does not depend on additional parameters.

The natural alternative to a spurious relationship is cointegration. Following Marinucci and Robinson (2001) and Robinson and Yajima (2002), we say that a $(p \times 1)$ vector z_t of NFI processes with filter matrix $\Delta(L)$ is cointegrated of orders $(d_1, d_2, \dots, d_p; \delta)$ with $d_1 \geq d_2 \geq d_3 \geq \cdots \geq d_{p-1} \geq d_p$ if at least $d_1 = d_2$, and there exists a $(p \times 1)$ vector h with nonnull first element such that the linear combination $h' z_t$ is a fractionally integrated process of order $0 \leq \delta < d_1$. Otherwise, i.e., if $\delta = d_1$ for all h, we say that the components of z_t are spuriously related. Cointegration is commonly thought of as a stationary relation between nonstationary variables. When $0 < \delta < \frac{1}{2}$, the linear combination $h' z_t$ is a stationary long-memory process and the slow decay of the effect of shocks allows for slow adjustments to equilibrium. The nonstationary case ($\delta > \frac{1}{2}$) is also worth entertaining. For instance, when $\frac{1}{2} < \delta < 1$, the error term will exhibit transitory memory, implying that a shock will have no permanent effect on $h' z_t$. Also, from a methodological point of view it is reasonable to consider a model that permits values of δ arbitrarily close to the memory parameters of the observables z_t.

Fractional cointegration has become an important and relevant topic in empirical analysis in recent years. A partial list includes Cheung and Lai (1993), Masih and Masih (1995), and Soofi (1998) who test the purchasing power parity hypothesis, Baillie and Bollerslev (1994) and Hassler, Marmol, and Velasco (2003) who investigate the memory of exchange rates, and Booth and Tse (1995), Masih and Masih (1998), and Dittmann (2000) who are concerned with the dynamic of interest rate future markets, exchange rate dynamics, and stock market prices, respectively. All of them find evidence of fractional cointegration in their data.

Following the initial suggestion of Engle and Granger (1987), when the series of interest are I(1), testing for cointegration in a single-equation framework is usually conducted by means of residual-based tests (cf. Phillips and Ouliaris (1990)). Residual-based tests rely on the residuals calculated from regressions among the levels of the relevant time series. They are designed to test the null of no cointegration by testing the null that there is a unit root in the residuals against the alternative that the root is less than unity, i.e., that the regression errors are I(0). Inference on I(1)/I(0) cointegrated systems in the framework of a fully parametric error correction mechanism (ECM) representation has been developed by Johansen (1988, 1991).

However, to the best of our knowledge, there is no test available in the literature that is powerful against fractionally cointegrated alternatives irrespective of the (generally unknown) memory properties of trending data. In this line, we investigate the properties of our Wald statistic under cointegration, using a triangular representation of NFI processes that encompasses both hypotheses by means of a single *correlation* parameter, ρ. Despite the well-known consistency of OLS coefficients under cointegration, our Wald statistic still converges to a well-defined distribution. This observation is the basis for a consistent test of cointegration comparing the OLS regression estimates with any other estimate of the long-run projection vector that shows opposite consistent properties under the competing hypotheses. This methodology, combined with information on the projection vector obtained through a semiparametric generalized least squares (GLS)-type estimate, permits consistent testing of the null hypothesis of no cointegration against the alternative of fractional cointegration, independently of knowledge of the memory of the original vector and the degree of cointegration. This is achieved using semiparametric estimates of the memory parameters of the components of the NFI vector and of the OLS residuals.

In this sense, our methodology does not require the specification of the short-run dynamics of the underlying processes because of the use of semiparametric estimates of memory parameters and long-run covariance matrix. This semiparametric aspect of modeling does not lead to an asymptotic loss of power, and permits the Wald statistics to diverge faster under the alternative of cointegration than when testing for a hypothesized cointegration vector. Thus we take advantage of the simultaneous (super) consistency of OLS estimates and of the inconsistency of GLS estimates under cointegration. Our test avoids in this way problems found in approaches based directly on memory estimates (Marinucci and Robinson (2001), Hassler, Marmol, and Velasco

(2003), Velasco (2003)) or on estimates of ρ, whose distribution under the null may be unknown or may have only semiparametric rates of divergence under the alternative. By means of simulations we show that for series of moderate size our method has comparable power to customary procedures under the unit root cointegration setup (which heavily exploits such knowledge) but maintains good properties in a general framework where other methods may fail.

The present paper is organized as follows. Section 2 introduces some preliminary theory and a triangular representation for fractional processes. Section 3 analyzes the behavior of the customary least squares statistics in the presence of no cointegrated NFI processes. We prove the invalidity of the customary OLS inference in this setup and suggest a local or narrow-band adjusted version in the frequency domain of the F-test. The asymptotic distribution of this Wald statistic is investigated for both spurious and cointegrated relationships. Section 4 gives a feasible and consistent version of the local Wald test for fractional cointegration and Section 5 examines the finite-sample properties of the test in a Monte Carlo experiment and illustrates our method by testing the cointegration hypothesis of nominal GNP and simple-sum ($M1$, $M2$, $M3$) monetary aggregates in the framework of the quantitative theory of money. Proofs of all results are outlined in the Appendix.

2. MODEL AND ASSUMPTIONS

2.1. *Preliminary Theory for Fractional Processes*

Let ε_t be an independent and identically distributed $(p \times 1)$ vector sequence of zero-mean random variables with $E(\varepsilon_1 \varepsilon_1') = I_p$, $E\|\varepsilon_1\|^{\mu} < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. For $d_i > \frac{1}{2}$, $i = 1, \ldots, p$, let

$$(1) \qquad z_t = \Delta(L)\{u_t \mathbb{1}_{t>0}(t)\} \qquad (t = 0, \pm 1, \pm 2, \ldots),$$

where $\mathbb{1}_A(\cdot)$ is the indicator function of the set A, $u_t = \sum_{j=-\infty}^{\infty} A_{t-j}\varepsilon_j$ is a linear process with long-run covariance matrix $\Omega = A(1)A(1)'$, $A(1) = \sum_{j=-\infty}^{\infty} A_j$, and the coefficient Δ_{ik} of L^k in the expansion of $(1-L)^{-d_i} = \sum_{k=0}^{\infty} \Delta_{ik} L^k$ is defined by

$$\Delta_{ik} = \frac{\Gamma(k+d_i)}{\Gamma(k+1)\Gamma(d_i)}, \quad \Gamma(s) = \int_0^{\infty} e^{-\eta}\eta^{s-1}\, d\eta \qquad (k = 1, 2, \ldots).$$

Therefore, z_t is a $(p \times 1)$ vector of NFI processes. For ease of exposition we assume that z_t is free of deterministic components. The case where z_t has both deterministic and NFI components can be analyzed, e.g., as in Marmol (1998). Truncation in u_t is necessary because the coefficients Δ_{ik} are not square-summable for $d_i > \frac{1}{2}$, $i = 1, \ldots, p$.

Define the normalizing matrix function D_n, for $d_z = (d_1, \ldots, d_p)$, as

$$D_n = \text{diag}\{n^{1/2-d_1}, \ldots, n^{1/2-d_p}\},$$

and let $z_n(r) = D_n z_{[nr]}$ for $0 \leq r \leq 1$, where $[nr]$ stands for the integer part of nr, n being the sample size. Let us further introduce the following assumptions.

ASSUMPTION A: $\sum_{j=-\infty}^{\infty} |j| \|A_j\| < \infty$ *and* $A(1)$ *is upper triangular with nonnegative diagonal elements.*

ASSUMPTION B: $\operatorname{rank}(A(1)) = p$.

ASSUMPTION C: $\mu > \max(4, 2/(2d_* - 1))$, $d_* = \min_{1 \leq i \leq p} d_i$.

The stationary linear specification for u_t in Assumption A entails a mild form of short-range dependence, whereas assuming that $E(\varepsilon_1 \varepsilon_1') = I_p$ and $A(1)$ is upper triangular causes essentially no loss of generality. See, e.g., Jansson and Haldrup (2002, p. 1322). Assumption B ensures that the limiting process will have nondegenerate finite-dimensional distributions. On the other hand, a larger d_* in Assumption C entails weaker moment conditions, at least for $d_* < 1$. A heuristic explanation is that a lower value of d_* implies a smaller normalization in $z_n(r)$ and hence tighter bounds are needed to obtain a well-defined limiting distribution. Assumptions A and C are satisfied, for example, whenever u_t has a Gaussian VARMA representation.

Under Assumptions A–C, Theorem 1 of Marinucci and Robinson (2000) implies that, as $n \to \infty$, $z_n(r) \Rightarrow B(d_z; r)$, where $\Rightarrow$ signifies convergence in the Skorohod J_1 topology of $D[0,1]^p$, the space of $\mathbb{R}^p$-valued vector functions on $[0, 1]$ whose components are continuous on the right with finite left limit, and

(2) $$B(d_z; r) = (0, \ldots, 0)' \quad \text{a.s.}, \qquad r = 0,$$

(3) $$B(d_z; r) = \int_0^r G(r, s)\, dB(s), \qquad r > 0,$$

with $B(r)$ being a $(p \times 1)$ Brownian motion with long-run covariance matrix Ω and with $G(r, s)$ being a $(p \times p)$ matrix with (i, j)th element $\Gamma(d_i)^{-1}(r-s)^{d_i - 1}$, $i, j = 1, \ldots, p$, for $0 \leq s \leq r$, and zero otherwise. Formally, $B(d_z; r)$ is defined in terms of a Holmgren–Riemann–Liouville fractional integral. It is a $(p \times 1)$ Gaussian vector process with almost surely continuous sample paths and non-independent (and nonstationary) increments. See Marinucci and Robinson (1999) for further details. In the particular case where $d_1 = d_2 = \cdots = d_p = d$,

(4) $$z_n(r) = n^{1/2-d} z_{[nr]} \Longrightarrow B(d; r) = \Gamma(d)^{-1} \int_0^r (r-s)^{d-1}\, dB(s), \qquad r > 0,$$

so that in the unit root case where $d = 1$ we obtain the well-known result $z_n(r) = n^{-1/2} z_{[nr]} \Rightarrow B(r)$, $r > 0$.

Consider now the partition $z_t = (y_t, x_t')'$, where y_t is a scalar NFI variate with memory parameter d_1 and x_t is an m vector of NFI processes with memory parameters $d_2, \ldots, d_p$ $(p = m + 1)$. It will be also useful in the sequel to work with Ω and $B(r)$ in partitioned format as

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{xy}' \\ \omega_{xy} & \Omega_{xx} \end{pmatrix}, \quad B(r) = \begin{pmatrix} B_y(r) \\ B_x(r) \end{pmatrix}, \tag{5}$$

with the partition conformable with $z_t = (y_t, x_t')'$. Further, in view of Assumption A and (5), we shall parameterize $A(1)$ as follows:

$$A(1) = \begin{pmatrix} \omega_{yy}^{1/2}(1-\rho^2)^{1/2} & \rho\overline{\omega}_{xy}'\Omega_{xx}^{-1/2} \\ 0 & \Omega_{xx}^{1/2} \end{pmatrix}, \tag{6}$$

where $\omega_{yy} > 0$, $\overline{\omega}_{xy}'$ is an m-vector satisfying $\overline{\omega}_{xy}'\Omega_{xx}^{-1}\overline{\omega}_{xy} = \omega_{yy}$, and

$$\rho^2 = \frac{\omega_{xy}'\Omega_{xx}^{-1}\omega_{xy}}{\omega_{yy}} \tag{7}$$

is the squared coefficient of multiple correlation computed from Ω, so that $0 \le \rho^2 \le 1$. The long-run covariance ω_{xy} is given by $\rho\overline{\omega}_{xy}$, where $\overline{\omega}_{xy}$ expresses the direction of the covariance while ρ measures the strength of the covariance. Consequently, $B(r) = A(1)W(r)$, where $W(r) = (W_y(r), W_x(r)')'$ denotes a $(p \times 1)$ vector of standard Brownian motions and

$$B_y(r) = \omega_{yy}^{1/2}(1-\rho^2)^{1/2}W_y(r) + \omega_{xy}'\Omega_{xx}^{-1/2}W_x(r), \tag{8}$$

$$B_x(r) = \Omega_{xx}^{1/2}W_x(r), \tag{9}$$

so that

$$B(d_z; r) = \int_0^r G(r,s)A(1)\,dW(s), \quad r > 0. \tag{10}$$

In addition to Assumptions A–C, let us introduce the following regularity conditions.

ASSUMPTION D: $\Omega_{xx} > 0$.

ASSUMPTION E: $d_1 = d_2 = \cdots = d_p = d > 1/2$.

Assumption D is fairly standard in the related literature and implies that x_t is a noncointegrated NFI process. Assumption E is a natural first research step. It is clearly stronger than needed in view of our definition of fractional cointegration and a higher degree of generality can be achieved by allowing different memory parameters. See Remark 2 below. Under Assumption E it follows from expression (10) that $B(d_z; r) = B(d; r) = (B_y(d; r), B_x(d; r)')' = A(1)W(d; r)$, where

$$W(d; r) = \Gamma(d)^{-1} \int_0^r (r - s)^{d-1}\, dW(s), \quad r > 0. \tag{11}$$

To complete the specification of the model, in the rest of the paper we shall assume that $d \in (\frac{1}{2}, \frac{3}{2})$, the most empirically relevant range. See, e.g., Gil-Alaña and Robinson (1997). Doing so will not alter most results in any interesting way and will simplify the exposition in Section 4.

