Econometric Theory, **24**, 2008, 216–255. Printed in the United States of America.
DOI: 10.1017/S0266466608080109

DISTRIBUTION-FREE TESTS OF FRACTIONAL COINTEGRATION

Javier Hualde
Universidad de Navarra

Carlos Velasco
Universidad Carlos III de Madrid

We propose tests of the null of spurious relationship against the alternative of fractional cointegration among the components of a vector of fractionally integrated time series. Our test statistics have an asymptotic chi-square distribution under the null and rely on generalized least squares–type of corrections that control for the short-run correlation of the weak dependent components of the fractionally integrated processes. We emphasize corrections based on nonparametric modelization of the innovations' autocorrelation, relaxing important conditions that are standard in the literature and, in particular, being able to consider simultaneously (asymptotically) stationary or nonstationary processes. Relatively weak conditions on the corresponding short-run and memory parameter estimates are assumed. The new tests are consistent with a divergence rate that, in most of the cases, as we show in a simple situation, depends on the cointegration degree. Finite-sample properties of the tests are analyzed by means of a Monte Carlo experiment.

1. INTRODUCTION

Inference problems on potentially cointegrated models involving fractionally integrated time series have recently received much attention in the econometric literature. Within this line of research, an important effort has been devoted to analyzing properties of estimates of the cointegrating parameters (Kim and Phillips, 2000; Robinson and Marinucci, 2001, 2003; Chen and Hurvich, 2003; Robinson and Hualde, 2003; Robinson and Iacone, 2005), and testing for cointegration or determination of the cointegrating rank (Robinson and Yajima, 2002; Breitung and Hassler, 2002; Chen and Hurvich, 2003; Hassler and Breitung,

We thank Helmut Lütkepohl and two referees for helpful comments and suggestions. We also thank participants at the NSF/NBER Time Series Conference at the University of Heidelberg, Germany, at the Unit Root and Cointegration Testing Conference at the University of Algarve, Faro, Portugal, and seminar participants at the Universidad de Navarra and Ente Luigi Einaudi for helpful comments. Javier Hualde's research is supported by the Spanish Ministerio de Educación y Ciencia through Juan de la Cierva and Ramón y Cajal contracts and ref. SEJ2005-07657/ECON. Carlos Velasco's research is supported by the Spanish Ministerio de Educación y Ciencia, ref. SEJ2004-04583/ECON. Address correspondence to Javier Hualde, Department of Quantitative Methods, School of Economics and Business Administration, Universidad de Navarra, Edificio Bibliotecas (Entrada Este), 31080 Pamplona (Navarra), Spain; e-mail: jhualde@unav.es.

2006; Marmol and Velasco, 2004 [MV hereafter]). All these problems have been tackled satisfactorily in the standard cointegrating setup with unit root levels and weakly dependent cointegrating errors (emphasized since Engle and Granger, 1987), but in the more general setting of fractional cointegration many difficulties arise, especially when no restrictions are imposed on the integration orders of the observables and/or possible cointegrating errors. In addition, although it is assumed that fractional cointegration describes a situation where a linear combination of the components of a (fractionally) integrated vector has reduced memory in some sense, there is no agreement about the precise specification of this idea. For example, Robinson and Yajima (2002) (RY hereinafter) present several definitions already proposed in the literature (Johansen, 1996; Flôres and Szafarz, 1996; Robinson and Marinucci, 2003) and offer a new one. Although it is true that all these definitions are identical if all the observables share the same integration order, there are important discrepancies among them when the vector of observables is composed of series with different integration orders. Furthermore, in a fractional cointegration framework, the real nature of the integration orders entails additional difficulties, because it seems unrealistic to assume knowledge of their precise values. Note that this is a distinctive feature from the traditional framework referred to earlier, where the knowledge of the integer degree of integration of the observables permits a variety of cointegration tests (see, e.g., Johansen, 1988; Phillips and Ouliaris, 1990).

In the general fractional cointegration setting there are relatively few proposals of testing for cointegration, and these are based on different testing strategies. Marinucci and Robinson (2001) proposed a Hausman-type procedure comparing different estimates of the memory of the observables, and recently Robinson (2005b) provided rigorous theoretical support to this idea. RY based their test on the analysis of the rank of a generalized long-run variance matrix of the weakly dependent error vector generating the fractionally integrated observables. They also developed a specific-to-general procedure for testing the necessary condition for cointegration of equality of at least some integration orders of the observables, and their cointegration testing procedure is able to determine the cointegrating rank (for a related approach, see also Chen and Hurvich, 2003). Breitung and Hassler (2002) proposed a test that extends Johansen (1988) and allows determination of the cointegrating rank of a vector of fractionally integrated processes. Hassler and Breitung (2006) test for the null of no cointegration by applying a modified Lagrange multiplier (LM) test to single equation regression residuals. Their statistic corrects for the endogeneity caused by the regressors and enjoys standard asymptotic properties. Lasak (2005) extends the likelihood ratio tests proposed by Johansen (1988) allowing for unknown cointegration order. But perhaps the closest idea to the test we propose in the present paper is the methodology designed by MV, which checks for the absence of cointegration comparing the ordinary least squares (OLS) and a generalized least squares (GLS) estimate of the cointegrating vector. These estimates enjoy opposite asymptotic properties under the null of spurious relationship and the alternative of cointegration. Their slope estimates are based in turn on memory estimates, which, as exploited in Marinucci and Robinson (2001), may be consistent only under one of such hypotheses or alternatively, may serve as a basis for comparing the memory of the levels and a possible cointegrating relationship. The estimation of memory parameters of the observables and cointegrating error is a feature also employed by Hualde and Robinson (2004, 2005, 2007) and Robinson and Hualde (2003) (RH hereinafter), who designed GLS-type estimates of the cointegrating parameter with standard asymptotic distribution (normal or mixed normal), leading to Wald-type test statistics with chi-squared null limiting distribution.

Based on these ideas, we concentrate in this paper on the problem of devising a general cointegration testing procedure with standard asymptotics valid for a general class of fractionally integrated processes. Thus, using a preliminary estimate of an appropriate projection vector obtained from the short-run structure of the fractionally integrated observables and differencing these series properly, we can recover the same type of standard asymptotics of the previously mentioned references under the null of no cointegration. This leads to a distribution-free test, hence avoiding the null nonstandard limit distribution of MV's test. Under the alternative of cointegration, the same reasoning of MV guarantees consistency of the test when cointegration induces consistency of OLS or narrow band (NB) regression and therefore of OLS or NB residual-based memory estimates.

We emphasize the use of test statistics based on semiparametric assumptions, although we will also comment on the precise circumstances in which parametric test statistics (which could enjoy better finite-sample properties), could be adequate. The semiparametric approach allows for a great deal of flexibility, accommodating situations with fractional processes of arbitrary positive memory (including simultaneous analysis of the stationary range, not covered by MV and the nonstationary one, not covered by RY) and dealing effectively with a vector series with components with different integration orders (which is not covered by MV; RY; Breitung and Hassler, 2002; or Hassler and Breitung, 2006). MV, who limited the maximum integration order allowed in their work to 3/2, indicate in their Remark 2 that their assumption of equal memory for all series is not critical, and they give some hints on how to proceed if this assumption is not satisfied. However, the practical implementation of their test statistic when this condition does not hold could be very difficult, because the null limiting distribution of their test statistic in this case would depend on all the different orders involved in the vector of observables. Also, calculation of the critical values from several integration order estimates could introduce important noise in the procedure. On the contrary, our method is not designed to check for the cointegrating rank (as does RY), being only valid for assessing statistically for the existence or not of cointegration. However, we expect similar consistency results in higher order rank cointegrated systems as in MV.

Therefore, although our test procedure is limited, we believe that given the previously mentioned improvements and extensions in some directions over existing works in the literature, it fills some relevant gaps.

The rest of the paper is organized as follows. In the next section we analyze a model for a vector of fractionally integrated time series that could potentially lead to fractional cointegration. Section 3 presents a simple (parametric) setting where the main ideas of our testing strategy are introduced. Section 4 deals with the general model mainly from a semiparametric perspective, commenting also about the plausibility of a parametric version of the proposed test statistics. Section 5 presents our assumptions and the asymptotic properties of these tests. Finally, Section 6 shows the finite-sample behavior of our test procedures. All proofs are relegated to the Appendix.