2.2. *A Triangular Representation*

Applying the well-known polynomial decomposition $A(L) = A(1) + \Delta\widetilde{A}(L)$ to (1), we have

$$z_t = \Delta(L)A(1)\varepsilon_t + \Delta(L)\widetilde{A}(L)\Delta\varepsilon_t, \quad t \geq 0, \tag{12}$$

where $\widetilde{A}(L) = \sum_{j=-\infty}^{\infty} \widetilde{A}_j L^j$ is a lag polynomial with coefficients $\widetilde{A}_j = -\sum_{i=j+1}^{\infty} A_i < \infty$ by Assumption A.

Taken together, Assumptions D–E and (6) imply that, for $t \geq 0$, $z_t = (y_t, x_t')'$ can be represented as

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \omega_{yy}^{1/2}(1-\rho^2)^{1/2} & \rho\overline{\omega}_{xy}'\Omega_{xx}^{-1/2} \\ 0 & \Omega_{xx}^{1/2} \end{pmatrix} \begin{pmatrix} y_t^* \\ x_t^* \end{pmatrix} + \Delta^{1-d}\widetilde{A}(L)\varepsilon_t, \tag{13}$$

with $(y_t^*, x_t^{*'})' := \Delta^{-d}\varepsilon_t$ being uncorrelated NFI processes of order d. Then

$$y_t - \beta_0' x_t = \omega_{yy}^{1/2}(1-\rho^2)^{1/2} y_t^* + (1, -\beta_0')e_t, \tag{14}$$

where $e_t = \Delta^{1-d}\widetilde{A}(L)\varepsilon_t$ is a fractionally integrated process of order $d-1$ and β_0 is the projection *or fundamental* vector (cf. Park, Ouliaris, and Choi (1988)) computed from Ω, i.e.,

$$\beta_0 = \Omega_{xx}^{-1}\omega_{xy} = \Omega_{xx}^{-1}\rho\overline{\omega}_{xy}. \tag{15}$$

Notice that in the representation (14), the gap between the memory parameter of z_t and that of the error term is always equal to 1. To encompass the possibility of fractional cointegration, we generalize (1) by assuming that the Wald representation of the vector z_t is given by

$$z_t = \Delta(L)\{v_t \mathbb{1}_{t>0}(t)\} \qquad (t = 0, \pm 1, \pm 2, \ldots), \tag{16}$$

$$v_t = \{A(L) + \Delta^{d-\delta}C(L)\}\varepsilon_t, \tag{17}$$

for some $\delta < d$, where $C(L) = \sum_{j=-\infty}^{\infty} C_j L^j$ is a lag polynomial with coefficients C_j satisfying

$$(18) \qquad \sum_{j=-\infty}^{\infty} |j| \|C_j\| < \infty.$$

Basically, this amounts to including a perturbation of memory $d - \delta$ in the d-memory filter $\Delta(L)$ of (1).

Under Assumptions A–E, $z_t = (y_t, x_t')'$ defined by (16)–(17) continues verifying that $n^{1/2-d} z_{[nr]} \Rightarrow B(d; r) = A(1)W(d; r)$ with the same long-run covariance matrix Ω, but now it can be represented as

$$(19) \qquad z_t = A(1)\Delta^{-d}\varepsilon_t + \Delta^{1-d}\widetilde{A}(L)\varepsilon_t + \Delta^{-\delta}C(L)\varepsilon_t.$$

Thus, z_t in (16) can be decomposed as the sum of a $(p \times 1)$ vector of NFI processes of order d plus a $(p \times 1)$ vector of fractionally integrated processes of order $(d - 1)$ plus a $(p \times 1)$ vector of fractionally integrated processes of order δ, and such that

$$(20) \qquad y_t - \beta_0' x_t = \omega_{yy}^{1/2}(1 - \rho^2)^{1/2} y_t^* + (1, -\beta_0')e_t^*,$$

where now $e_t^* = \Delta^{1-d}\widetilde{A}(L)\varepsilon_t + \Delta^{-\delta}C(L)\varepsilon_t$ is a fractionally integrated process of order $\delta^* := \max\{d - 1, \delta\}$.

Since x_t is noncointegrated by Assumption D, the series $(y_t, x_t')'$ are fractionally cointegrated $(d, d, \ldots, d; \delta^*)$ with (unique) cointegrating vector $(1, -\beta_0')'$ if and only if $\rho^2 = 1$. When $|\rho| < 1$, $\{y_t - \beta_0' x_t\}$ is an NFI process of order d and the series $(y_t, x_t')'$ are spuriously related. In the extreme case where $\rho = 0$, $\omega_{xy} = 0$ so that y_t and x_t are asymptotically independent NFI processes of order d and $\beta_0 = 0$. This is the situation where the *maximal* degree of spuriousness is obtained.

3. REGRESSIONS WITH NFI PROCESSES

3.1. *Spurious OLS Inference*

We first study the asymptotic properties of basic OLS statistics for spuriously related NFI processes. Let $\widehat{\alpha}$ and $\widehat{\beta}$ be the OLS estimators in the multiple regression

$$(21) \qquad y_t = \widehat{\alpha}' q_t + \widehat{\beta}' x_t + res.,$$

where $q_t = (1, \ldots, t^{m_q - 1})'$ for some $m_q \geq 1$ ($m_q = 0$ would denote the absence of q_t in (21)), and let F be the standard F-statistic used to test the null hypothesis $H_0 : \beta = \beta_0$ based on the regression (21).

By defining the normalizing matrix

$$\Psi_n = \begin{pmatrix} \operatorname{diag}(n^{1/2}, \ldots, n^{m_q - 1/2}) & 0_{m_q \times m} \\ 0_{m \times m_q} & n^d I_m \end{pmatrix},$$

it is not difficult to prove that, in the spurious case, under Assumptions A–E, as $n \to \infty$,

$$(22) \qquad n^{-d}\Psi_n \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} - \beta_0 \end{pmatrix} \Rightarrow \omega_{yy}^{1/2}(1-\rho^2)^{1/2}\left(\int_0^1 Q(r)Q(r)'\,dr\right)^{-1}\left(\int_0^1 Q(r)W_y(d;r)\,dr\right),$$

so that $\widehat{\beta}$ is not a consistent estimator of β_0, and

$$(23) \qquad n^{-1}F \Rightarrow \frac{1}{m}\frac{\|\int_0^1 \widetilde{W}_x^q(d;r)W_y^q(d;r)\,dr\|^2}{\int_0^1 W_y^{Q*}(d;r)^2\,dr},$$

where $Q(r)' = (q(r)', B_x(d;r)')$, $Q_*(r)' = (q(r)', W_x(d;r)')$, $q(r) = (1, r, \ldots, r^{m_q-1})'$,

$$\widetilde{W}_x^q(d;r) = \left(\int_0^1 W_x^q(d;s)W_x^q(d;s)'\,ds\right)^{-1/2} W_x^q(d;r),$$

$$W_x^q(d;r) = W_x(d;r) - \left(\int_0^1 W_x(d;s)q(s)'\,ds\right)\left(\int_0^1 q(s)q(s)'\,ds\right)^{-1} q(r),$$

and $W_y^q(d;r)$ and $W_y^{Q*}(d;r)$ are defined similarly.

The F-statistic diverges at the rate n for all $d > \frac{1}{2}$. A similar result was first obtained by Jansson and Haldrup (2002) in the particular $d = 1$ case. The failure of standard OLS inference can be expected because the variance estimate based on the residual sum of squares (RSS) does not fully take into account the residual autocorrelation. The persistence or degree of serial dependence in the residuals of a spurious regression is equivalent to that of the levels of the data. Consequently, any underestimation of it would inevitably lead to spuriously significant test statistics.

Notice, however, that $n^{-1}F$ has a nondegenerate limiting distribution free of unknown parameters (apart from d, m, and m_q). In particular, it does not depend on ρ. Thus, after proper normalization, one could obtain F-type statistics with asymptotically pivotal distributions. This possibility is explored in the next subsection.

3.2. *Modified Wald Statistic*

From a frequency domain point of view, the persistence in the sample paths of trending NFI processes is reflected by periodograms with very high power at low frequencies. Due to this low frequency dominance, inference can be carried out in only a narrow band around zero frequency, neglecting high frequency behavior (cf. Robinson and Marinucci (2000, 2001)), so that we can define the class of *local frequency domain least squares estimates*

$$\widehat{\beta}_M = \left(\sum_{j=-M}^{M} I_{xx}(\lambda_j) \right)^{-1} \sum_{j=-M}^{M} I_{xy}(\lambda_j), \tag{24}$$

where $1 \leq M \leq n/2$, $\lambda_j = 2\pi j/n$ are the Fourier frequencies, and

$$I_{ab}(\lambda_j) = w_a(\lambda_j) w_b(-\lambda_j)', \quad w_a(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^{n} a_t \exp(it\lambda_j),$$

is the (cross) periodogram matrix of any vectors a_t and b_t. The OLS coefficients satisfy $\widehat{\beta} = \widehat{\beta}_{[n/2]}$ (assuming n is odd) by symmetry and Parseval's formula, but if $Mn^{-1} + M^{-1} \to 0$ as $n \to \infty$, $\widehat{\beta}_M$ is the *narrow band least squares* estimator of Robinson and Marinucci (2000, 2001), which neglects the information at high frequencies. Dropping the zero frequency in the summations in (24) would account for estimation with intercept by mean correction, while the symmetric sum around $j = 0$ eliminates the imaginary parts of the periodogram matrices.

For studentization of LS coefficients we propose the following feasible estimate of the variance–covariance matrix of $\widehat{\beta}_M$:

$$\widehat{V}_M = \left(\sum_{j=-M}^{M} I_{x_q x_q}(\lambda_j) \right)^{-1} \sum_{j=-M}^{M} I_{x_q x_q}(\lambda_j) I_{\widehat{\xi}\widehat{\xi}}(\lambda_j) \left(\sum_{j=-M}^{M} I_{x_q x_q}(\lambda_j) \right)^{-1}, \tag{25}$$

with $I_{x_q x_q}(\lambda_j)$ being the periodogram of the series $x_{q,t} = x_{q,t}(M)$ obtained after detrending x_t by q_t, namely, $x_{q,t} = x_t - \sum_{j=-M}^{M} I_{xq}(\lambda_j)(\sum_{j=-M}^{M} I_{qq}(\lambda_j))^{-1} q_t$, and where $I_{\widehat{\xi}\widehat{\xi}}(\lambda_j)$ stands for the residual periodogram computed with the observed residuals $\widehat{\xi}_t = \widehat{\xi}_t(M) = y_t - (\widehat{\alpha}'_M q_t + \widehat{\beta}'_M x_t)$. The use of statistics similar to $\widehat{V}_M$ is common in autocorrelation-robust frequency domain studentization for weakly dependent stationary time series regressions, with $I_{\widehat{\xi}\widehat{\xi}}(\lambda_j)$ replaced by a smoothed estimate of the error spectral density, and where proper standardization requires consideration of all frequencies. See, e.g., Robinson and Velasco (1997).

Next result analyzes the effects of focusing on a degenerating band around the origin in our spurious framework.

LEMMA 1: *Under Assumptions* A–E, *with* $\rho^2 < 1$, *if* $\widehat{\alpha} = \widehat{\alpha}_{[n/2]}$, $\widehat{\beta} = \widehat{\beta}_{[n/2]}$ *denote the OLS estimates and* $\widehat{V} = \widehat{V}_{[n/2]}$, *then*

$$n^{-d}\Psi_n \begin{pmatrix} \widehat{\alpha} - \widehat{\alpha}_M \\ \widehat{\beta} - \widehat{\beta}_M \end{pmatrix} = O_p(M^{1-2d} + M^{-1}) = o_p(1) \quad as \quad M, n \to \infty;$$

$$\widehat{V} - \widehat{V}_M = O_p\big(M^{1-4\min\{1,d\}}\big) = o_p(1) \quad as \quad M, n \to \infty.$$

Notice that the result is valid as long as M increases with n as slowly as desired and it could also be a fixed proportion of n. Two immediate corollaries of Lemma 1 are that $\widehat{\beta}_M$ has the same limiting distribution as $\widehat{\beta}$ (see (22)), and that its corresponding usual F-statistic is also $O_p(n)$. Thus, though narrow band OLS is able to control some bias terms in the asymptotic distribution of regression estimates for some cointegrated NFI vectors (cf. Robinson and Marinucci (2001)) no apparent gains to alleviate the spurious problem seem to be achieved by using restricted versions in the frequency domain of OLS. In a similar fashion to standard OLS statistics, the asymptotic behavior of $\widehat{V}$ is only determined by the very low frequencies. Therefore, Lemma 1 allows us to restrict the analysis to full band statistics with $M = [n/2]$ without loss of generality.

Using $\widehat{V}$, computed with the OLS residuals, we propose the following Wald or adjusted F-statistic:

$$\mathcal{W}(\widehat{\beta}, \beta_0) := \frac{1}{m}(\widehat{\beta} - \beta_0)'\widehat{V}^{-1}(\widehat{\beta} - \beta_0). \tag{26}$$

By contrast with the customary F-statistic, constructed using the usual (time-domain) RSS, the Wald statistic $\mathcal{W}$ has a well-defined limiting distribution as given in the next theorem.