2. A POTENTIALLY FRACTIONALLY COINTEGRATED MODEL

Throughout the paper we consider the $p \times 1$ vector of fractionally integrated time series z_t given by

$$z_t = \Delta_p^{-1}(\underline{\delta})\{u_t 1(t > 0)\}, \tag{1}$$

where $1(\cdot)$ denotes the indicator function, $\Delta_p(\underline{\delta}) = diag\{\Delta^{\delta_0}, \Delta^{\delta_1}, \ldots, \Delta^{\delta_l}\}$, $p = l + 1$, with

$$\delta_0 = \max_{i \in \{1,\ldots,l\}} \delta_i > 0, \qquad \min_{i \in \{1,\ldots,l\}} \delta_i \geq 0; \tag{2}$$

the fractional difference operator $\Delta^{-\alpha}$ is defined in terms of $\Delta = 1 - L$, where L is the lag operator, and the formal series,

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

for any real $\alpha \neq -1, -2, \ldots$, where Γ is the gamma function and $\Gamma(0)/\Gamma(0) = 1$. The process u_t is a multivariate weakly dependent (perhaps only asymptotically) covariance stationary process for which specific regularity conditions will be given in Section 4. The truncation in (1) is necessary when $\delta_i \geq 0.5$, because the weights $a_j(\delta_i)$ are not square summable in this case, but it leads to nonstationary series for all $\delta_i \neq 0$, though asymptotically stationary for $\delta_i < 0.5$. The same type of model has been used by Robinson and Marinucci (2001) and RH, among others, to study the properties of parameter estimates in fractionally cointegrated systems. Other works on fractional cointegration have used an alternative definition of fractional nonstationarity built on long-memory stationary increments. In this case the levels are constructed as partial sums of such increments and are denoted Type I nonstationary processes by Marinucci and Robinson (1999), whereas z_t is termed Type II. We will refer to the ith component of z_t as an $I(\delta_i)$ process.

Condition (2) has several implications. First, we do not deal with antipersistent observables, which are processes with negative memory. These are rare in practice, and although cointegration involving this type of processes is possible in the strict sense, it does not have empirical appeal. For similar reasons we set $\max_i \delta_i > 0$, and so we deliberately avoid cointegration among $I(0)$ exclusively. More importantly, the first component of z_t is assumed to have the highest integration order. As will be seen later, the identification of the component with the highest integration degree is one of the key requirements of our procedure. This is certainly slightly restrictive, but it seems unavoidable in view of the great generality that our framework permits. Note that this requirement does not cause any difficulty in the traditional framework where cointegration among processes with the same integration order is considered, and in practical terms, it is always possible to base the choice of the particular component with highest memory on consistent estimates of the individual integration orders of the observables.

We adopt the following definition of cointegration given by Flôres and Szafarz (1996).

DEFINITION 1. *We say that z_t is cointegrated if there exists a $p \times 1$ vector $\alpha \neq 0$ such that $\alpha' z_t \sim I(\gamma)$ with $\gamma < \delta_0$ and at least a nonzero scalar component of α multiplies one component of z_t with integration order equal to δ_0.*

Obviously this definition implies that for z_t to be cointegrated it is necessary that at least one observable apart from the first one have integration order δ_0. Note that this definition could miss some cointegrating relations (for a good example, see, e.g., RY) where variables $I(\delta_0)$ are not present. However, we do not find this worrying, because in those particular cases the $I(\delta_0)$ variables would not be involved in any relation of cointegration, and so they could be removed from the model and we could interpret the new vector of observables without these variables in terms of Definition 1. Note that the test statistics that will be presented in subsequent sections test for the null of no cointegration against the alternative of cointegration. Thus, the purpose of introducing Definition 1 here is to identify the precise type of cointegrating relations our test is able to assess. Of course, our test will be able to detect more restrictive versions of cointegration (e.g., Marinucci and Robinson, 2001, where $\gamma < \min_i \delta_i$).

If $f(\lambda)$ represents the spectral density matrix of u_t (or of its covariance stationary approximation if u_t is only asymptotically stationary; see the discussion that follows for a definition), a necessary and sufficient condition for the existence of cointegration among the elements of z_t is that $f(0)$ be singular. If, on the contrary, $f(0)$ is full rank, we say that z_t is spuriously related, because any nontrivial linear combination with a nonzero component multiplying an $I(\delta_0)$ element of z_t is also $I(\delta_0)$. This discussion makes apparent that the weak dependence structure of the innovation u_t is essential to design any inferential procedure on the existence of cointegration. When dealing with multivariate fractionally integrated processes like z_t, the weakly dependent error input process u_t is usually viewed as depending only on a vector of short-memory parameters. However, in a cointegrating framework it usually depends also on memory parameters, this dependence possibly vanishing if there is no cointegration in the model. These ideas are nicely captured by the following structure. First, we partition the basic vectors as $z_t = (y_t, x_t')'$, $u_t = (u_{yt}, u_{xt}')'$, x_t, u_{xt} being $l \times 1$ vectors, noting that $y_t \sim I(\delta_0)$, the critical condition in (2) playing a role here. Next, denoting for any scalar or vector sequence ζ_t,

$$\zeta_t(c) = \Delta^c\{\zeta_t 1(t > 0)\},$$

suppose that there exist a weakly dependent covariance stationary scalar process v_{yt}, a real number γ such that $0 \leq \gamma \leq \delta_0$, and an $l \times 1$ vector $\beta \neq 0$ such that

$$u_{yt} = \beta' x_t(\delta_0) + v_{yt}(\delta_0 - \gamma). \tag{3}$$

Model (1) with (3) and $\gamma < \delta_0$ leads to a multivariate extension of the bivariate cointegrated system involving Type II fractionally integrated processes considered in Hualde and Robinson (2007) and RH, which for this case is

$$y_t = \beta' x_t + v_{yt}(-\gamma), \tag{4}$$

$$x_t = \Delta_l^{-1}(\bar{\delta})\{u_{xt} 1(t > 0)\}, \tag{5}$$

where $\Delta_l(\bar{\delta}) = diag\{\Delta^{\delta_1}, \ldots, \Delta^{\delta_l}\}$. Therefore, in view of Definition 1, testing the hypothesis of no cointegration against that of cointegration in the previous framework can be formulated in terms of the memory parameters, and so

$$H_0 : \delta_0 = \gamma \quad \text{vs.} \quad H_1 : \delta_0 > \gamma. \tag{6}$$

Note that assuming that u_{xt} is a covariance stationary $I(0)$ process and $\gamma < \delta_0$, under (3), u_{yt} is only asymptotically stationary because of the truncation on $v_{yt}(\delta - \gamma)$. Other asymptotically stationary elements of the linear combination forming u_{yt} may arise if some of the components of x_t have integration orders strictly smaller than δ_0. However, based on the following definition we could easily obtain the covariance stationary approximation of u_{yt}.

DEFINITION 2. *Given $\alpha < \frac{1}{2}$, let ξ_t be a covariance stationary $I(0)$ process and*

$$\zeta_t = \sum_{j=0}^{t} a_j(\alpha) \xi_{t-j}.$$

Then, we say that $\tilde{\zeta}_t$ is the covariance stationary approximation of ζ_t if

$$\tilde{\zeta}_t = \sum_{j=0}^{\infty} a_j(\alpha)\xi_{t-j}. \tag{7}$$

Note that setting the difference equation

$$\Delta^{\alpha}\zeta_t = \xi_t, \tag{8}$$

both ζ_t, $\tilde{\zeta}_t$, are solutions of (8) (given certain initial conditions in the case of ζ_t), but although ζ_t is not stationary, $\tilde{\zeta}_t$ is the stationary solution of (8), which exists because $\sum_{j=0}^{\infty} a_j^2(\alpha) < \infty$ for $\alpha < \frac{1}{2}$, and the process (7) is well defined. For $\alpha \geq \frac{1}{2}$, (8) does not have a stationary solution. The covariance stationary approximation of u_{yt} is given by

$$\tilde{u}_{yt} = \beta'\{\Delta^{\delta_0}\Delta_l^{-1}(\bar{\delta})\}u_{xt} + \Delta^{\delta_0-\gamma}v_{yt}.$$

Here, it is interesting to analyze the connection between β in (4) and $f(\lambda)$, the spectral density matrix of $(\tilde{u}_{yt}, u'_{xt})'$. Partition f as

$$f(\lambda) = \begin{pmatrix} f_{yy}(\lambda) & f_{yx}(\lambda) \\ f_{xy}(\lambda) & f_{xx}(\lambda) \end{pmatrix}$$

and also let $\beta = (\beta_1', \beta_2')'$, where β_1 and β_2 are $\bar{l}_1 \times 1$ and $\bar{l}_2 \times 1$ vectors corresponding to components of x_t with integration orders equal to δ_0 or smaller than δ_0, respectively, with $\bar{l}_1 + \bar{l}_2 = l$. If $\gamma < \delta_0$, $f(0)$ is singular, and if $\delta_i = \delta_0$ for all $i = 1, \ldots, l$, and there is no cointegration among the elements of x_t, β is the fundamental vector (cf. Park, Ouliaris, and Choi, 1988)

$$f_{xx}^{-1}(0)f_{xy}(0) = \beta.$$

If on the contrary $\bar{l}_1 < l$,

$$f_{xx}^{-1}(0)f_{xy}(0) = (\beta_1', 0_{\bar{l}_2}')',$$

where 0_q is a $q \times 1$ vector of zeros.

3. THE TEST PROCEDURE IN THE WHITE NOISE CASE

We find it convenient to present the basic ideas behind our test strategy in a simple setting, which will be generalized in several dimensions in Section 4. In particular, throughout this section we will consider the case where all the observables share the same integration order, denoted by δ. This condition is certainly restrictive, but it is also introduced by Breitung and Hassler (2002), Hassler and Breitung (2006), and MV and effectively also by RY, which only tests for cointegration among subsets of variables with the same integration order. Also, we will focus on the case where in (3) the $p \times 1$ vector $w_t = (v_{yt}, u'_{xt})'$ is independent and identically distributed (i.i.d.) with zero mean and nonsingular covariance matrix

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega'_{xy} \\ \omega_{xy} & \Omega_{xx} \end{pmatrix},$$

where we assume that Ω_{xx} is also positive definite. Furthermore, the i.i.d. condition of w_t will be taken as known, and so the procedure described in this section is parametric. For these reasons, this section could be considered of reduced empirical relevance, but, on the contrary, we find it very informative for grasping the intuition behind our test methodology.

Using (3) we find that $(\tilde{u}_{yt}, u'_{xt})'$ has spectral density matrix

$$f(\lambda) = \frac{1}{2\pi} \begin{pmatrix} \omega_{yy}|h(\lambda)|^2 + 2\beta'\omega_{xy}\operatorname{Re}\{h(\lambda)\} + \beta'\Omega_{xx}\beta & \omega'_{xy}h(\lambda) + \beta'\Omega_{xx} \\ \omega_{xy}h(-\lambda) + \Omega_{xx}\beta & \Omega_{xx} \end{pmatrix},$$

where $h(\lambda) = (1 - e^{i\lambda})^{\delta-\gamma}$. Then, when $\gamma = \delta$, $(u_{yt}, u'_{xt})' = (\tilde{u}_{yt}, u'_{xt})'$ is a white noise sequence with nonsingular constant spectral density matrix, $f(\lambda) = f(0)$, which does not depend on γ or δ because $h(0) = 1$ in this case. However, when $\gamma < \delta$, we find that $f(0)$ is singular because then $h(0) = 0$, so that z_t is cointegrated.

In view of (6), estimates of δ and γ can be useful to derive hypothesis tests of the null of no cointegration. Although the values of the nuisance parameters δ and β are in general unknown, these could be estimated from data, and from these estimates, in turn, we may be able to estimate γ consistently from residuals, as is discussed later. However, following the route of MV, we use such estimates through a procedure that takes advantage of the divergence of the sample moments of z_t, which, for example, also leads to nonstandard asymptotic properties of usual statistics, such as OLS coefficients. Note that the estimation of γ from residuals is inherent to our approach, and so our test, although nonstandard, could be regarded as a "residual-based regression test."

We define the projection vector

$$\eta = f_{xx}^{-1}(0)f_{xy}(0), \tag{9}$$

noting that

$$\begin{aligned} \eta &= \Omega_{xx}^{-1}\omega_{xy} + \beta, && \text{under } H_0, \\ &= \beta, && \text{under } H_1. \end{aligned}$$

Letting c, h, be any possible value or estimate of the parameters γ, η, define also the fractionally differenced residuals

$$v_t(c, h) = y_t(c) - h'x_t(c),$$

which are one of the key elements of our first approximation to the cointegration test problem. Under H_0, it is evident that $v_t(\gamma,\eta)$ is a white noise with variance

$$\omega_{y.x} = 2\pi f_{yy}(0)(1-\rho_{y.x}^2), \tag{10}$$

where $\rho_{y.x}^2$ is the squared coefficient of multiple correlation between u_{yt} and u_{xt} given by

$$\rho_{y.x}^2 = \frac{f_{yx}(0)f_{xx}^{-1}(0)f_{xy}(0)}{f_{yy}(0)}. \tag{11}$$

More importantly, under H_0,

$$E(x_t v_t(\gamma,\eta)) = 0, \tag{12}$$

for all t, whereas, noting that for any $l \times 1$ vector ζ,

$$v_t(\gamma,\zeta) = (\eta-\zeta)' x_t(\gamma) + v_{yt},$$

we have that under H_1, $v_t(\gamma,\zeta)$ is $I(\delta-\gamma)$ and correlated with x_t when $\zeta \neq \eta$. Thus, it appears that a sensible strategy for testing (6) is to base our procedure on an appropriately normalized version of the sample counterpart of (12) using consistent estimates of γ and η under the null. Under the alternative, however, inconsistent estimation of η guarantees that (12) fails and the residuals v_t are no longer $I(0)$.

Thus, setting

$$\tau_n(c,h) = \sum_{t=1}^{n} x_{t-1} v_t(c,h) \tag{13}$$

(we explain later in this section why x_{t-1} replaces the "more natural" x_t in (13)), it can be shown that under additional regularity conditions (to be detailed in the next section),

$$n^{-1/2}\tau_n(\gamma,\eta) \to_d N(0, \omega_{y.x} E(\tilde{x}_t \tilde{x}_t')), \tag{14}$$

if $\delta < \frac{1}{2}$, where

$$\tilde{x}_t = \sum_{j=0}^{\infty} a_j(\delta) u_{x,t-j}.$$

By contrast, denoting by $\Rightarrow$ convergence in the Skorohod topology on the appropriate metric space,

$$n^{-\delta}\tau_n(\gamma,\eta) \Rightarrow \int_0^1 W_x(r;\delta)\, dW_{y.x}(r), \tag{15}$$

when $\delta > \frac{1}{2}$. Here the (Type II, Marinucci and Robinson, 1999) fractional Brownian motion (fBm) $W_x(r;\delta)$ is defined as

$$W_x(r;\delta) = \Gamma(\delta)^{-1} \int_0^r (r-s)^{\delta-1}\, dW_x(s)$$

in terms of the last l components of the $p \times 1$ vector Brownian motion (Bm) $W(r) = (W_y(r), W_x'(r))'$ with covariance matrix $2\pi f(0)$, and the univariate Bm

$$W_{y.x}(r) = W_y(r) - W_x'(r)\eta$$

is independent of W_x. The right side of (15) is a mixed normal distribution, and so in view of this result and also (14), it is expected that an appropriately normalized statistic based on $\tau_n(\gamma,\eta)$ has a χ^2 limiting distribution irrespective of whether $\delta < \frac{1}{2}$ or $\delta > \frac{1}{2}$. In fact, defining for $b \neq 0$ the statistic