THEOREM 1: *Under Assumptions* A–E, *with* $\rho^2 < 1$, *as* $n \to \infty$,

$$\mathcal{W}(\widehat{\beta}, \beta_0) \Rightarrow \mathcal{W}_\infty := \frac{1}{m}\int_0^1 W_y(d;r)W_x^q(d;r)'\,dr\, V^{-1}\int_0^1 W_x^q(d;r)W_y(d;r)\,dr, \tag{27}$$

where

$$V := \int_0^1 \gamma_R(s)\big\{\boldsymbol{\gamma}_{x_q x_q}(s) + \boldsymbol{\gamma}_{x_q x_q}(s)' + \boldsymbol{\gamma}_{x_q x_q}(1-s) + \boldsymbol{\gamma}_{x_q x_q}(1-s)'\big\}\,ds,$$

$$\gamma_R(s) := \int_0^{1-s} W_y^{Q*}(d;r)W_y^{Q*}(d;r+s)\,dr,$$

$$\boldsymbol{\gamma}_{x_q x_q}(s) := \int_0^{1-s} W_x^q(d;r)W_x^q(d;r+s)'\,dr.$$

The normalization $\widehat{V}$ is not consistent for any limit matrix for nonstationary series, but provides an adequate automatic normalization so that $\mathcal{W}(\widehat{\beta}, \beta_0)$ is $O_p(1)$, with a nondegenerated limiting distribution free of nuisance parameters (apart from d, m, and m_q). Note that $\mathcal{W}_\infty = \mathcal{W}_\infty(d, m, m_q)$ is defined in terms of $W_x^q(d, r)$ and not using $\widetilde{W}_x^q(d, r)$ as in (23) because the normalization is included in V.

The fact that $\widehat{\beta} - \beta_0$ and $\widehat{V}$ weakly converge in such a way that $\mathcal{W}(\widehat{\beta}, \beta_0)$ is $O_p(1)$, however, prevents $\mathcal{W}(\widehat{\beta}, \beta_0)$ from providing a consistent testing device for hypotheses on β_0, since it can be checked that $\mathcal{W}(\widehat{\beta}, \beta)$ remains $O_p(1)$ for any β.

REMARK 1: It is possible to studentize the OLS estimates with a variety of heteroscedasticity and autocorrelation consistent (HAC) robust procedures, including the following time domain estimate,

$$\widetilde{V} = n\left(\sum_{t=1}^{n} x_{t,q}x'_{t,q}\right)^{-1} \times \sum_{j=0}^{n-1} \mathbb{I}_0(j)\widehat{\gamma}_{\widehat{\xi}\widehat{\xi}}(j)\{\widehat{\boldsymbol{\gamma}}_{x_q x_q}(j) + \widehat{\boldsymbol{\gamma}}_{x_q x_q}(j)'\}\left(\sum_{t=1}^{n} x_{t,q}x'_{t,q}\right)^{-1},$$

where $\mathbb{I}_0(0) = \frac{1}{2}$; $\mathbb{I}_0(j) = 1$, $j \neq 0$, $\widehat{\gamma}_{\widehat{\xi}\widehat{\xi}}(j) = n^{-1}\sum_{t=1}^{n-j}\widehat{\xi}_t\widehat{\xi}_{t+j}$, and $\widehat{\boldsymbol{\gamma}}_{x_q x_q}(j) = (1/n)\sum_{t=1}^{n-j} x_{t,q}x'_{t+j,q}$ (see, e.g., Robinson (1998)). $\widetilde{V}$ is asymptotically equivalent to $\widehat{V}$ for weakly dependent time series, but for NFI processes the asymptotic distribution of $\widetilde{V}$ is slightly different from that of $\widehat{V}$. See Marmol and Velasco (2002) for a similar proposal.

REMARK 2: The assumption of equal memory for all series is not critical for these results and Assumption E can be relaxed to allow for different values of $d_z = (d_1, d_2, \ldots, d_p)$. Assume for instance that the vector z_t can be partitioned as $z_t = (y_t, x'_{1t}, x'_{2t})'$, where y_t is a scalar NFI variate with memory parameter d_1, and x_{1t} and x_{2t} are m_1- and m_2-dimensional vectors of NFI processes ($m_1 + m_2 = m$) with memory parameters $\underline{d}$ and $\overline{d}$, respectively, with $\overline{d} > \underline{d} > \frac{1}{2}$. According to our definition of fractional cointegration, a necessary condition for cointegration in this fractional setting is that $d_1 = \overline{d}$. Let

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{y1} & \omega_{y2} \\ \omega_{1y} & \Omega_{11} & \Omega_{12} \\ \omega_{2y} & \Omega_{21} & \Omega_{22} \end{pmatrix},$$

conformably with z_t, where the diagonal submatrices Ω_{11} and Ω_{22} are assumed to be positive definite such that x_{1t} and x_{2t} are not permitted to be individually cointegrated. Then, it can be proved as in Theorem 1 that, under Assumptions A–C, as $n \to \infty$, when the elements of z_t are spuriously related of order $\overline{d}$, the Wald statistic has a well-defined asymptotic distribution depending upon the number of NFI processes of orders $\underline{d}$ and $\overline{d}$ in the system.

3.3. *Wald Test under Cointegration*

We now address the asymptotic analysis of our Wald statistic with a HAC-type studentization under cointegration. For this, let $z_t = (y_t, x_t')'$ be fractionally cointegrated of orders $(d, d, \ldots, d; \delta^*)$, $\delta^* < d$, so that Ω is singular and, in view of Assumption D, $\rho^2 = 1$ and $\beta_0 = \Omega_{xx}^{-1}\overline{\omega}_{xy}$. In this case, (20) becomes

$$(28) \qquad y_t - \beta_0' x_t = \xi_t,$$

where $\xi_t = (1, -\beta_0')e_t^*$.

ASSUMPTION F: (i) *z_t is fractionally cointegrated (so that $\rho^2 = 1$) with $d > \delta > \frac{1}{2}$;*
(ii) $\omega_{\xi\xi.x} = \omega_{\xi\xi} - \omega_{x\xi}'\Omega_{xx}^{-1}\omega_{x\xi} > 0$;
(iii) $(1, -\beta_0')C(1)(1, 0_{1\times m})' > 0$.

With condition (i) we focus on cointegrated systems with nonstationary error terms to guarantee that $\delta^* = \delta$ in our triangular representation, so that the error term ξ_t essentially behaves as an NFI process of order δ with innovation $c_t := (1, -\beta_0')C(L)\varepsilon_t$. See, however, Remark 3 in Section 4 for the stationary case. Assumption F(ii) states that the fractional cointegration is regular in the sense of Park (1992, Definition 2.3). Assumption F(iii) entails that the corresponding limiting process has nondegenerate finite-dimensional distributions.

For $z_t^* = (\xi_t, x_t')'$, let

$$\Omega^* = \begin{pmatrix} \omega_{\xi\xi} & \omega_{x\xi}' \\ \omega_{x\xi} & \Omega_{xx} \end{pmatrix}$$

stands for the long-run covariance matrix of $(c_t, \Delta^d x_t')'$ with Ω^* partitioned in the obvious way. Define the normalizing matrix function $D_n^* = \mathrm{diag}\{n^{1/2-\delta}, n^{1/2-d}I_m\}$, and let $z_n^*(r) = D_n^* z_{[nr]}^*$, for $0 \le r \le 1$. Then, under Assumptions A–F, as $n \to \infty$, $z_n^*(r) \Rightarrow (B(\delta; r), B(d; r)')'$ using the obvious notation. Using the Continuous Mapping Theorem (CMT), it follows that

$$n^{-\delta}\Psi_n \begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} - \beta_0 \end{pmatrix} \Rightarrow \left(\int_0^1 Q(r)Q(r)'\,dr\right)^{-1} \left(\int_0^1 Q(r)B_\xi(\delta; r)\,dr\right),$$

obtaining the well-known result that $\widehat{\beta}$ is consistent for β_0 under fractional cointegration. It is also possible to show that $\widehat{\alpha}_M$ and $\widehat{\beta}_M$ have the same asymptotic distribution under fractional cointegration than their full-band counterparts if $M^{-1} + Mn^{-1} \to 0$ (cf. Robinson and Marinucci (2001)). This consistency of $\widehat{\beta}$, however, has no effect on the convergence rate of $\mathcal{W}(\widehat{\beta}, \beta_0)$, whose asymptotic behavior under fractional cointegration is described in the following result. Proof of this theorem follows as in Theorem 1 and is thus omitted.

THEOREM 2: *Under Assumptions* A–F, *as* $n \to \infty$,

$$\mathcal{W}(\widehat{\beta}, \beta_0) \Rightarrow \mathcal{W}^*_\infty := \frac{1}{m} \int_0^1 B_\xi(\delta; r) B_x^q(d; r)' \, dr \, V^{*-1} \int_0^1 B_x^q(d; r) B_\xi(\delta; r) \, dr$$

and

$$V^* := \int_0^1 \gamma_R^*(s) \{ \boldsymbol{\gamma}^*_{x_q x_q}(s) + \boldsymbol{\gamma}^*_{x_q x_q}(s)' + \boldsymbol{\gamma}^*_{x_q x_q}(1-s) + \boldsymbol{\gamma}^*_{x_q x_q}(1-s)' \} \, ds,$$

$$\gamma_R^*(s) := \int_0^{1-s} B_\xi(\delta; r) B_\xi(\delta; r+s) \, dr,$$

$$\boldsymbol{\gamma}^*_{x_q x_q}(s) := \int_0^{1-s} B_x^q(d; r) B_x^q(d; r+s)' \, dr.$$

Under cointegration $\widehat{\beta}$ is $n^{d-\delta}$-consistent for β_0 but $\widehat{V}^{-1}$ diverges as $n^{2(d-\delta)}$, producing an automatic normalization of the Wald statistic $\mathcal{W}(\widehat{\beta}, \beta_0)$ as in the spurious case, though with a different asymptotic distribution. By contrast with the spurious case, this result would allow consistent testing of the value of β_0 under cointegration, since $\mathcal{W}(\widehat{\beta}, \beta)$ diverges as $n^{2(d-\delta)}$ as n increases for any $\beta \neq \beta_0$.

In our single equation framework we are interested in testing the hypothesis that there exists a cointegrating vector. Indeed, since there are p variables contained in z_t, there could be up to $p^+ < p$ linearly independent $(p \times 1)$ vectors $H = (h_1, h_2, \ldots, h_{p^+})$ such that $H' z_t$ is a fractionally integrated $(p^+ \times 1)$ vector of orders $(\delta_1, \delta_2, \ldots, \delta_{p^+})$, say, with $d > \delta^+ (:= \max\{\delta_1, \delta_2, \ldots, \delta_{p^+}\}) > \frac{1}{2}$. Then it is simple to generalize our triangular representation allowing for $p^+ \geq 1$ cointegrating relationships. In contrast to the standard $\mathrm{I}(1)/\mathrm{I}(0)$ cointegration case (cf. Wooldridge (1991), Johansen (1992)) it can be proved under mild regularity conditions that, when $p^+ > 1$, the single equation OLS estimates $\widehat{\beta}$ do not provide a consistent estimate of a suitable linear combination of the cointegrating relations, though they remain bounded in probability. In the particular case of a common error memory $\delta^+ = \delta_i$, $i = 1, \ldots, p^+$, the OLS residuals $\widehat{\xi}_t$ can be shown to approximate a linear combination of $\mathrm{NFI}(\delta^+)$ processes as in the single equation setup. Nonetheless, not all the conditioning regressors in (21) will satisfy Assumption D, affecting the limit distributions of $\widehat{V}$ and $\mathcal{W}(\widehat{\beta}, \beta_0)$.

4. A TEST OF COINTEGRATION

We are interested in testing the null hypothesis

$$H_0 : (y_t, x_t')' \text{ are spuriously related NFI processes of order } d > \tfrac{1}{2}$$

against (nonstationary) fractional cointegration alternatives where $\delta > \frac{1}{2}$,

$$H_1 : (y_t, x_t')' \text{ are fractionally cointegrated of orders } (d, d, \ldots, d; \delta),\ \tfrac{1}{2} < \delta < d.$$

For convenience of analysis, herein we restrict the analysis to a form of weak cointegration where $\delta > \frac{1}{2}$ under H_1. For $\delta \leq \frac{1}{2}$, the asymptotic theory changes, but our test can be shown to remain consistent; see Remark 3.

In previous sections, we have illustrated the asymmetric behavior of the *numerator* and *denominator* of $\mathcal{W}(\widehat{\beta}, \beta_0)$ under spurious and cointegrated relationships, in spite of the fact that $\mathcal{W}(\widehat{\beta}, \beta_0)$ converges to well-defined distributions under both hypotheses. Suppose now that an estimate $\widehat{\beta}_0$ of β_0 is available, consistent under H_0, but inconsistent under H_1. Then, using the OLS residuals, we compute $\widehat{V}$ and

$$\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0) = (\widehat{\beta} - \widehat{\beta}_0)'\widehat{V}^{-1}(\widehat{\beta} - \widehat{\beta}_0),$$

so $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0) - \mathcal{W}(\widehat{\beta}, \beta_0) = o_p(1)$ under H_0. However, under H_1, the numerator $\widehat{\beta} - \widehat{\beta}_0$ of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0)$ is bounded away from zero, but the denominator $\widehat{V}$ computed with OLS residuals is still $o_p(1)$. This suggests a consistent cointegration test by means of comparing $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0)$ with the appropriate critical value of the $\mathcal{W}_\infty$ distribution, so the test has correct asymptotic type-I error (cf. Theorem 1), but from Theorem 2, $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0)$ diverges as $n \to \infty$ because $\widehat{V}^{-1}$ diverges while $\widehat{\beta}_0$ is inconsistent for β_0.

To develop this idea, we propose in this section a semiparametric narrow band estimate $\widehat{\beta}_0^N$, depending on a bandwidth number N, which diverges with n under H_1. This permits $\widehat{V}^{-1}$ and $\widehat{\beta}_0^N$ to diverge simultaneously under cointegration, accelerating the rate of divergence of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ under the alternative hypothesis. Our estimate of $\widehat{\beta}_0^N$ depends in turn on knowledge of the memory d and $\delta(\leq d)$ of levels and errors, so we also discuss semiparametric estimation of such parameters, which permits us to keep modelling assumptions to a minimum.