$$\Xi_n(b,c,h) = \frac{\tau_n'(c,h)\left(\sum_{t=1}^{n} x_{t-1}x_{t-1}'\right)^{-1}\tau_n(c,h)}{b},$$

it is straightforward to show that

$$\Xi_n(\omega_{y.x},\gamma,\eta) \to_d \chi_l^2 \quad \text{under } H_0.$$

As mentioned before, one of the key elements of this test procedure is the residual $v_t(\gamma,\eta)$, which is constructed from the differenced processes $y_t(\gamma)$ and $x_t(\gamma)$. Note that under H_0, the observables y_t, x_t, are filtered by their integration order because $\gamma = \delta$, whereas under H_1 they are underdifferenced and will not deliver $I(0)$ residuals if η is not estimated consistently. An argument against this strategy could be that differencing in possibly cointegrated frameworks is usually not appropriate and could imply a loss of power. However, Hualde and Robinson (2007) and RH have found that "proper" differencing in cointegrated models leads to estimates of the cointegrating vector with optimal asymptotic properties. This is precisely the type of filtering we propose in our cointegration tests, although of course it is not obvious that optimal properties in estimation would automatically be translated into testing situations, and, undoubtedly, further research would be needed to explore this connection.

These results can serve as a basis for a distribution-free test of the null of no cointegration based on rejecting H_0 for large values of Ξ_n compared with a χ_l^2 distribution, once consistent estimates of the unknown γ, η, and $\omega_{y.x}$ are found. As will be seen in the next section, under correlated $I(0)$ innovations we should replace the basic OLS-type fluctuations τ_n by those of alternative statistics that preserve a similar orthogonality property to that achieved by $v_t(\gamma,\eta)$ with x_t by accounting for such weak dependence in a general framework.

Obviously, $\Xi_n(\omega_{y.x}, \gamma, \eta)$ is an infeasible statistic because, in general, both the elements of $f(0)$ and γ are unknown. However, given an estimate $\hat{\delta}$ of δ, we can easily estimate the elements of $f(0)$ by

$$\hat{f}_{yy}(0) = \frac{1}{2\pi n} \sum_{t=1}^{n} y_t^2(\hat{\delta}), \qquad \hat{f}_{xy}(0) = \frac{1}{2\pi n} \sum_{t=1}^{n} x_t(\hat{\delta}) y_t(\hat{\delta}),$$

$$\hat{f}_{xx}(0) = \frac{1}{2\pi n} \sum_{t=1}^{n} x_t(\hat{\delta}) x_t'(\hat{\delta}), \tag{16}$$

and then from (9)–(11) obtain easily corresponding estimates of η and $\omega_{y.x}$, respectively. The $\hat{\delta}$ term could be recovered from levels y_t or x_t or from (asymptotically) stationary increments Δy_t or Δx_t, with a rate of convergence

$$\hat{\delta} = \delta + O_p(n^{-\kappa}), \qquad \kappa > 0. \tag{17}$$

Most analyses of usual parametric and semiparametric memory estimates use the alternative Type I definition of nonstationary processes, but they can be shown to have the same properties under (1) using the techniques of Robinson (2005a) and Velasco (2004). Thus, under (17), building on the results of RH, it is not difficult to show that the estimates $\hat{\omega}_{y.x}$, $\hat{\eta}$ of $\omega_{y.x}$ and η based on (16) are $\sqrt{n}$-consistent because of their parametric nature.

To obtain consistent estimates of γ we can use the OLS or NB residuals

$$\hat{v}_t = y_t - x_t' \hat{\beta}$$

to get

$$\hat{\gamma} = \gamma + O_p(n^{-\kappa}), \qquad \kappa > 0,$$

under both hypotheses. If H_0 is true, then $\gamma = \delta$, and because $\hat{\beta}$ is inconsistent for β, then $\hat{v}_t$ is a linear combination (with stochastic coefficients) of $I(\delta)$ processes in a noncointegrating direction, so that $\hat{\gamma}$ is expected to be a consistent estimate of δ. We give a richer justification of this fact in the following section. On the contrary, under H_1, $\gamma < \delta$, we have that $\hat{\beta}$ is consistent (note that the OLS could be inconsistent if $\delta < \frac{1}{2}$, but the NB suffices), and so residuals $\hat{v}_t$ are approximately $I(\gamma)$ and can be used to estimate consistently $\gamma < \delta$. See, e.g., Velasco (2003) and Hassler, Marmol, and Velasco (2006), which justified residual semiparametric memory estimation under weak assumptions for Type I fractional processes.

Then, proceeding as in RH, given $\sqrt{n}$-consistent estimates $\hat{\eta}$, $\hat{\omega}_{y.x}$, the rate in (17) is sufficient to show that under H_0

$$\Xi_n(\hat{\omega}_{y.x}, \hat{\gamma}, \hat{\eta}) - \Xi_n(\omega_{y.x}, \gamma, \eta) = o_p(1),$$

so that our feasible test statistics share the same (first-order) asymptotic properties as the infeasible ones,

$$\Xi_n(\hat{\omega}_{y.x}, \hat{\gamma}, \hat{\eta}) \to_d \chi_l^2 \quad \text{under } H_0,$$

for both $\delta < \frac{1}{2}$ and $\delta > \frac{1}{2}$. When $\delta < \frac{1}{2}$, the effect of the estimation of η is negligible thanks to the utilization of the regressor x_{t-1} in τ_n instead of other alternatives, such as x_t. However in a general parametric framework a similar strategy seems not to be possible; see Remark 4 in Section 5. On the contrary, when $\delta > \frac{1}{2}$ it is indifferent to use x_t or x_{t-1} in (13).

Finally, to conclude the analysis of the white noise situation we present a brief justification of the different sources of power of the test. These ideas are again better described in this simple setting, although similar reasoning would apply to the general test procedures we present in Section 4. First, under H_1, $v_t(\gamma, \eta)$ is still an $I(0)$ process uncorrelated to x_{t-1}, but, as mentioned before, for any $\zeta \neq \eta$, $v_t(\gamma, \zeta)$ is $I(\delta - \gamma)$ and correlated with x_{t-1}. Thus, following the Robinson and Marinucci (2001) results, we can obtain the following sharp rates for $\tau_n(\gamma, \zeta)$ under H_1:

$$\begin{aligned} n^{-\max\{1/2,\delta\}}\tau_n(\gamma,\zeta) &= O_p(n^{1/2}), \qquad \delta < \tfrac{1}{2}, \\ &= O_p(n^{1-\delta}), \qquad \delta > \tfrac{1}{2}, \qquad 2\delta - \gamma < 1, \\ &= O_p(n^{1-\delta}\log n), \qquad \delta > \tfrac{1}{2}, \qquad 2\delta - \gamma = 1, \\ &= O_p(n^{\delta-\gamma}), \qquad \delta > \tfrac{1}{2}, \qquad 2\delta - \gamma > 1, \end{aligned}$$

all diverging with n, in contrast with (14) and (15) under the null. Thus, the key is to employ an estimate of η consistent under the null but inconsistent under the alternative. Following MV's ideas, there are ways to increase these divergence rates under H_1, for example, by proposing consistent estimates of η under H_0 that diverge under the alternative. For our simple model, this could be

$$\tilde{\eta} = \left(\sum_{t=1}^{n} x_t(\hat{\delta}) x_t'(\hat{\delta}) \right)^{-1} \sum_{t=1}^{n} x_t(\hat{\gamma}) y_t(\hat{\gamma}), \tag{18}$$

for which the following sharp rates can be derived:

$$\begin{aligned} \tilde{\eta} &= O_p(1), \qquad \delta - \gamma < \tfrac{1}{2}, \\ &= O_p(\log n), \qquad \delta - \gamma = \tfrac{1}{2}, \\ &= O_p(n^{2(\delta-\gamma)-1}), \qquad \delta - \gamma > \tfrac{1}{2}, \end{aligned}$$

so that, in case $\delta - \gamma \geq \frac{1}{2}$, the divergence rate of the feasible τ_n under H_1 can be improved upon by using $\tilde{\eta}$ instead of $\hat{\eta}$. Note that under H_0, it is asymptotically equivalent to use $\hat{\eta}$ or $\tilde{\eta}$, as both are $\sqrt{n}$-consistent estimates of η.