In the spirit of the Hausman (1978) principle, our approach relies upon comparing two estimates of the projection vector β_0 with different properties under the competing hypotheses. However, our test has several important differences with respect to the standard case where both estimates are $n^{1/2}$-consistent under the null hypothesis, with asymptotic normal distributions but different efficiency. First, in our framework, limiting distributions are not standard and rates of convergence of basic statistics change under the different hypotheses, so we use an automatic normalization by $\widehat{V}$. Second, instead of comparing efficient vs. inefficient estimates under the null, in our test we compare consistent vs. inconsistent estimates. As a consequence, the asymptotic distribution of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0)$ is the same as that of $\mathcal{W}(\widehat{\beta}, \beta_0)$ under H_0. Furthermore, under the alternative, $\widehat{\beta}_0$ is not only inconsistent for β_0, but diverges, improving the rate at which $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0)$ diverges when cointegration is detected.

4.1. *Memory Estimation*

We consider H_0 and H_1 simultaneously with $z_t^* := (\xi_t, x_t')'$ and $\xi_t := y_t - \beta_0' x_t$ with $z_t = (y_t, x_t')'$ given by (16) and (17) in both cases. Most theoretical work on semiparametric inference on long memory and fractional time series has been developed under the assumption that the observed series, differenced an integer number of times if necessary, are exactly covariance stationary. Then, under Assumptions A and B, we define the stationary version of the increments Δz_t^* of the NFI vector z_t^*, as

$$\Delta s_t^* = \text{diag}\{\Delta^{1-\delta}, \Delta^{1-d}, \dots, \Delta^{1-d}\} v_t^* \qquad (t = 0, \pm 1, \pm 2, \dots),$$

now feasible with infinite fractional filters since $(1-\delta), (1-d) \in (-\frac{1}{2}, \frac{1}{2})$. The vector process $v_t^* = (v_t^\xi, v_t^{x\prime})'$ is covariance stationary, where $v_t^\xi = [A^y(L) + \Delta^{d-\delta} C^y(L)]\varepsilon_t - \beta_0' v_t^x$ and $v_t^x = [A^x(L) + \Delta^{d-\delta} C^x(L)]\varepsilon_t$, with $A(L)$ and $C(L)$ partitioned in the obvious way, and since $d - \delta > 0$, v_t^* is integrated of order zero. It can be checked that Δs_t^* converges in a mean square sense to Δz_t^* as t grows because the initial conditions play asymptotically no role when the degree of integration is less that $\frac{1}{2}$, though their effect may last for a long period. This initial conditions problem might affect the properties of estimates and test statistics based on the observed, only asymptotically stationary, Δz_t^*, compared to those based on the stationary Δs_t^* (see, e.g., Velasco (2002)).

Under Assumptions A–F and (18) (allowing for $\rho^2 \leq 1$) the spectral density matrix of Δs_t^* satisfies

$$H_{\Delta s^* \Delta s^*}(\lambda) = \frac{1}{2\pi} \Lambda \Omega^* \bar{\Lambda}\big(1 + O(\lambda + \lambda^{d-\delta})\big) \qquad (29)$$

as $\lambda \to 0^+$, where $\Lambda := \text{diag}\{e^{i\pi(\delta-1)/2}\lambda^{1-\delta}, e^{i\pi(d-1)/2}\lambda^{1-d}, \dots, e^{i\pi(d-1)/2}\lambda^{1-d}\}$, and $\bar{\Lambda}$ is the complex conjugate of Λ. See Lobato and Robinson (1996) and Lobato (1999).

Under this set of assumptions we obtain that each element of $A(\lambda) := \sum_{j=0}^{\infty} A_j \exp(ij\lambda)$ and $C(\lambda)$, defined similarly, is differentiable in $(0, \pi]$ with

$$(30) \qquad \frac{d}{d\lambda} A_a(\lambda) = O\big(\lambda^{-1} \| A_a(\lambda) \|\big), \quad \lambda \to 0^+,$$

where A_a is the ath row of $A(\lambda)$. This implies that each element of $H_{\Delta s^* \Delta s^*}(\lambda)$ is differentiable in $(0, \pi]$ and

$$\frac{d}{d\lambda} H_{ab}(\lambda) = O\big(\lambda^{-1} |H_{ab}(\lambda)|\big) \quad \text{as} \quad \lambda \to 0^+ \qquad (a, b = 1, \ldots, p).$$

For the estimation of d, we minimize the concentrated local Gaussian likelihood (cf. Robinson (1995a), Lobato (1999)),

$$(31) \qquad Y_x^M(d) := \log |\widehat{\Omega}_{xx}^M(d)| - \frac{2m(d-1)}{M} \sum_{j=1}^{M} \log(\lambda_j),$$

with

$$(32) \qquad \widehat{\Omega}_{xx}^M(d) := \frac{2\pi}{M} \sum_{j=1}^{M} \Lambda_j^{-1}(d) \operatorname{Re}\{I_{\Delta x \Delta x}(\lambda_j)\} \Lambda_j^{-1}(d),$$

where Re stands for real part. Note that $\Lambda_j(d) := \operatorname{diag}\{\lambda_j^{1-d}, \ldots, \lambda_j^{1-d}\}$ is approximately, for low frequencies, the (square root) spectrum of Δx_t, up to a constant. Here M grows slower than n as $n \to \infty$. $\widehat{\Omega}_{xx}^M(d)$ can be replaced by any of its diagonal terms (and the factor m in (31) by 1) to obtain an initial univariate consistent estimate.

An efficient semiparametric estimate $\widehat{d}_M$ can be calculated through a Newton–Raphson iteration

$$(33) \qquad \widehat{d}_M^{(2)} = \widehat{d}_M^{(1)} - \left(\left. \frac{\partial^2 Y_x^M(d)}{\partial d^2} \right|_{\widehat{d}_M^{(1)}} \right)^{-1} \left(\left. \frac{\partial Y_x^M(d)}{\partial d} \right|_{\widehat{d}_M^{(1)}} \right),$$

starting from a root-M consistent estimate $\widehat{d}_M^{(1)}$, obtained, e.g., by the previous univariate Gaussian (or by a log-periodogram) procedure applied to Δx_{1t} or any other component of Δx_t (or Δy_t).

This method gives a better-than-$\log n$ consistent estimate $\widehat{d}_M$, i.e., $\log n(\widehat{d}_M - d) \xrightarrow{p} 0$ (see Robinson (1994, 1997)), if we could use as input Δs_t^* and $M^{-1} \log^2 n \to 0$. However, in our setup we have to check that the use of Δx_t makes no asymptotic difference for consistency of estimates of d. This is formalized in the next lemma. We concentrate only on the Gaussian semiparametric estimate, but a similar result is possible for the log-periodogram estimate under stronger conditions on the distribution of z_t. See Hassler, Marmol, and Velasco (2003).

LEMMA 2: *Under Assumptions* A–F, (18), *and*

$$\{M^{d-2} + M^{\epsilon-1}\log n\}\log^2 n + Mn^{-1} \to 0 \tag{34}$$

as $n \to \infty$, *for some* $\epsilon > 0$, *then* $\log n(\widehat{d}_M - d) \xrightarrow{p} 0$, *where*

$$\widehat{d}_M = \arg \min_{\tau\in[\nabla_1,\nabla_2]} Y_x^M(\tau)$$

and the true value $d \in [\nabla_1, \nabla_2] \subset (\frac{1}{2}, \frac{3}{2})$.

The choice of M for memory estimation is only slightly more restricted than for the study of the properties of the (local) LS $\widehat{\beta}_M$ in Lemma 1, but note that (34) holds if we choose $M \sim Kn^a$ for any $0 < a < 1$ and some positive constant K. The mean square optimal rate is given by $M \sim n^{4/5}$ for many processes with regular spectral densities, including the increments of the stationary versions of NFI processes with Assumption A strengthened to 2-summability (see Robinson and Henry (1996)). However, in our setup with innovation process v_t as in (17), where $d \in (\frac{1}{2}, \frac{3}{2})$ and $\delta > \frac{1}{2}$ so that $d - \delta < 1$, the optimal rate of M for estimation of d is at most $n^{2(d-\delta)/[2(d-\delta)+1]}$ under both H_0 and H_1. This rate can be arbitrarily slow for δ close to d and never better than $n^{2/3}$.

The memory δ of the errors ξ_t is estimated by means of the increments $\Delta\widehat{\xi}_t$ of the OLS residuals through

$$\begin{aligned} \widehat{\delta}_M &= \arg \min_{\tau\in[\nabla_1,\nabla_2]} Y_\xi^M(\tau); \\ Y_\xi^M(\delta) &:= \log \widehat{\Omega}^M_{\widehat{\xi}\widehat{\xi}}(\delta) - \frac{2(\delta-1)}{M}\sum_{j=1}^{M}\log(\lambda_j). \end{aligned} \tag{35}$$

Because under the alternative the OLS estimate is consistent for β_0 with possibly a fast rate of convergence, the $\widehat{\xi}_t$ are asymptotically close to ξ_t and $\widehat{\delta}_M$ is $\log n$-consistent for the memory $\delta < d$ of the cointegration errors. By contrast, under H_0, the residuals $\widehat{\xi}_t$ are a linear combination of NFI(d) processes (conditionally on the estimate $\widehat{\beta}$) so their memory estimate $\widehat{\delta}_M$, based on increments $\Delta\widehat{\xi}_t$, is $\log n$-consistent for d under H_0 and an appropriate choice of the bandwidth M; cf. (34).

4.2. *Consistency of the Test*

Since $\beta_0 = \Omega_{xx}^{-1}\omega_{xy}$, we consider separately the estimation of Ω_{xx} and ω_{xy}. To estimate Ω_{xx} we use the GLS-type of estimate (32) obtained in the estimation of d, where the computation and consistency of $\widehat{\Omega}^N_{xx}(d)$ requires a

$\log n$-consistent estimate of d. We allow in our notation for possibly different values of the bandwidth numbers, M and N. Then, with $\widehat{d}_M$ defined as in the previous subsection, and using the same techniques, it is straightforward to show that

$$\widehat{\Omega}_{xx}^N(\widehat{d}_M) \xrightarrow{p} \Omega_{xx}, \tag{36}$$

where N satisfies the same conditions of M in (34), and could be equal to or different from M.

For the estimation of ω_{xy} we use a similar strategy by means of $\widehat{\omega}_{xy}^N(\widehat{\delta}_M)$, where

$$\widehat{\omega}_{xy}^N(\delta) := \frac{2\pi}{N} \sum_{j=1}^{N} \operatorname{Re} I_{\Delta x \Delta y}(\lambda_j) \lambda_j^{2(\delta-1)}. \tag{37}$$

$\widehat{\omega}_{xy}^N(\widehat{\delta}_M)$ is consistent for ω_{xy} under H_0 and the assumptions of Lemma 2 because the memory of errors is equal to d under H_0. However the errors are NFI(δ), $\delta < d$, under H_1 so the estimates $\widehat{\delta}^M$ and $\widehat{\omega}_{xy}^N(\delta)$ have different properties under H_1.

Then, our proposal to complete the estimate of β_0 is

$$\widehat{\beta}_0^N = \widehat{\beta}_0^N(\widehat{d}_M, \widehat{\delta}_M) := \widehat{\Omega}_{xx}^N(\widehat{d}_M)^{-1} \widehat{\omega}_{xy}^N(\widehat{\delta}_M),$$

from which it follows at once that

$$\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N) \Rightarrow \mathcal{W}_\infty \quad \text{under } H_0 \tag{38}$$

from the consistency of $\widehat{\beta}_0^N$. In the next theorem we summarize the conditions for the consistency of our Wald test for cointegration testing.

THEOREM 3: *Under Assumptions* A–F, (18), *and* (34), *for both M and N,*

$$\textit{prob} \liminf_{n\to\infty} n^{6(\delta-d)} N^{4(d-\delta)} \mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N) > 0, \quad \textit{under } H_1,$$

if $d - \delta < \frac{1}{2}$ and

$$\textit{prob} \liminf_{n\to\infty} n^{6(\delta-d)} N \mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N) > 0, \quad \textit{under } H_1,$$

if $d - \delta \geq \frac{1}{2}$, so the test based on rejecting H_0 of no cointegration for large values of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ is consistent for H_1.

The divergence rate of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ under H_1 can be arbitrarily slow if d is very close to δ. In practice we expect lower power the higher N but with $Nn^{-1} \to 0$ as $n \to \infty$. Note that the choices of N and M are not linked, so M can be chosen so as to produce the most reliable estimates of d and δ as possible, because M does not enter into the divergence rate of $\mathcal{W}$ under H_1.

The consistency of the $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ test is based on the different properties of the residual memory estimate $\widehat{\delta}_M$, and thus of $\widehat{\beta}_0^N$, under the null and alternative hypotheses. A direct semiparametric Wald test based on the memory $\widehat{\delta}_M$ estimated from residuals compared to $\widehat{d}_M$, estimated from original data, could diverge with n at most at the rate $M^{\min\{1,2(d-\delta)\}}$ under H_1 (cf. Velasco (2003)). However the properties of such a test are yet unknown under the null hypothesis, and could depend on the particular regression estimate used to obtain these residuals. Similar inefficiency problems might arise for the Hausman test based on comparing different semiparametric estimates of d proposed in Marinucci and Robinson (2001). Alternatively, the inclusion of $\widehat{\delta}_M$ in our Wald statistic through $\widehat{\beta}_0^N$, allows us to outperform this semiparametric rate up to $n^{6(d-\delta)}N^{4(\delta-d)}$, which, taking $M = N$, is always faster than $M^{\min\{1,2(d-\delta)\}}$ because $Nn^{-1} \to 0$.