The second source of power is due to the fact that $\rho_{y.x}^2 = 1$ under H_1, so that provided one can get $\sqrt{n}$-consistent estimates of $\rho_{y.x}^2$ under H_1, $\hat{\rho}_{y.x}^2 - 1 = O_p(n^{-1/2})$, and noting the denominator of $\Xi_n(\hat{\omega}_{y.x}, \hat{\gamma}, \hat{\eta})$, this rate also adds to previously reported divergence rates under H_1. Thus, in this case, the exact divergence rates of the test statistic under H_1 are

$$
\begin{aligned}
\Xi_n(\hat{\omega}_{y.x}, \hat{\gamma}, \tilde{\eta}) &= O_p(n^{3/2}), \qquad \delta < \tfrac{1}{2},\\
&= O_p(n^{5/2-2\delta}), \qquad \delta > \tfrac{1}{2}, \qquad 2\delta - \gamma < 1,\\
&= O_p(n^{5/2-2\delta}\log^2 n), \qquad \delta > \tfrac{1}{2}, \qquad 2\delta - \gamma = 1,\\
&= O_p(n^{2(\delta-\gamma)+1/2}), \qquad \delta > \tfrac{1}{2}, \qquad 2\delta - \gamma > 1, \qquad \delta - \gamma < \tfrac{1}{2},\\
&= O_p(n^{2(\delta-\gamma)+1/2}\log n), \qquad \delta > \tfrac{1}{2}, \qquad \delta - \gamma = \tfrac{1}{2},\\
&= O_p(n^{4(\delta-\gamma)-1/2}), \qquad \delta > \tfrac{1}{2}, \qquad \delta - \gamma > \tfrac{1}{2}.
\end{aligned}
$$

4. THE GENERAL COINTEGRATION TEST

The arguments used to construct the test in the previous section are only valid when, under H_0, the weakly dependent vector u_t is an i.i.d. process, and, moreover, this circumstance is known to the researcher, so the procedure was essentially parametric. This case was adequate to illustrate the idea behind our test procedure but undoubtedly is very restrictive. Thus, throughout this section we will work under a condition that imposes some regularity on the dynamics of u_t (see also Assumptions A–E in MV), while keeping, as in the white noise situation, the modelization proposed in (1). The main distinctive feature of our approach now will be that under correlated $I(0)$ innovations we replace the basic OLS-type fluctuations τ_n by those of alternative statistics that preserve a similar orthogonality property to that achieved by $v_t(\gamma, \eta)$ with x_t (cf. (12)), by accounting for such weak dependence in a general framework. As will be seen, there are different ways of achieving this, but we will emphasize the use of semiparametric procedures over parametric ones. There are three important reasons that drive this choice. First, a parametric procedure requires knowledge, up to a finite vector of unknown parameters, of the model generating u_t. Here, even if ways of testing for this have been proposed in the literature, this knowledge could be difficult to justify especially when the dimensionality of u_t is high. In practice, the researcher could take the approach of fitting to u_t a relatively large vector autoregressive moving average (VARMA) process, but estimation of a large number of parameters could entail difficulties. Furthermore, corresponding asymptotic theory developed in the next section holds in a fully parametric approach only if all the observables are purely nonstationary. Sometimes this requirement is not very strong, because it is widely assumed that nonstationary processes have a very important role in economics, but undoubtedly it introduces an additional limitation. Finally, from a practical point of view, it seems that the identification of the parametric structure of u_t is only feasible if all the observables share the same integration order (see the discussion that follows for an explanation of this point).

Thus, a semiparametric approach that, although still stressing the fractionally integrated nature of z_t, does not assume any parametric model for u_t could be certainly preferable. Fortunately, this approach allows us to propose test statistics that are valid for any nonnegative value of δ_i, $i = 0, \dots, l$, excluding $\frac{1}{2}$, and so basically no a priori knowledge of the type discussed before is needed to apply the following techniques.

Denote by I_k the $k \times k$ identity matrix, denote by $\|\cdot\|$ the euclidean norm, and consider that a function $g(x)$ (defined on an interval I) satisfies a Lipschitz condition of order α ($g \in Lip(\alpha)$) if there exist two positive constants M, α, such that $|g(x) - g(y)| \leq M|x - y|^{\alpha}$ for all $x, y \in I$. We set the following condition that will characterize the short-run dynamics of u_t.

Assumption 1. For $0 \leq \gamma \leq \delta_0$, there exists an $l \times 1$ vector $\beta \neq 0$ such that (3) holds and the process $w_t = (v_{yt}, u'_{xt})'$, $t = 0, \pm 1, \dots$ has representation

$$w_t = A(L)\varepsilon_t,$$

where

$$A(s) = I_p + \sum_{j=1}^{\infty} A_j s^j$$

and the A_j are $p \times p$ matrices such that

(a)

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

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

(c) the ε_t are i.i.d. vectors with mean zero, positive definite covariance matrix Ω, and $E\|\varepsilon_t\|^q < \infty$, $q \geq 4$, $q > 2/(2\min_{i:\delta_i > 1/2} \delta_i - 1)$.

Assumption 1 is sufficient to apply the functional limit theorem of Marinucci and Robinson (2000), which will be needed to obtain the asymptotic null distribution of our test statistics. The conditions on the process w_t set by this assumption are identical to those in Assumption 1 of RH and hold for stationary and invertible autoregressive moving average (ARMA) processes. Under H_0,

$$u_t = C(L)w_t, \qquad C(L) = \begin{pmatrix} 1 & \beta'\Delta^{\delta_0}\Delta_l^{-1}(\bar{\delta}) \\ 0_l & I_l \end{pmatrix},$$

so that in general the spectral density of u_t depends on memory parameters. In the important case where $\delta_i = \delta_0$ for all $i \in \{1,\ldots,l\}$, this dependence disappears, and $C(L)$ reduces to

$$C(L) = C = \begin{pmatrix} 1 & \beta' \\ 0_l & I_l \end{pmatrix}.$$

In this particular case $f(\lambda)$ inherits the smoothness properties of $A(e^{i\lambda})$, but, if the equality of the orders of the observables does not hold, the presence of components like $(1 - e^{i\lambda})^a$ for some $a > 0$ in $f(\lambda)$ affects severely the smoothness of f, and this could have important effects on the properties of the estimate of f, which is required to obtain our feasible test statistics.

As mentioned before, when short-run correlation is allowed in the basic $I(0)$ input of the fractional processes, some sort of prewhitening or previous orthogonalization should be performed to maintain a test statistic with standard asymptotic distribution. With this purpose, we use the random fluctuations of GLS type inspired by Hualde and Robinson (2005), controlling for the short-run correlation of the weakly dependent u_t. Thus, we propose frequency-domain test statistics that we find more natural in our semiparametric setting. Defining, for any sequences a_t, b_t (possibly identical to a_t), the discrete Fourier transform and (cross-)periodogram as

$$w_a(\lambda) = \frac{1}{(2\pi n)^{1/2}} \sum_{t=1}^{n} a_t \exp(i\lambda t), \qquad I_{ab}(\lambda) = w_a(\lambda) w_b'(-\lambda),$$

$$I_a(\lambda) = I_{aa}(\lambda),$$

and

$$p(\lambda) = \zeta' f(\lambda)^{-1},$$

where $\zeta = (1, 0_l')'$, given any real function $g(\lambda)$, $\lambda \in [-\pi, \pi]$, we consider statistics based on

$$\hat{\tau}_m(c, \bar{d}, g) = \sum_{j=0}^{m} s_j \operatorname{Re} w_x(-\lambda_j) g(\lambda_j) w_{z(c,\bar{d})}(\lambda_j),$$

$$\hat{\tau}_m^0(c, \bar{d}, g) = \sum_{j=0}^{m} s_j \operatorname{Re} w_x(-\lambda_j) g(0) w_{z(c,\bar{d})}(\lambda_j), \tag{19}$$

where $\lambda_j = 2\pi j/n$ are the Fourier frequencies, $\bar{d} = (d_1, \ldots, d_l)'$, $z_t(c, \bar{d}) = (y_t(c), x_{1t}(d_1), \ldots, x_{lt}(d_l))$, m is a sequence tending to infinite such that $m \leq n/2$, and $s_j = 1, j = 0, n/2$; $s_j = 2$, otherwise, and in all cases we will set

$$g(\lambda) = \hat{p}(\lambda) = \zeta' \hat{f}(\lambda)^{-1},$$

where $\hat{f}(\lambda)$ is a nonparametric estimate of $f(\lambda)$ for which precise conditions will be imposed later.