In the case where the number of cointegrating relationships is unknown, OLS residuals still can be used to obtain consistent semiparametric estimates of δ, but when $p^+ > 1$, $\widehat{d}_M$ is no longer consistent due to the failure of Assumption D on x_t (as exploited for z_t by Robinson and Marinucci (2001)), so it can be a good policy in general to base $\widehat{d}_M$ on a restricted set of variables known not to be cointegrated, possibly just one single series. Then, despite the fact that $\widehat{V}^{-1}$ does not diverge as $n \to \infty$ when $p^+ > 1$, the $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ test still inherits the consistency properties of the residual based $\widehat{\delta}_M$.

For unit root series, $d = 1$, Krämer and Marmol (2004) have proved that standard ordinary residual-based tests are consistent under fractional cointegration alternatives, showing that the Z_α test diverges faster than the (t-type) augmented Dickey–Fuller (ADF) test under fractionally cointegrated alternatives (see Section 5.1 for details on these tests). Specifically, they find that the divergence rate of the ADF test is $O_p(n^{1-\delta})$, whereas the divergence rate of the Z_α test is $O_p(n^{2(1-\delta)})$ if $\frac{1}{2} < \delta < 1$. By contrast, when $\frac{1}{2} < \delta < d = 1$, the square root of our $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ test statistic diverges at the rate $n^{1-\delta}(nN^{-1})^{2(1-\delta)}$, which is always faster than that of the ADF test (because $Nn^{-1} \to 0$) and also than Z_α if $Nn^{-1/2} \to 0$. Hence, the fact that $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ takes advantage of the possibly simultaneous superconsistency of OLS estimates and of the (semiparametric) inconsistency of GLS estimates under cointegration, entails no asymptotic loss of power compared with these *parametric* procedures in spite of the semiparametric nature of our modelization.

REMARK 3: For convenience of analysis in the previous two subsections we have considered a form of weak cointegration where $\delta > \frac{1}{2}$ in H_1. In this sense, the possibility of having stationary equilibrium errors may affect the asymptotic properties of our tests in various ways. In the first place it might be possible that $d - 1 \geq \delta$, so the cointegration vector in general is of memory

$\delta^* = \max\{\delta, d-1\}$. Then the rate of convergence of OLS estimates of β has a different form when $d+\delta^* \le 1$, $\delta^* > 0$, and the studentization by $\widehat{V}$ also has different properties. Regarding memory estimates, we can only hope that $(\widehat{\delta}^*_M - 1) \xrightarrow{p} -\frac{1}{2}$ because of the use of (noninvertible) increments of regression residuals, so the divergence rate of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ also would be affected by this fact. If knowledge of $d < 1$ can be used (so $\delta < 1$), the estimate of δ^* can be constructed in terms of original residuals $\widehat{\xi}_t$ or on tapered observations (see Velasco (2003)), increasing the asymptotic power. Nonetheless, in the Monte Carlo experiment in the next section we show that the empirical power of the Wald test is not much affected if δ is only slightly smaller than $\frac{1}{2}$.

5. EXPERIMENTAL AND EMPIRICAL EVIDENCE

5.1. *Monte Carlo Simulations*

To evaluate the finite sample properties of our proposal for fractional cointegration testing we have simulated several configurations of the following bivariate DGP,

$$(39) \qquad y_t = \zeta x_t + \xi_t,$$

where, for $t = 1, \dots, n$,

$$(40) \qquad \begin{pmatrix} \xi_t \\ x_t \end{pmatrix} = \begin{pmatrix} (1-L)^{-\delta} & 0 \\ 0 & (1-L)^{-d} \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \mathbb{1}_{t>0}(t)$$

so that $\Delta^d y_t = \zeta u_{2t} + \Delta^{d-\delta} u_{1t} := v_{1t}$, and

$$(41) \qquad \begin{pmatrix} (1-\phi L)u_{1t} \\ u_{2t} \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix} \right), \qquad \gamma = \frac{\omega_{12}}{\omega_1 \omega_2}.$$

Assumption (41) on the dynamics of $(u_{1t}, u_{2t})'$ is not very restrictive in our fractional framework, since many economic time series seem to be well represented as fractional white noise or finite autoregressive processes. The regression equation is

$$(42) \qquad y_t = \widehat{\alpha} + \widehat{\beta} x_t + \widehat{\xi}_t.$$

The spurious case comes out by setting $\zeta = 0$ (so that $y_t \equiv \xi_t$ and $\rho = \gamma$), $\delta = d$, and $|\rho| < 1$. As for the alternative of fractional cointegration, we set $\zeta = 1$ and $\delta = d - \frac{1}{4}$. We also normalize $\omega_1 = 1$, so that ω_2 stands for the signal-to-noise ratio. The parameter γ controls the value of ρ under H_0 and the endogeneity of regressors under H_1. The autoregressive parameter ϕ introduces some flexibility in the short properties of u_{1t}. The complete experimental design is given by

$$d = \{.7, 1, 1.3\}, \quad \gamma = \{0, .5, .8\}, \quad \omega_2 = \{.5, 1, 2\}, \quad \phi = \{0, .3, .6\}.$$

Since the estimation of $\widehat{d}_M$, $\widehat{\delta}_M$, and $\widehat{\beta}_0$ is invariant to ω_2 under the null of no cointegration, we fixed $\omega_2 = 1$ when evaluating the empirical size of $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$. The number of replications is 100,000 under H_0 and half of this number under H_1, and the nominal size is 5%. For each combination and replication we obtain first x_t and ξ_t, and then y_t, $t = 1, 2, \ldots, n$, by means of (39). We report the results for two sample sizes $n = 100$ and 200 (400 is also used for size simulations) and three values of $N = n^{.5}, n^{.6}, n^{.7}$ for the computation of $\widehat{\beta}_0^N = \widehat{\beta}_0^N(\widehat{d}_M, \widehat{\delta}_M)$. The choice $N = n^{.5}$ is important because for slower growing N than $n^{.5}$ our Wald test is asymptotically more efficient than the Z_α test, but finite sample performance of semiparametric estimates could deteriorate if N is too small. For the LS estimates $\widehat{\beta}$ and $\widehat{V}$ we always used full band OLS regressions, because the differences when using a narrow band (e.g., those defined by the bandwidths N proposed for $\widehat{\beta}_0$) were negligible, as predicted by our asymptotic theory.

The bandwidth M used for the computation of $(\widehat{d}_M, \widehat{\delta}_M)$ was fixed to $M = n^{2/3}$ for all cases and choices of N, because this choice turned out to be secondary in terms of performance of $\widehat{\beta}_0$ and $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$. Since the optimal choice of M depends on the smoothness of the spectral density of u_t and on the degree of cointegration, we used this value which is the best under our model (as far as $\delta > \frac{1}{2}$) and is also optimal under standard $\mathrm{I}(1)/\mathrm{I}(0)$ cointegration.

Following expression (42), we first regress y_t on a constant and x_t, obtaining $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\xi}_t = y_t - \widehat{\alpha} - \widehat{\beta} x_t$, and $\widehat{V}$ using (25). Then, we estimate the memory parameter of x_t. Because it is a nonstationary process, we first-difference x_t prior to d estimation, and then add unity. For the computation of $\widehat{d}_M$ we use narrow band Gaussian estimates minimizing the univariate version of (31) based on Δx_t, and we employ the same strategy for $\widehat{\delta}_M$, this time using the differenced LS residuals $\Delta\widehat{\xi}_t$. We report results for joint semiparametric estimation of (d, δ) by means of a second step (33) starting from the previous univariate Gaussian estimates, all with the same $M = n^{2/3}$. This method improved finite sample performance for high values of γ, where correlation among the series can be used successfully in joint memory estimation. Finally, we compute $\widehat{\beta}_0^N$ and $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ as in Section 4 and compare the observed value with the corresponding critical value of $\mathcal{W}_\infty$.

The distribution of $\mathcal{W}_\infty = \mathcal{W}_\infty(d, m, m_q)$ has been simulated for $d = .6(.1)1.4$, $m = 1, 2, 3$, and $m_q = 0, 1, 2$ using independent Gaussian series for z_t of length 1,000. The results of 100,000 replications are summarized in Table I by means of the coefficients of polynomial OLS regressions of (the fourth root of) the 1%, 5%, and 10% sample quantiles for each value of d on $(1, d, d^2, d^3)$. This particular transformation was used to obtain approximately homoskedastic errors in the polynomial regression. This table can be used with great precision to obtain critical values for any value of $d \in (\frac{1}{2}, \frac{3}{2})$, replacing d by $\widehat{d}_M$.

Empirical size and power results are summarized in Tables II and III, respectively.

TABLE Ia

ASYMPTOTIC CRITICAL VALUES OF $\mathcal{W}_\infty$, $\alpha = 1\%$

m		const.	d	d^2	d^3
1	no constant or trend	−4.377	15.169	−6.871	1.136
	constant	1.624	−2.402	5.932	−1.945
	trend	1.495	−.465	1.945	−.391
2	no constant or trend	−4.852	15.874	−6.880	1.597
	constant	1.239	−1.813	5.525	−1.517
	trend	1.308	−.268	1.510	.134
3	no constant or trend	−5.737	18.609	−9.609	3.088
	constant	.981	−.856	3.875	−.292
	trend	1.396	−.916	2.193	.183

TABLE Ib

ASYMPTOTIC CRITICAL VALUES OF $\mathcal{W}_\infty$, $\alpha = 5\%$

m		const.	d	d^2	d^3
1	no constant or trend	−.798	3.931	1.879	−1.317
	constant	1.824	−2.518	4.371	−1.256
	trend	1.388	−.271	1.169	−.250
2	no constant or trend	−1.571	5.614	1.119	−.841
	constant	1.703	−2.794	5.094	−1.345
	trend	1.373	−.662	1.634	−.198
3	no constant or trend	−2.135	6.940	.367	−.283
	constant	1.240	−1.336	3.333	−.361
	trend	1.196	−.230	1.062	.250

TABLE Ic

ASYMPTOTIC CRITICAL VALUES OF $\mathcal{W}_\infty$, $\alpha = 10\%$

m		const.	d	d^2	d^3
1	no constant or trend	.430	.303	4.163	−1.818
	constant	1.600	−1.761	3.074	−.831
	trend	1.148	.326	.332	−.017
2	no constant or trend	−.459	2.379	3.075	−1.330
	constant	1.647	−2.388	4.037	−.985
	trend	1.141	.020	.737	.012
3	no constant or trend	−.981	3.533	2.593	−.951
	constant	1.349	−1.559	3.128	−.402
	trend	1.148	−.134	.830	.186

Notes: The numbers given in each line of the table are the coefficients of an OLS regression of the fourth root of simulated quantiles of the F-statistic on $(1, d, d^2, d^3)$, $d = .6(.1)1.4$, under the three alternative specifications, with original data, demeaned data, and linear detrended data. In all cases, 100,000 simulations of $m + 1$ independent Gaussian series of length 1,000 with zero mean and unit variance are fractionally integrated and then full band statistics are computed.

TABLE II

EMPIRICAL SIZE OF 5% $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ TESTS

d	ρ	ϕ	N $(n=100)$ $n^{.5}$	$n^{.6}$	$n^{.7}$	N $(n=200)$ $n^{.5}$	$n^{.6}$	$n^{.7}$	N $(n=400)$ $n^{.5}$	$n^{.6}$	$n^{.7}$
.7	.0	.0	6.08	5.24	4.86	5.55	4.85	4.81	5.03	4.82	4.81
		.3	5.23	4.81	4.72	4.95	4.70	4.75	4.79	4.74	4.66
		.6	3.92	3.86	3.85	4.01	4.06	4.06	4.24	4.30	4.15
	.5	.0	10.13	7.13	5.33	10.54	7.17	5.26	10.15	7.27	5.20
		.3	7.04	5.47	5.12	7.43	5.63	5.01	7.54	5.79	4.88
		.6	4.22	4.24	5.15	4.48	4.71	6.03	4.63	4.88	6.72
	.8	.0	19.22	11.83	7.12	21.62	13.10	6.50	22.00	14.22	6.89
		.3	12.53	7.54	6.91	15.06	8.57	5.94	16.59	9.93	5.79
		.6	5.56	5.42	10.39	7.24	5.34	12.64	9.61	4.99	13.31
1.0	.0	.0	6.16	5.83	5.72	5.51	5.36	5.41	5.17	5.15	5.26
		.3	5.33	5.24	5.18	5.06	5.04	5.11	4.89	4.94	5.02
		.6	3.72	3.71	3.70	3.89	3.98	3.98	4.10	4.16	4.17
	.5	.0	7.71	6.52	5.87	7.17	6.04	5.48	6.43	5.63	5.21
		.3	5.92	5.40	5.38	5.59	5.13	5.04	5.14	4.90	4.87
		.6	3.69	3.97	4.62	3.56	4.01	4.69	3.38	3.80	4.61
	.8	.0	11.74	8.38	6.52	11.86	8.13	5.90	10.63	7.64	5.63
		.3	7.71	5.93	6.25	8.12	5.70	5.38	7.17	5.26	4.83
		.6	3.86	4.63	7.84	3.85	3.92	7.51	3.43	2.77	6.55
1.3	.0	.0	5.15	5.04	5.08	4.96	4.90	4.97	4.68	4.74	4.79
		.3	4.52	4.55	4.59	4.62	4.58	4.68	4.43	4.52	4.58
		.6	3.36	3.41	3.47	3.61	3.74	3.76	3.74	3.88	3.88
	.5	.0	6.04	5.38	5.04	5.89	5.23	5.05	5.50	5.07	4.82
		.3	4.73	4.53	4.63	4.81	4.60	4.70	4.57	4.51	4.50
		.6	3.44	3.76	4.37	3.38	3.81	4.36	3.21	3.62	4.26
	.8	.0	8.59	6.68	5.55	8.64	6.60	5.35	7.98	6.28	5.23
		.3	5.86	4.96	5.41	6.06	4.88	4.91	5.53	4.61	4.61
		.6	3.46	4.47	7.20	3.21	3.80	6.60	2.72	2.76	5.67

Overall, from Table II we learn that $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ has good size properties in small to medium sample sizes and even for moderate values of ρ. This is specially so in the mean-averting case where $d \geq 1$, the asymptotic approximation improving with sample size. The lower choice of N leads in general to overrejection of H_0 for the largest ρ (and $d < 1$) but the other two achieve more conservative tests, except when $\phi = .6$, where the situation is reversed. As regards power, it follows from Table III that the higher the signal-to-noise ratio ω_2 and the larger the correlation γ, the higher the power of the statistic $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$, as expected. As predicted by the results in Section 4, the power is in general quite large for the sample sizes considered and increases the smaller N. We only report results for the two smaller values of ϕ, but Table III makes clear that a

TABLE III

EMPIRICAL POWER OF 5% $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ TESTS

			$\phi = 0$						$\phi = .3$					
			N ($n = 100$)			N ($n = 200$)			N ($n = 100$)			N ($n = 200$)		
d	γ	ω_2	$n^{.5}$	$n^{.6}$	$n^{.7}$	$n^{.5}$	$n^{.6}$	$n^{.7}$	$n^{.5}$	$n^{.6}$	$n^{.7}$	$n^{.5}$	$n^{.6}$	$n^{.7}$
.7	.0	.5	31.5	20.4	8.2	51.1	34.2	15.1	16.6	10.1	4.8	29.4	17.3	7.1
		1.0	49.6	43.8	13.9	70.8	56.4	28.9	28.7	18.0	6.9	49.2	33.0	13.8
		2.0	67.2	52.2	22.6	84.4	74.4	47.4	45.9	30.9	11.7	67.8	53.3	26.0
	.5	.5	51.2	38.8	19.0	73.9	63.6	42.1	33.1	23.3	10.5	55.9	44.2	24.4
		1.0	63.1	50.2	24.4	82.9	75.0	54.0	42.8	30.8	13.3	67.4	56.1	33.0
		2.0	74.8	62.8	32.0	89.5	84.2	66.6	55.7	41.2	18.6	78.2	66.6	44.7
	.8	.5	66.5	54.4	29.6	88.3	82.7	65.6	45.8	34.9	34.5	72.0	63.8	42.0
		1.0	74.6	63.4	33.6	92.1	88.3	73.7	53.5	41.2	20.7	78.2	70.7	48.3
		2.0	82.5	72.7	38.9	94.9	92.4	81.7	63.7	50.0	25.0	84.4	77.9	56.6
1.0	.0	.5	22.6	14.7	7.5	35.0	21.4	10.0	11.3	7.6	4.7	17.9	10.2	5.2
		1.0	37.2	25.1	11.4	56.0	38.6	18.7	19.0	11.8	6.0	32.1	19.2	8.6
		2.0	55.0	40.0	18.0	74.8	59.6	33.1	31.5	20.2	8.6	51.2	34.8	15.8
	.5	.5	39.0	28.6	15.7	58.6	44.7	28.1	22.7	16.1	9.2	38.4	28.0	17.0
		1.0	51.0	38.4	20.1	71.9	58.5	37.4	30.0	21.1	10.1	50.0	37.2	21.8
		2.0	64.9	51.4	26.3	83.7	73.2	50.7	41.1	29.2	13.7	63.6	50.4	29.3
	.8	.5	54.9	42.6	25.4	78.2	66.0	46.6	33.8	26.2	16.1	56.5	45.6	30.7
		1.0	65.1	51.7	28.6	86.6	76.3	54.8	40.6	30.9	17.5	65.8	53.8	35.0
		2.0	76.3	63.3	35.6	92.9	86.3	66.1	50.4	38.4	19.8	75.9	61.9	45.0
1.3	.0	.5	16.6	11.0	6.3	24.8	15.1	7.6	8.4	6.0	4.2	12.8	7.6	4.5
		1.0	27.0	18.2	9.0	40.9	26.9	13.5	13.5	8.7	5.0	22.4	13.4	6.6
		2.0	41.7	29.4	14.0	59.3	43.8	23.9	22.2	14.2	6.7	37.1	24.1	11.3
	.5	.5	28.9	21.4	12.9	43.2	32.0	21.1	16.5	12.2	7.7	27.7	20.5	13.4
		1.0	38.6	28.4	16.0	55.8	42.8	27.6	21.8	15.7	8.9	36.5	26.7	16.5
		2.0	51.1	39.0	21.0	69.7	56.7	37.7	30.0	21.3	10.9	48.6	36.7	21.9
	.8	.5	43.0	33.5	21.5	62.6	50.2	36.4	25.8	20.7	13.6	43.4	35.0	24.8
		1.0	52.4	40.9	24.3	73.0	60.3	42.7	31.0	24.0	14.6	51.3	41.2	27.8
		2.0	64.2	51.3	28.5	83.4	72.4	52.3	39.0	29.3	16.2	62.1	50.2	32.9

large ϕ makes detection of cointegration much more difficult, since these values increase the finite sample persistence of ξ_t and, therefore, of the observed residuals $\widehat{\xi}_t$.

For completeness, we have also studied the finite sample behavior of two well-known classical residual-based tests, namely, the ADF test, which amounts to calculating the OLS t-test of the null hypothesis $a = 0$ in the regression

$$\Delta\widehat{\xi}_t = \widehat{a}\widehat{\xi}_t + \sum_{i=1}^{k} \widehat{\varphi}_i \Delta\widehat{\xi}_{t-i} + errors, \tag{43}$$

and the Z_α test,

$$(44) \qquad Z_\alpha = n(\widehat{a}-1) - \left(n^{-1}\sum_{s=1}^{l} w_{sl} \sum_{t=s+1}^{n} \widehat{b}_t \widehat{b}_{t-s}\right)\left(n^{-2}\sum_{t=2}^{n} \widehat{\xi}_{t-1}^2\right)^{-1},$$

where $w_{sl} = 1 - s/(l+1)$ and $\widehat{b}_t = \widehat{\xi}_t - \widehat{a}\widehat{\xi}_{t-1}$. These tests, although not originally designed to test against fractional cointegration, are consistent against fractionally cointegrated alternatives. See Krämer and Marmol (2004). The order k in the ADF regression was determined using the Akaike information criterion, whilst the lag truncation parameter, l, of the Z_α test was determined by means of Andrews' (1991) data-dependent formula assuming that $\Delta\widehat{\xi}_t - \overline{\Delta\widehat{\xi}}$ follows an AR(1) process. Both procedures are fairly stable with respect to ω_2, so in Tables IV and V we only present the results for $\omega_2 = 1$.

It turns out that the Z_α test has poor size properties except in the unit root $d = 1$ case, displaying vast rejection percentages if $d < 1$ and almost zero rejection percentages if $d > 1$. See Table IV. In terms of power, note that even in the case $d = 1$, when $\phi = .3$ the Z_α test has almost no power and our Wald test compares favorably for many parameter combinations. On the other hand, the ADF test exhibits only moderate size distortions, but it has low power properties. See Table V. Our study, then, confirms previous findings of Gonzalo and Lee (1998) and Dittmann (2000) on the lack of reliability of classical cointegration tests in the presence of fractionally integrated error terms. By contrast, our test has good size properties and shows comparable power to that of ADF and Z_α tests in spite of its semiparametric nature and the fact that, in contrast with ADF and Z_α, we treat d and δ as nuisance parameters to be estimated with only moderate efficiency losses in the particular unit root ($d = 1$) case.

TABLE IV

EMPIRICAL SIZE OF 5% RESIDUAL-BASED TESTS

		$\phi = 0$				$\phi = .3$			
		$n = 100$		$n = 200$		$n = 100$		$n = 200$	
d	ρ	ADF	Z_α	ADF	Z_α	ADF	Z_α	ADF	Z_α
.7	.0	8.7	62.3	13.0	81.3	8.3	27.7	12.2	53.3
	.5	8.5	61.2	12.7	80.8	8.1	29.8	11.9	55.4
	.8	8.1	60.4	12.1	79.9	7.8	40.5	12.2	53.3
1.0	.0	5.8	5.1	5.3	5.4	5.7	1.9	5.3	2.6
	.5	5.6	4.8	5.2	5.2	5.7	2.1	5.1	2.7
	.8	5.6	4.7	5.0	5.0	5.6	3.6	5.0	4.0
1.3	.0	5.3	.2	3.3	.1	5.3	.1	3.4	.1
	.5	5.1	.2	3.3	.1	5.2	.1	3.4	.1
	.8	5.1	.2	3.3	.1	5.3	.2	3.3	.1

TABLE V

EMPIRICAL POWER OF 5% RESIDUAL-BASED TESTS

		$\phi=0$				$\phi=.3$			
		$n=100$		$n=200$		$n=100$		$n=200$	
d	γ	ADF	Z_α	ADF	Z_α	ADF	Z_α	ADF	Z_α
.7	.0	17.4	99.7	37.4	100.0	15.6	88.7	33.9	99.7
	.5	15.6	99.6	35.1	100.0	14.4	86.6	31.5	99.7
	.8	11.3	99.6	27.9	100.0	10.9	80.1	24.3	94.4
1.0	.0	9.5	46.2	12.1	64.4	9.1	14.8	11.6	31.1
	.5	9.2	47.8	12.1	66.9	8.8	14.0	11.3	31.1
	.8	8.3	52.9	11.4	73.5	7.7	12.5	10.2	31.4
1.3	.0	6.7	2.3	5.1	1.9	6.7	.7	5.2	.9
	.5	6.5	2.6	5.0	2.3	6.4	.7	5.0	1.0
	.8	6.2	3.7	4.9	3.7	6.1	.7	4.9	1.2

5.2. *Empirical Evidence*

In order to judge the empirical applicability of our testing procedure, herein we are concerned with testing one of the most important implications of the quantitative theory of money, namely, the stability of the (*inverse of*) velocity of circulation of money (the ratio of money stock to nominal income). The dynamic properties behavior of the velocity of money has attracted a great deal of attention in the literature given its implications for the monetarist position. The justification for the stable velocity of circulation is that people wish to hold as little cash as possible in order to carry out their transactions. The velocity of circulation of money will therefore be as fast as the existing money-holding-technology permits and therefore relatively stable. Our empirical work employs the data set of Engle and Granger (1987). All data series are seasonally adjusted, quarterly observations of (*the log of*) U.S. simple-sum ($M1$, $M2$, and $M3$) monetary aggregates and (*the log of*) GNP, covering the period 1959:1 to 1981:2 ($n=90$). We also considered the extended sample period 1959:1 to 2000:1 and the fact that financial innovations of the 1970's and the deregulation of the early 1980's might have affected the low-frequency properties of the series, obtaining similar results which are not reported for the sake of brevity.

In their seminal paper, Engle and Granger tested whether velocity is stationary by means of the ADF test assuming the standard $\mathrm{I}(1)/\mathrm{I}(0)$ cointegration setup. Only for $M2$ was the test significant (at the 5% level). For the other aggregates they reject cointegration and the stationarity of velocity. In Table VI we present the results of applying our general testing strategy to Engle and Granger's data set. In order to judge robustness of our procedure to the bandwidth N, we chose the same grid of values as in the simulation, namely, $N=\{n^{.5}, n^{.6}, n^{.7}\}=\{9,14,23\}$. These values provide a reasonable balance between size and power, according to our Monte Carlo experiments and avoid

TABLE VI

(LOG) MONEY VERSUS (LOG) GNP, 1959:1–1981:2

M	$\widehat{d}_M$	$\widehat{\delta}_M$	$\widehat{\beta}$	N	$\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$
			$M1$		
20	1.386	.949	.630	9	2610.44*
				14	600.38*
				23	197.45*
			$M2$		
20	1.378	1.176	.966	9	357.16*
				14	763.42*
				23	1714.42*
			$M3$		
20	1.366	1.133	1.083	9	128.98*
				14	672.35*
				23	2064.84*

Notes: The superscript * indicates statistical significance for the null hypothesis of no cointegration at the 1 percent level.

higher frequencies that could be contaminated by seasonal or other cyclical behaviors. M was fixed to $n^{2/3}$.

Our estimates $\widehat{d}_M$, $\widehat{\delta}_M$, and $\widehat{\beta}$ are broadly comparable to those obtained by Robinson and Marinucci (1998) for the same data set. In Table VI, $\widehat{d}_M$ stands for the Gaussian estimator of the memory parameter of log(GNP). Robinson and Marinucci failed to reject the null of equal orders of integration in a clear majority of cases for $\log(M1)/\log(\text{GNP})$, $\log(M2)/\log(\text{GNP})$, and $\log(M3)/\log(\text{GNP})$.

The integration order of log(GNP) turns out to be about 1.3 and estimates $\widehat{\delta}_M$ are strongly inconsistent with stationarity, ranging from .95 to 1.17. For these values of d and δ, our Monte Carlo experiments show that ADF and Z_α are not reliable tests, with very low power to reject the null of no cointegration. In our case we reject such a null for all three monetary aggregates at 1% level, uniformly in N.