Denoting by $[a]$ the integer part of a, note that because of the symmetry properties of the Fourier transform, we have that, for example,

$$\hat{\tau}_{[n/2]}(c, \bar{d}, \hat{p}) = \sum_{j=1}^{n} w_x(-\lambda_j) \hat{p}(\lambda_j) w_{z(c,\bar{d})}(\lambda_j),$$

which, because of the presence of all the Fourier frequencies, could be referred to as a full band statistic. When $m/n \to 0$ as $n \to \infty$, $\hat{\tau}_m(c, \bar{d}, g)$ only considers frequencies on a shrinking neighborhood around zero and for this reason is usually referred to as an NB statistic. Some of the results that follow will also apply to cases where $m/n \to K < \frac{1}{2}$ as $n \to \infty$, but these do not have much intuitive appeal and are hardly stressed in the literature. The expression $\hat{\tau}_m^0(c, \bar{d}, \hat{p})$ is related to what Hualde and Robinson (2005) denoted as a "zero-frequency" statistic, because the weighting factor $\hat{p}$ is only evaluated at zero frequency, so that, strictly speaking, the GLS weighting is not correct but only approximate, noting that if f is smooth around 0 and $m/n \to 0$ this approximation should be appropriate. As we show subsequently, under certain conditions on m, the statistic leads to the same asymptotic results as if the weighting factors are evaluated at the different Fourier frequencies.

We now propose our semiparametric test statistics. Defining $\hat{q}(\lambda) = \zeta' \hat{f}(\lambda)^{-1} \zeta$,

$$\hat{b}_m = \sum_{j=0}^{m} s_j \operatorname{Re} \hat{q}(\lambda_j) I_x(\lambda_j), \qquad \hat{b}_m^0 = \hat{q}(0) \sum_{j=0}^{m} s_j I_x(\lambda_j),$$

we will reject the null of no cointegration for large values of

$$\hat{\Upsilon}_m(c, \bar{d}, g) = \hat{\tau}_m'(c, \bar{d}, g) \hat{b}_m^{-1} \hat{\tau}_m(c, \bar{d}, g),$$

$$\hat{\Upsilon}_m^0(c, \bar{d}, g) = \hat{\tau}_m^{0\prime}(c, \bar{d}, g) (\hat{b}_m^0)^{-1} \hat{\tau}_m^0(c, \bar{d}, g),$$

where the unknowns $(c, \bar{d}, g)$ are replaced by appropriate estimates of $(\gamma, \bar{\delta}, p)$.

Alternatively, considering the known function $f(\lambda; h)$, $h \in \mathbb{R}^k$, where for a $k \times 1$ vector θ of unknown parameters $f(\lambda, \theta) = f(\lambda)$, it is straightforward to design a parametric version of our test statistics. To develop this extension, it is important to take into account important aspects that differ from the semiparametric situation. First, although theoretically it is possible to carry out the analysis of the parametric case allowing the integration orders of the components of z_t to differ, in practical terms this is not very relevant, because, if this is the case, it is not feasible to identify the parametric model driving u_t under H_0. Although the model for u_{xt} is identifiable from residuals based on fractional differences of x_t (from estimates of the respective integration orders of its components), u_{yt} is a linear combination of $I(0)$ and overdifferenced x_t's under H_0, if some integration orders are smaller than δ_0, which makes it practically infeasible to identify the parametric structure of the whole vector u_t on which the orthogonalization we need to apply is based. Fortunately, if all observables share the integration order δ_0, u_{yt} is a linear combination of $I(0)$ processes, and its parametric structure could be recovered. Note that assuming common memory does not imply any loss of generality with respect to previous works but undoubtedly is a limitation of the approach.

Then, under the assumption of common memory, a natural parametric statistic that exploits all the information contained in f and could be the basis of our test procedure is

$$\tilde{\tau}_n(c,h) = \sum_{j=1}^{n} w_x(-\lambda_j) p(\lambda_j;h) w_{z(c)}(\lambda_j),$$

where $p(\lambda;h) = \zeta' f^{-1}(\lambda;h)$. The feasibility of the test now depends on estimates of γ and θ to replace c and h, respectively, in $\tilde{\tau}_n(c,h)$. Finally, defining

$$\tilde{b}_n(h) = \sum_{j=1}^{n} q(\lambda_j;h) I_x(\lambda_j),$$

our parametric test statistic is

$$\tilde{\Upsilon}_n(c,h) = \tilde{\tau}_n'(c,h)\tilde{b}_n^{-1}(h)\tilde{\tau}_n(c,h).$$

Given a consistent estimate of γ, calculating the residual vector $z_t(\hat{\gamma})$ it is possible to identify the parametric model driving u_t, and, building on the results of RH, it is simple to show that parametric estimates of θ based on $z_t(\hat{\gamma})$ enjoy the same asymptotic properties as those based on u_t, for which $\sqrt{n}$-consistency and asymptotic normality are fully developed in the multivariate framework (see, e.g., Dunsmuir and Hannan, 1976; Dunsmuir, 1979). Here, methods that estimate simultaneously short- and long-memory parameters could be also useful. For example, inference in multivariate fractionally integrated vectors has also been pursued recently by Gil-Alaña (2003), extending the work of Robinson (1994), and in (possibly) cointegrated systems by Dueker and Startz (1998) and Hassler and Breitung (2006).

5. ASYMPTOTIC PROPERTIES OF COINTEGRATION TESTS

To derive the asymptotic properties of our test statistics we need first some conditions on the estimates of the integration orders and $f(\lambda)$. Thus, we impose the following condition.

Assumption 2. Under the null and the alternative hypotheses, there exist a $K < \infty$ and estimates $\hat{\gamma}$, $\hat{\bar{\delta}}$ of γ, $\bar{\delta}$, respectively, such that

$$|\hat{\gamma}| + \|\hat{\bar{\delta}}\| \leq K \qquad \textbf{(20)}$$

and $\kappa > 0$ such that

$$\hat{\gamma} = \gamma + O_p(n^{-\kappa}), \qquad \hat{\bar{\delta}} = \bar{\delta} + O_p(n^{-\kappa}), \qquad \textbf{(21)}$$

where, as $n \to \infty$,

$$n^{-\kappa} m^{1-\max\{\min\{\delta_1,\ldots,\delta_l,1\},1/2\}} \log m \to 0. \qquad \textbf{(22)}$$

There are several important remarks related to this assumption. First, (20) is not restrictive if our semiparametric estimates are optimizers of corresponding loss functions over compact sets. Next, the likability of (21) and (22) for $\hat{\gamma}$, which was briefly described in Section 3, is definitely not a trivial issue. Under H_0, the residuals (e.g., OLS or NB) on which the estimation of γ should be based are a linear combination (with stochastic coefficients) of fractionally integrated processes with dominant order δ_0. The presence of these stochastic coefficients complicates matters substantially, and although a very detailed analysis goes beyond the scope of the present paper, we offer a brief justification of why (21) holds for a particular estimate of γ, the Gaussian semiparametric, proposed by Künsch (1987) and analyzed by Robinson (1995a). First, suppose that all observables share the same integration order δ_0 and for simplicity let $\frac{1}{2} < \delta_0 < \frac{3}{2}$. Denote by $\hat{\beta}$ the OLS or NB estimate of β. Under H_0, it can be shown that $\hat{\beta}$ converges weakly, and so $\hat{\beta} = O_p(1)$ and $\|\hat{\beta}\| > 0$ with probability tending to one. Define $\hat{b} = (1, \hat{\beta}')'/\|(1, \hat{\beta}')\|$, so that $\|\hat{b}\| = 1$ and $\hat{v}_t(\hat{b}) = \hat{b}' z_t$. Clearly, under Assumption 1, the spectral density of Δz_t behaves like

$$f_{\Delta z}(\lambda) \sim G_{\Delta z} \lambda^{-2(\delta_0 - 1)} \quad \text{as } \lambda \to 0,$$

for a certain $p \times p$ matrix $G_{\Delta z}$, which is positive definite under H_0. Then, following the arguments in Chen and Hurvich (2006), replacing the true constant of the spectral error sequence by the random quantity $\hat{b}' G_{\Delta z} \hat{b}$, which is strictly positive (with probability tending to one) by the positive definiteness of $G_{\Delta z}$, all the results on consistency of $\hat{\gamma}$ hold following the results of Robinson (1995a) and Lobato (1999). It can also be obtained that $\hat{\gamma} - \gamma = O_p(m^{-1/2})$, where m satisfies the usual restriction

$$\frac{1}{m} + \frac{m^{1+2\rho}}{n^{2\rho}} \log^2 m \to 0 \quad \text{as } n \to \infty,$$

with $\rho \in (0, \rho^*)$, where, as in Robinson (1995a), $\rho^* \in (0, 2]$ is the parameter related to the smoothness of the spectral density of Δz_t around frequency zero.

Thus, in the most favorable case the least strict bound for the convergence rate of $\hat{\gamma}$ is $n^{-2/5}\log^{1/5} n$ (if $\rho^* = 2$).

When the integration orders of the observables are not all equal, the situation is much more complicated. In this case, those components of $\hat{\beta}$ corresponding to processes with integration order smaller than δ_0 are typically not bounded in probability, invalidating in principle the estimation of γ based on the residuals $y_t - \hat{\beta}'x_t$. For the sake of an easy exposition, collect those components of $\hat{\beta}$ in an $l^* \times 1$ vector $\hat{\beta}^*$, where $l^* \in \{1,\ldots,l-1\}$, and correspondingly let x_t^* be the vector of components of x_t with integration order strictly smaller than δ_0. To describe the procedure we propose in this situation let $\hat{\beta}$ also be an NB estimate. Then, it can be easily shown that if $\delta_0 < \frac{1}{2}$ (see, e.g., Robinson and Marinucci, 2003), the rate of divergence of $\hat{\beta}^*$ can be made arbitrarily small (up to a power rate) by letting the rate at which the NB bandwidth increases be arbitrarily close to (but smaller than) n. Thus, letting the bandwidth employed in the estimation of γ grow more slowly than that used in the NB estimation of β, the contribution of $\hat{\beta}^{*\prime}x_t^*$ to the spectral density of the residuals $y_t - \hat{\beta}'x_t$ can be made negligible, and so the terms with δ_0 memory dominate. Note that this strategy is only valid if $\delta_0 < \frac{1}{2}$, but if we suspect that, for example, $\frac{1}{2} < \delta_0 < \frac{3}{2}$, the same reasoning applies if we estimate β from the NB regression of Δy_t on Δx_t, obtaining $\hat{\beta}$, and then estimating γ from $y_t - \hat{\beta}'x_t$. Higher δ_0's could be treated by estimating β from higher integer differences of the observables.

Under H_1, if $\delta_0 = \delta_i$ for all $i \in \{1,\ldots,l\}$, the first part of (21) is well known for an estimate of γ based on OLS or NB residuals (see, e.g., Velasco, 2003). Here, OLS residuals are not a good proxy of the true cointegrating errors if $\delta_0 < \frac{1}{2}$ (so there is the so-called stationary cointegration), but NB residuals suffice. When $\delta_i < \delta_0$ for some $i \in \{1,\ldots,l\}$, if $\gamma < \min_i \delta_i$, our estimate of β (calculated from integer differences of the observables) will be consistent (although its rate of convergence could be very slow), because taking integer differences of the observables the cointegrating structure is preserved. Thus, estimating γ from $y_t - \hat{\beta}'x_t$ will lead to a consistent estimate of γ under H_1, and the test will gain power. If $\gamma < \delta_0$ but $\gamma \geq \delta_i$ for some $i \in \{1,\ldots,l\}$, some of the components of $\hat{\beta}$ could diverge, but as under H_0, this rate of divergence could be made arbitrarily small, and similarly, the estimation of γ is not going to be affected if we restrict the rate of growth of the bandwidth employed in this estimation accordingly.

The conditions for $\hat{\bar{\delta}}$ are satisfied for standard semiparametric estimates of $\bar{\delta}$, based on the corresponding components of z_t. Finally, (22) reflects a trade-off between the rate of growth of m and the smoothness of f through the positive relation between this smoothness and κ. Note however that even if κ is very small, (22) could be satisfied by constraining the rate of growth of m appropriately. This is of primal importance, because, as mentioned before, depending on the values of the memory parameters of the observables, the lack of smoothness of f is a very realistic possibility. If m grows at the same rate as n and

$\delta_0 > \frac{1}{2}$, almost $\sqrt{n}$-consistent estimates of the orders might be necessary, and so some bias-reduction estimation procedures might be required (see Hualde and Robinson, 2004).

We impose on $\hat{f}$ either of the following conditions, which will be used for $\hat{\tau}_m(c, \bar{d}, g)$ and $\hat{\tau}_m^0(c, \bar{d}, g)$, respectively.

Assumption 3. Uniformly in j, there exist $\varkappa > 0$, $\phi > 0$, such that

$$\hat{f}(\lambda_j) - f(\lambda_j) = O_p(n^{-\varkappa}),$$

$$\hat{f}(\lambda_{j+1}) - f(\lambda_{j+1}) - (\hat{f}(\lambda_j) - f(\lambda_j)) = O_p(n^{-\phi}),$$

where, as $n \to \infty$,

$$n^{-\varkappa} m^{1-\max\{\min\{\delta_1,\dots,\delta_l,1\},1/2\}} \to 0, \tag{23}$$

$$n^{-\phi} m^{2-\max\{\min\{\delta_1,\dots,\delta_l,1\},1/2\}} \to 0.$$

Assumption 4. There exists $\varkappa > 0$ such that

$$\hat{f}(0) - f(0) = O_p(n^{-\varkappa}),$$

for which (23) is satisfied.

Both assumptions are unprimitive but Hualde and Robinson (2004) justified them rigorously under general conditions for particular estimates of f. Note that these estimates could be based on residuals $z_t(\hat{\delta})$, for a certain estimate of δ consistent under both hypotheses, or alternatively on residuals $z_t(\hat{\gamma}, \hat{\hat{\delta}})$, which under H_0 behave similarly to $z_t(\hat{\delta})$ but under H_1 could lead to divergent estimates of f, which could add power to the test. As for the estimates of the orders, m could be restricted appropriately to deal with the lack of smoothness of f.

We do not consider the specific case where components of x_t have an integration order equal to $\frac{1}{2}$, for which we introduce the following condition.

Assumption 5. $\delta_i \neq \frac{1}{2}$ for all $i = 0, \dots, l$.

To get a neat asymptotic theory, without loss of generality, we reorder the variables in x_t according to

$$\delta_1 \geq \delta_2 \geq \cdots \geq \delta_l \geq 0.$$

Thus, we set

$$\bar{\delta} = (\bar{\delta}_1', \bar{\delta}_2')', \quad \text{where } \bar{\delta}_1 = (\delta_1, \dots, \delta_{l_1})', \qquad \bar{\delta}_2 = (\delta_{l_1+1}, \dots, \delta_l)',$$

with $\delta_i > \frac{1}{2}$, $i = 1, \dots, l_1$ and $\delta_i < \frac{1}{2}$, $i = l_1 + 1, \dots, l$, where $l_1 = l$ indicates that all integration orders of the x's are strictly larger than $\frac{1}{2}$, $l_1 = 0$ meaning that all the orders are smaller than $\frac{1}{2}$.