6. CONCLUSIONS

We have developed in this paper a general theory for cointegrated and spuriously related NFI processes. We propose consistent tests of these hypotheses based on the observation that the Wald statistic of OLS coefficients, appropriately normalized, converges weakly under both hypotheses, in particular to a distribution $\mathcal{W}_\infty$, which only depends on d under the null. Our test exploits the different properties of OLS coefficients and our normalization $\widehat{V}$ based on OLS residuals under the two hypotheses, by means of the comparison of the OLSE $\widehat{\beta}$ with a further estimate of the projection vector β_0. Our proposal is to use an estimate $\widehat{\beta}_0$ that diverges for cointegrated series for which $\widehat{\beta}$ is consistent, so we can improve the divergence rate of the adjusted Wald statistic. However both aspects depend on the distance $d-\delta$, reflecting the fact that closer alternatives are much more difficult to detect for a given sample size. This also shows that, although the null hypothesis is composite when parameterized by means of the parameter ρ ($\rho^2<1$), the alternative is also composite when stated as $\delta<d$ vs. the simple null of $\delta=d$.

Our tests are intuitive and simple to implement, and show good power properties, even when compared to unit root cointegration tests, which assume the information on the degree of integration of the original series. These tests only require the choice of a narrow band of frequencies that contains the relevant information about the long run dynamics of the time series vector considered. This feature is common to most devices used in econometrics to correct for autocorrelation of unknown form. However the semiparametric character of the methodology not only guarantees the suitability of our tests for general processes without resort to difficult-to-justify parametric or distributional assumptions, but hopefully provides desirable power properties in applications.

Dept. of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain; francesc.marmol@uc3m.es

and

Dept. of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain; Carlos.Velasco@uc3m.es.

Manuscript received October, 2003; final revision received February, 2004.

APPENDIX: MATHEMATICAL PROOFS

PROOF OF LEMMA 1: Assuming that n is odd (otherwise replace the lower limit in the sums by $1-[n/2]$) we have that

$$\begin{pmatrix}\widehat{\alpha}_M\\ \widehat{\beta}_M\end{pmatrix}=\left(\sum_{j=-[n/2]}^{[n/2]} I_{gg}(\lambda_j)-2\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{gg}(\lambda_j)\right)^{-1}$$

$$\times\left(\sum_{j=-[n/2]}^{[n/2]} I_{gy}(\lambda_j)-2\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{gy}(\lambda_j)\right),$$

where Re stands for real part. Now, as $n\to\infty$,

$$\text{(A.1)}\qquad \Psi_n^{-1}\sum_{j=-[n/2]}^{[n/2]} I_{gg}(\lambda_j)\Psi_n^{-1}=\Psi_n^{-1}\sum_{j=1}^{n} I_{gg}(\lambda_j)\Psi_n^{-1}=\frac{1}{4\pi}\Psi_n^{-1}\sum_{t=1}^{n} g_t g_t'\Psi_n^{-1}\Rightarrow\frac{1}{4\pi}A,$$

$$n^{-d}\Psi_n^{-1}\sum_{j=-[n/2]}^{[n/2]} I_{gy}(\lambda_j)=\frac{n^{-d}}{4\pi}\Psi_n^{-1}\sum_{t=1}^{n} g_t y_t\Rightarrow\frac{1}{4\pi}B,$$

where $A=\int_0^1 Q(r)Q(r)'\,dr$ and $B=\int_0^1 Q(r)B_y(d;r)\,dr$ are nondegenerated random variables.

From the proof of Proposition 4.1 of Robinson and Marinucci (1998) we have that $EI_{zz}(\lambda_j)=O(|\lambda_j|^{-2d})$ if $\frac{1}{2}<d<1$ and $O(|\lambda_j|^{-2d}|j|^{2(d-1)})$ if $d\geq 1$, so that, $|j|<n/2$,

$$E\|I_{xx}(\lambda_j)\|=O\big(n^{2d}|j|^{-2\min\{d,1\}}\big) \tag{A.2}$$

as $\lambda_j\to 0$, using Cauchy–Schwarz for each element in $I_{xx}(\lambda_j)$, implying that

$$\begin{aligned} E\left\|\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{xx}(\lambda_j)\right\| &\leq \sum_{j=M+1}^{[n/2]} E\|I_{xx}(\lambda_j)\| \\ &= O\left(\sum_{j=M+1}^{n} n^{2d}|j|^{-2\min\{d,1\}}\right)=O\big(n^{2d}M^{1-2\min\{d,1\}}\big) \end{aligned}$$

as $M\to\infty$. Then using that, for $a=1,2,\ldots$, $|j|<n/2$,

$$I_{t^a t^a}(\lambda_j)=O(|\lambda_j|^{-2}n^{2a-1})=O(n^{2a+1}|j|^{-2})$$

(see, e.g., Robinson and Marinucci (2001)), so for $\psi_n=\operatorname{diag}(n^{1/2},\ldots,n^{m_f-1/2})$,

$$\left\|\psi_n^{-1}\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{qq}(\lambda_j)\psi_n^{-1}\right\| \leq \sum_{j=M+1}^{[n/2]}\|\psi_n^{-1}I_{qq}(\lambda_j)\psi_n^{-1}\|=O\left(\sum_{j=M+1}^{n}|j|^{-2}\right)=O(M^{-1}),$$

and proceeding in a similar way for the cross terms, we obtain that

$$\begin{aligned} E\left\|\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{qy}(\lambda_j)\right\| &\leq \sum_{j=M+1}^{[n/2]}\big(n^{-2d}EI_{yy}(\lambda_j)\|\psi_n^{-1}I_{qq}(\lambda_j)\psi_n^{-1}\|\big)^{1/2} \\ &= O\left(\sum_{j=M+1}^{[n/2]}|j|^{-\min\{d,1\}-1}\right) \\ &= O\big(M^{-\min\{d,1\}}\big)=O(M^{-d}+M^{-1}). \end{aligned}$$

Then finally,

$$\Psi_n^{-1}\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{gg}(\lambda_j)\Psi_n^{-1}=O_p\big(M^{1-2\min\{d,1\}}+M^{-d}+M^{-1}\big)=O_p(M^{1-2d}+M^{-1}),$$

$$n^{-d}\Psi_n^{-1}\sum_{j=M+1}^{[n/2]}\operatorname{Re} I_{gy}(\lambda_j)=O_p\big(M^{1-2\min\{d,1\}}+M^{-d}+M^{-1}\big)=O_p(M^{1-2d}+M^{-1}).$$

Collecting all the previous results, it follows that, as $M\to\infty$,

$$\begin{aligned} \begin{pmatrix}\widehat{\alpha}_M\\ \widehat{\beta}_M\end{pmatrix} &= \left(\sum_{j=-[n/2]}^{[n/2]} I_{gg}(\lambda_j)\right)^{-1}[1+O_p(M^{1-2d}+M^{-1})] \\ &\quad\times\left(\sum_{j=-[n/2]}^{[n/2]} I_{gy}(\lambda_j)\right)[1+O_p(M^{1-2d}+M^{-1})] \\ &= \begin{pmatrix}\widehat{\alpha}_{[n/2]}\\ \widehat{\beta}_{[n/2]}\end{pmatrix}[1+O_p(M^{1-2d}+M^{-1})]. \end{aligned}$$

Now we can write

$$I_{\widehat{\xi}\widehat{\xi}}(\lambda_j) = I_{yy}(\lambda_j) - 2\widehat{\beta}_M' \mathrm{Re}[I_{x_q y}(\lambda_j)] + \widehat{\beta}_M' I_{x_q x_q}(\lambda_j)\widehat{\beta}_M.$$

Then we can obtain that, as $M \to \infty$, $M \leq n/2$, $\widehat{V}_M = \widehat{V}_{[n/2]} + o_p(1)$, because taking expectations of pairs of (cross) periodograms,

$$\sum_{j=M}^{[n/2]} \mathrm{Re}\, I_{x_q x_q}(\lambda_j)\big\{I_{yy}(\lambda_j) - 2\widehat{\beta}_M' \mathrm{Re}[I_{x_q y}(\lambda_j)] + \widehat{\beta}_M' I_{x_q x_q}(\lambda_j)\widehat{\beta}_M\big\}$$

$$= O_p\left(n^{4d} \sum_{j=M}^{[n/2]} j^{-4\min\{1,d\}}\right) = O_p\big(n^{4d} M^{1-4\min\{1,d\}}\big) = o_p(n^{4d}),$$

noting that by our Definition 5.1 and Theorem 5.1 of Robinson and Marinucci (2001) we can control the fourth-order cumulant contribution to those expectations, and that our NFI processes satisfy such definition, since u_t is a well-behaved linear process with i.i.d. innovations and smooth higher-order spectral densities. Then, the same result holds after detrending, replacing I_{xx} by $I_{x_q x_q}$, etc., using similar methods. *Q.E.D.*

PROOF OF THEOREM 1: We can write using standard equalities that

$$\widehat{V}_{[n/2]} = \left(\sum_{j=-[n/2]}^{[n/2]} I_{x_q x_q}(\lambda_j)\right)^{-1} \sum_{j=-[n/2]}^{[n/2]} I_{x_q x_q}(\lambda_j) I_{\widehat{\xi}\widehat{\xi}}(\lambda_j) \left(\sum_{j=-[n/2]}^{[n/2]} I_{x_q x_q}(\lambda_j)\right)^{-1}$$

$$= n\left(\sum_{t=1}^{n} x_{t,q} x_{t,q}'\right)^{-1}$$

$$\times \left(\sum_{j=0}^{n-1} \mathbb{I}_0(j)\widehat{\gamma}_{\widehat{\xi}\widehat{\xi}}(j)\big\{\widehat{\boldsymbol{\gamma}}_{x_q x_q}(j) + \widehat{\boldsymbol{\gamma}}_{x_q x_q}(j)' + \widehat{\boldsymbol{\gamma}}_{x_q x_q}(n-j) + \widehat{\boldsymbol{\gamma}}_{x_q x_q}(n-j)'\big\}\right)$$

$$\times \left(\sum_{t=1}^{n} x_{t,q} x_{t,q}\right)^{-1},$$

where $\mathbb{I}_0(0) = \frac{1}{2}$, $\mathbb{I}_0(j) = 1$, $j \neq 0$, $\widehat{\gamma}_{\widehat{\xi}\widehat{\xi}}(j) = n^{-1}\sum_{t=1}^{n-|j|} \widehat{\xi}_t \widehat{\xi}_{t+j}$, and $\widehat{\boldsymbol{\gamma}}_{x_q x_q}(j) = \frac{1}{n}\sum_{t=1}^{n-|j|} x_{t,q} x_{t+|j|,q}'$. Using the CMT we have that

$$n\Psi_n^{-1}\widehat{\boldsymbol{\gamma}}_{x_q x_q}([ns])\Psi_n^{-1} \Rightarrow \Omega_{xx}^{1/2}\boldsymbol{\gamma}_{x_q x_q}(s)\Omega_{xx}^{1/2}$$

and

$$n^{1-2d}\widehat{\gamma}_{\widehat{\xi}\widehat{\xi}}([ns]) \Rightarrow \omega_{yy}(1-\rho^2)^{1/2}\gamma_R(s).$$

Thus, a direct application of the CMT yields

$$n^{-4d} \sum_{t=1}^{n} x_{t,q} x_{t,q}' \widehat{V}_N \sum_{t=1}^{n} x_{t,q} x_{t,q}' \Rightarrow V.$$

The asymptotic distribution of $\mathcal{W}(\widehat{\beta}_M, \beta_0)$ follows from (22), Lemma 1, and the CMT. *Q.E.D.*

PROOF OF LEMMA 2: Following Robinson (1995a, Theorem 1) and Velasco (2003, Theorem 1) we consider $r_t := \Delta x_t - \Delta s_t$ as a residual approximation and then use the methods developed in these references. We consider for simplicity of notation that x_t is scalar and bound first the expectation of the periodogram of r_t at Fourier frequencies by Theorems 1 and 5 of Velasco (2002). We have that

$$I_{rr}(\lambda) = I_{\Delta s \Delta s}(\lambda) - 2\,\mathrm{Re}\, I_{\Delta x \Delta s}(\lambda) + I_{\Delta x \Delta x}(\lambda), \tag{A.4}$$

while by Theorem 1 of Robinson (1995b), $j = 1, \ldots, M$, $M/n \to 0$ as $n \to \infty$,

$$E[I_{\Delta s \Delta s}(\lambda_j)] = H_{\Delta x \Delta x}(\lambda_j)\{1 + O(j^{-1}\log n)\}$$

and, on the other hand, by Velasco (2002, Theorems 1 and 5, for $d > 1$ and $d < 1$, respectively),

$$E[I_{\Delta x \Delta x}(\lambda_j)] = H_{\Delta x \Delta x}(\lambda_j)\{1 + O(\alpha_j)\}$$

where $\alpha_j := j^{-1}\log n + j^{\gamma-1} + n^{-1}[j^{2\gamma} + \lambda_j^{2\gamma}]$, and $\gamma = d - 1 \in (-\frac{1}{2}, \frac{1}{2})$. Using the same methods, it is straightforward to obtain that $E[I_{\Delta x \Delta s}(\lambda_j)] = H_{\Delta x \Delta x}(\lambda_j)\{1 + O(\alpha_j)\}$, so that, using (A.4),

$$E[I_{rr}(\lambda_j)] = O(H_{\Delta x \Delta x}(\lambda_j)\alpha_j),$$

and $E[I_{r\Delta s}(\lambda_j)] = O(H_{\Delta x \Delta x}(\lambda_j)\alpha_j)$ using the fact that $I_{r\Delta s}(\lambda) = I_{\Delta x \Delta s}(\lambda) - I_{\Delta s \Delta s}(\lambda)$ and cancelling the leading terms of each expectation.

Now, set $\nabla = \nabla_1 - 1$ when $d < \frac{1}{2} + \nabla_1$ and $\gamma \geq \nabla > \gamma - \frac{1}{2}$ otherwise, as in Robinson (1995a). Then the lemma follows from Robinson (1997, Theorem 3) if

$$\sum_{n=1}^{M-1}\left(\frac{n}{M}\right)^{2(\nabla-\gamma)+1}\frac{1}{n^2}\left|\sum_{j=1}^{n} R_j\right| \xrightarrow{p} 0, \tag{A.5}$$

where $R_j := [I_{\Delta x \Delta x}(\lambda_j) - I_{\Delta s \Delta s}(\lambda_j)]/h(\lambda_j) = [-2\,\mathrm{Re}\, I_{r\Delta s}(\lambda_j) + I_{rr}(\lambda_j)]/h(\lambda_j)$, $h(\lambda_j) := |\lambda_j|^{2\gamma}$, and additionally, for $\psi > 0$ arbitrarily small,

$$\log^2 n \sum_{n=1}^{M-1}\left(\frac{n}{M}\right)^{1-2\psi}\frac{1}{n^2}\left|\sum_{j=1}^{n} R_j\right| \xrightarrow{p} 0 \tag{A.6}$$

and

$$\frac{\log^2 n}{M}\sum_{j=1}^{n} R_j \xrightarrow{p} 0. \tag{A.7}$$

When $d \geq \frac{1}{2} + \nabla_1$ we also need to show that

$$\frac{1}{M}\sum_{j=1}^{M}(a_j - 1)R_j \xrightarrow{p} 0, \tag{A.8}$$

with $a_j = (j/h)^{2(\nabla-\gamma)}$, $1 \leq j \leq h$, and $a_j = (j/h)^{2(\nabla_1-\gamma+1)}$, $h < j \leq M$, $h = \exp(M^{-1}\sum_1^M \log j)$.

As regards (A.5), notice that, for a generic positive constant $K < \infty$, the left-hand side of (A.5) is bounded by

$$KM^{2(\gamma-\nabla)-1}\sum_{j=1}^{M} j^{2(\nabla-\gamma)}|R_j| \quad \text{for } \nabla < \gamma, \tag{A.9}$$

and by

$$K\frac{\log M}{M}\sum_{j=1}^{M}|R_j| \quad \text{for } \nabla = \gamma. \tag{A.10}$$

The left-hand side of (A.9) is bounded by

$$KM^{2(\gamma-\nabla)-1}\left\{2\sum_{j=1}^{M} j^{2(\nabla-\gamma)}\frac{|I_{r\Delta s}(\lambda_j)|}{h(\lambda_j)}+\sum_{j=1}^{M} j^{2(\nabla-\gamma)}\frac{I_{rr}(\lambda_j)}{h(\lambda_j)}\right\},$$

and, taking expectations of each term in the summands, the order of magnitude turns out to be

$$O_p\left(M^{2(\gamma-\nabla)-1}\sum_{j=1}^{M} j^{2(\nabla-\gamma)}\alpha_j\right)$$
$$=o_p(1)+O_p(M^{-1}\log M\log n)+O_p(n^{-1}\{M^{2\gamma}+M^{2\gamma}n^{-2\gamma}\}),$$

and this is $o_p(1)$ using (34). The bounds for (A.10) and (A.6) follow similarly.

The left-hand side of (A.7) is $o_p(1)$ because it is bounded by

$$\frac{\log^2 n}{M}\sum_{j=1}^{M}|R_j| \le \frac{\log^2 n}{M}\sum_{j=1}^{M}\left\{2\frac{|I_{r\Delta s}(\lambda_j)|}{h(\lambda_j)}+\frac{I_{rr}(\lambda_j)}{h(\lambda_j)}\right\}$$
$$=O_p\left(\frac{\log^2 n}{M}\sum_{j=1}^{M}\alpha_j\right)$$
$$=O_p(M^{\gamma-1}\log M+M^{-1}\log M\log n)\log^2 n$$
$$+O_p(n^{-1}\{M^{2\gamma}+M^{2\gamma}n^{-2\gamma}\}\log M)\log^2 n,$$

which tends to zero with (34).

On the other hand, using (A.7), the left-hand side of (A.8) is bounded by

$$\frac{1}{M}\sum_{j=1}^{M}a_j|R_j|+o_p(1),$$

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

$$O_p\left(M^{2(\gamma-\nabla)-1}\sum_{j=1}^{M} j^{2(\nabla-\gamma)}\alpha_j+M^{2(\gamma-\nabla)-1}\sum_{j=1}^{M} j^{2(\nabla-\gamma)}\alpha_j\right),$$

and using the same arguments and (34) we can show that this is $o_p(1)$. *Q.E.D.*

PROOF OF THEOREM 3: In case of cointegration, $\delta>\frac{1}{2}$, we obtain that $\log n(\widehat{\delta}_M-\delta)\xrightarrow{p}0$ (cf. Lemma 2 and Velasco (2003)), and using that $\log n(\widehat{d}_M-d)\xrightarrow{p}0$, this yields under H_1 when $d-\delta<\frac{1}{2}$,

$$\left(\frac{n}{N}\right)^{2(\delta-d)}\widehat{\omega}_{xy}^{N}(\widehat{\delta}_M) = \Omega_{xx}\beta_0\frac{n^{2(\delta-d)}}{N^{1+2(\delta-d)}}\sum_{j=1}^{N}\lambda_j^{2(\delta-d)}(1+o_p(1))+O_p\left(\left(\frac{n}{N}\right)^{2(\delta-d)}\right)$$
$$\xrightarrow{p}\Omega_{xx}\beta_0K(\delta,d)$$

for some constant $K(\delta,d)>0$, so

$$\left(\frac{n}{N}\right)^{2(\delta-d)}\widehat{\beta}_0^N\xrightarrow{p}\beta_0K(\delta,d),$$

i.e. $\widehat{\beta}_0^N$ diverges with n and N. It follows that, under the assumptions of Lemma 2,

$$\text{prob}\lim_{n\to\infty} \inf n^{6(\delta-d)} N^{4(d-\delta)} \mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N) > 0,$$

under H_1. When $d - \delta > \frac{1}{2}$, reasoning in the same way,

$$N n^{2(\delta-d)} \widehat{\omega}_{xy}^N(\widehat{\delta}_M) \xrightarrow{p} \Omega_{xx} \beta_0 K^*(\delta, d),$$

for some $K^*(\delta, d) > 0$, so $N n^{2(\delta-d)} \widehat{\beta}_0^N \xrightarrow{p} \beta_0 K^*(\delta, d)$ and $\mathcal{W}(\widehat{\beta}, \widehat{\beta}_0^N)$ diverges at the rate $n^{6(d-\delta)} N^{-1}$ (which tends to infinity since $Nn^{-1} \to 0$ and $6(d-\delta) > 3$) with an additional $\log^{-1} N$ term when $d - \delta = \frac{1}{2}$. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BAILLIE, R. T., AND T. BOLLERSLEV (1994): "Cointegration, Fractional Cointegration, and Exchange Rate Dynamics," *The Journal of Finance*, 49, 737–745.

BOOTH, G. G., AND Y. TSE (1995): "Long Memory in Interest Rate Future Markets: A Fractional Cointegration Analysis," *The Journal of Future Markets*, 15, 563–584.

CHEUNG, Y. E., AND K. S. LAI (1993): "A Fractional Cointegration Analysis of Purchasing Power Parity," *Journal of Business and Economic Statistics*, 11, 103–122.

DITTMANN, I. (2000): "Residual-Based Tests for Fractional Cointegration: A Monte Carlo Study," *Journal of Time Series Analysis*, 21, 615–647.

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

GIL-ALAÑA, L. A., AND P. M. ROBINSON (1997): "Testing of Unit Root and Other Nonstationary Hypotheses in Macroeconomic Time Series," *Journal of Econometrics*, 80, 241–268.

GONZALO, J., AND T. LEE (1998): "Pitfalls in Testing for Long Run Relationships," *Journal of Econometrics*, 86, 129–154.

GRANGER, C. W. J., N. HYUNG, AND Y. JEON (2001): "Spurious Regressions with Stationary Series," *Applied Economics*, 33, 899–904.

GRANGER, C. W. J., AND P. NEWBOLD (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 74, 111–120.

GRANGER, C. W. J., AND N. SWANSON (1997): "An Introduction to Stochastic Unit-Root Processes," *Journal of Econometrics*, 80, 35–62.

HALDRUP, N. (1994): "The Asymptotics of Single-Equation Cointegration Regressions with I(1) and I(2) Variables," *Journal of Econometrics*, 63, 153–181.

HASSLER, U., F. MARMOL, AND C. VELASCO (2003): "Residual Log-Periodogram Inference for Long-Run Relationships," *Journal of Econometrics*, forthcoming.

HAUSMAN, J. (1978): "Misspecification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

JANSSON, M., AND N. HALDRUP (2002): "Regression Theory for Nearly Cointegrated Time Series," *Econometric Theory*, 18, 1309–1335.

JOHANSEN, S. (1988): "Statistical Analysis and Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

——— (1992): "Cointegration in Partial Systems and the Efficiency of Single Equation Analysis," *Journal of Econometrics*, 52, 389–402.

KRÄMER, W., AND F. MARMOL (2004): "The Power of Residual-Based Tests for Cointegration when Residuals Are Fractionally Integrated," *Economics Letters*, 82, 63–69.

LOBATO, I. N. (1999): "A Semiparametric Two-Step Estimator in a Multivariate Long Memory Model," *Journal of Econometrics*, 90, 129–153.
LOBATO, I. N., AND P. M. ROBINSON (1996): "Averaged Periodogram Estimation of Long Memory," *Journal of Econometrics*, 73, 303–324.
MARINUCCI, D., AND P. M. ROBINSON (1999): "Alternative Forms of Fractional Brownian Motion," *Journal of Statistical Planning and Inference*, 80, 111–122.
——— (2000): "Weak Convergence of Multivariate Fractional Processes," *Stochastic Processes and Their Applications*, 86, 103–120.
——— (2001): "Semiparametric Fractional Cointegration Analysis," *Journal of Econometrics*, 105, 225–247.
MARMOL, F. (1998): "Spurious Regression Theory with Nonstationary Fractionally Integrated Processes," *Journal of Econometrics*, 84, 232–250.
MARMOL, F., AND C. VELASCO (2002): "Trend Stationarity versus Long Range Dependence in Time Series Analysis," *Journal of Econometrics*, 108, 25–42.
MASIH, A. M. M., AND R. MASIH (1998): "A Fractional Cointegration Approach to Testing Mean Reversion Between Spot and Forward Exchange Rates: A Case of High Frequency Data with Low Frequency Dynamics," *Journal of Business Finance and Accounting*, 25, 987–1003.
MASIH, R., AND A. M. M. MASIH (1995): "A Fractional Cointegration Approach to Empirical Tests of *PPP*: New Evidence and Methodological Implications from an Application to the Taiwan/US Dollar Relationship," *Review of World Economics*, 131, 673–694.
PARK, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–143.
PARK, J. Y., S. OULIARIS, AND B. CHOI (1988): "Spurious Regressions and Tests for Cointegration," CAE Working Paper, 88–07, Cornell University.
PHILLIPS, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340.
PHILLIPS, P. C. B., AND S. OULIARIS (1990): "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193.
ROBINSON, P. M. (1994): "Semiparametric Analysis of Long-Memory Time Series," *The Annals of Statistics*, 22, 515–539.
——— (1995a): "Gaussian Semiparametric Estimation of Long Range Dependence," *The Annals of Statistics*, 23, 1630–1661.
——— (1995b): "Log-Periodogram Regression of Time Series with Long Range Dependence," *The Annals of Statistics*, 23, 1048–1072.
——— (1997): "Large Sample Inference for Nonparametric Regression with Dependent Errors," *The Annals of Statistics*, 25, 2054–2083.
——— (1998): "Inference-Without-Smoothing in the Presence of Nonparametric Autocorrelation," *Econometrica*, 66, 1163–1182.
ROBINSON, P. M., AND M. HENRY (1996): "Bandwidth Choice in Gaussian Semiparametric Estimation of Long Range Dependence," in *Athens Conference on Applied Probability and Time Series Analysis, Vol. II, Lecture Notes in Statistics*, 115. New York: Springer-Verlag.
ROBINSON, P. M., AND D. MARINUCCI (1998): "Semiparametric Frequency Domain Analysis of Fractional Cointegration," STICERD Discussion Paper EM/98/348, London School of Economics.
——— (2000): "The Averaged Periodogram for Nonstationary Vector Time Series," *Statistical Inference for Stochastic Processes*, 3, 149–160.
——— (2001): "Narrow-Band Analysis of Nonstationary Processes," *The Annals of Statistics*, 29, 947–986.
ROBINSON, P. M., AND C. VELASCO (1997): "Autocorrelation Robust Inference," in *Handbook of Statistics, Vol. 15 (Robust Inference)*, ed. by G. S. Maddala and C. R. Rao. Amsterdam: North-Holland, 267–298.
ROBINSON, P. M., AND Y. YAJIMA (2002): "Determination of Cointegrating Rank in Fractional Systems," *Journal of Econometrics*, 106, 217–241.

SOOFI, A. S. (1998): "A Fractional Cointegration Test of Purchasing Power Parity: The Case of Selected Members of OPEC," *Applied Financial Economics*, 8, 559–566.

TSAY, W.-J., AND C.-F. CHUNG (2000): "The Spurious Regression of Fractionally Integrated Processes," *Journal of Econometrics*, 96, 155–182.

VELASCO, C. (2002): "Log-Periodogram Inference for Nonstationary Integrated and Differenced Processes," Preprint, Universidad Carlos III de Madrid.

——— (2003): "Gaussian Semiparametric Estimation of Fractional Cointegration," *Journal of Time Series Analysis*, 24, 345–378.

WOOLDRIGE, J. (1991): "Notes on Regression with Difference-Stationary Data," Preprint, Michigan State University.