Finally, we impose some conditions on the bandwidth m.

Assumption 6. If $l_1 < l$, for ϵ in Assumption 1,

$$m^{\max_{i\in\{l_1+1,\dots,l\}}\delta_i - 1/2}\log^{1/2} n + m^{3+2\epsilon}/n^{2+2\epsilon} \to 0, \quad \text{as } n \to \infty.$$

Assumption 7. Assumption 6 holds and if $l_1 > 0$

$$m/n^{\min_{i\in\{1,\dots,l_1\}}\delta_i} \to 0, \quad \text{as } n \to \infty.$$

We introduce some additional notation. Let $D(L) = C(L)A(L)$, $W(r)$ be the $p \times 1$ Bm with covariance matrix Ω,

$$\overline{W}(r) = 2\pi\zeta'(D(1)^{-1})'\Omega^{-1}W(r),$$

and define the $l_1 \times 1$ column vector

$$\hat{W}(r,\bar{\delta}_1) = \left[(0,i_j')\frac{1}{\Gamma(\delta_j)}\int_0^r (r-s)^{\delta_j-1}D(1)\,dW(s)\right]_{j=1,\dots,l_1},$$

where i_j is an $l \times 1$ vector of zeros except 1 in the jth position. Denoting by $f_{ij}(\lambda)$ the (i,j)th element of $f(\lambda)$, let

$$\bar{f}_{22}(0) = \left[\frac{f_{ij}(0)}{1-\delta_i-\delta_j}\right]_{\substack{i=l_1+1,\dots,l \\ j=l_1+1,\dots,l}};$$

finally set

$$\Lambda(n) = diag(n^{-\delta_1},\dots,n^{-\delta_{l_1}}, m^{-1/2}\lambda_m^{\delta_{l_1+1}},\dots,m^{-1/2}\lambda_m^{\delta_l}).$$

We next present the null limiting distribution of the statistic $\hat{\tau}_m(\hat{\gamma},\hat{\bar{\delta}},\hat{p})$.

THEOREM 1. *Under Assumptions 1–3, 5, 6 and H_0, as $n \to \infty$,*

$$\Lambda(n)\hat{\tau}_m(\hat{\gamma},\hat{\bar{\delta}},\hat{p}) \Rightarrow \begin{pmatrix} \int_0^1 \hat{W}(r;\bar{\delta}_1)\,d\overline{W}(r) \\ Z \end{pmatrix}, \quad \textbf{(24)}$$

where Z is an $(l-l_1)\times 1$ vector of random variables normally distributed with $E(Z) = 0$ and

$$\mathrm{Var}(Z) = \frac{1}{2}q(0)\bar{f}_{22}(0),$$

which for all r is independent from $\hat{W}(r;\bar{\delta}_1)$ and $\overline{W}(r)$. Under Assumptions 1, 2, 4, 5, 7 and H_0, an identical result to (24) applies for $\hat{\tau}_m^0(\hat{\gamma},\hat{\bar{\delta}},\hat{p})$.

The proof of Theorem 1 is given in the Appendix. Denoting by $\hat{\Upsilon}_m^*$ any of the $\hat{\Upsilon}_m(\hat{\gamma},\hat{\bar{\delta}},\hat{p})$, $\hat{\Upsilon}_m^0(\hat{\gamma},\hat{\bar{\delta}},\hat{p})$, we have the following corollary, which is a straightforward consequence of Theorem 1.

COROLLARY 1. *Under the conditions of Theorem 1 and H_0, as $n \to \infty$, $\hat{\Upsilon}_m^* \to_d \chi_l^2$.*

Remark 1. The distribution on the right of (24) is mixed normal because $\hat{W}(r;\bar{\delta})$ and $\overline{W}(r)$ are not correlated; hence the corresponding test statistics have chi square null asymptotic distribution.

Remark 2. In the most important case emphasized in the literature, that is, when all the observables share the same integration order (see, e.g., RY; MV), we could simplify our statistic substantially by replacing the process $z_t(\hat{\gamma},\hat{\bar{\delta}})$ by simply $z_t(\hat{\gamma})$, where the same filtering is applied to all the observables. Note that in this case $z_t(\hat{\gamma})$ is a good proxy for u_t under H_0. However, if we allow the integration orders to vary across the components of z_t, $z_t(\hat{\gamma})$ would have some overdifferenced components under H_0, and the orthogonalization in (19) with $g(\lambda) = \hat{p}(\lambda)$ would not be correct. This problem is avoided by considering $z_t(\hat{\gamma},\hat{\bar{\delta}})$ instead, noting that this modification should not imply any loss of power because under H_1, $y_t(\hat{\gamma})$ is underdifferenced, and so it is the leading component in $z_t(\hat{\gamma},\hat{\bar{\delta}})$.

Remark 3. The "zero-frequency" statistic has a direct interpretation relative to that proposed for the white noise situation. Clearly

$$\hat{\tau}^0_{[n/2]}(\gamma,\bar{\delta},\hat{p}) = \frac{1}{\hat{\omega}_{y.x}} \sum_{t=1}^{n} x_t(y_t(\gamma) - \hat{f}_{yx}(0)\hat{f}_{xx}^{-1}(0)\Delta_t(\bar{\delta})x_t),$$

although, in view of our assumptions, this statistic enjoys nice properties only when $\delta_i > 1$, $i = 0,\dots,l$, because otherwise the incorrect treatment of the short-memory components by weighting only at frequency zero heavily distorts its asymptotic behavior.

For the analysis of the parametric test we set the following conditions.

Assumption 8.

(a) Assumption 1 holds;
(b) $f(\lambda;h)$ satisfies the technical smoothness conditions imposed in Assumption 2 of RH;
(c) there exist estimates $\hat{\gamma}$, $\hat{\theta}$ of γ, θ, respectively, for which Assumption 2 (for $\hat{\gamma}$, without the need of (22)) holds and

$$\hat{\theta} = \theta + O_p(n^{-1/2});$$

(d) $\delta_i = \delta$ for all $i \in \{0,1,\dots,l\}$, where $\delta > \frac{1}{2}$.

Thus, we have the following theorem, whose proof is given in the Appendix, and a corollary, which is a straightforward consequence of Theorem 2.

THEOREM 2. *Under Assumption 8 and H_0 an equivalent result to (24) holds for $\tilde{\tau}_n(\hat{\gamma},\hat{\theta})$.*

COROLLARY 2. *Under the conditions of Theorem 2 and H_0, as $n \to \infty$,*

$\tilde{\Upsilon}_n \to_d \chi^2_l$.

Remark 4. Theorem 2 uncovers another important limitation of the parametric setting, because the result is only given for the $\delta > \frac{1}{2}$ case and so one should know a priori the (purely) nonstationary condition of the observables. When $\delta < \frac{1}{2}$, it can be shown that under our conditions $n^{-1/2}\tilde{\tau}_n(\gamma, \theta)$ is asymptotically normal, but even if $\hat{\theta}$ is $\sqrt{n}$-consistent the asymptotic distribution of the properly normalized statistic $\tilde{\tau}_n(\gamma, \hat{\theta})$ differs from that of the infeasible one, unlike in the semiparametric setting where feasible and infeasible statistics share the same limiting distribution. Here, it should be possible to determine that $n^{-1/2}\tilde{\tau}_n(\hat{\gamma}, \hat{\theta})$ is asymptotically normal, but the asymptotic variance of the normalized statistic is not the same as in the case where θ is known and depends on the particular form of $\hat{\theta}$. Dealing appropriately with the $\delta < \frac{1}{2}$ case was the precise reason why x_{t-1} replaces (the more natural) x_t in (13), but in our general setting the problem of calculating the limiting distribution of the normalized feasible statistic is complicated, because $\hat{\theta}$ is generally implicitly defined, although letting u_t be a finite vector autoregressive (VAR) process the task is simpler (see Hualde and Robinson, 2007).

Remark 5. Theorem 2 uses results from RH, the main distinguishing feature now being that the requirement on the estimate of the order δ (under H_0) is much less stringent than in RH. In particular, RH derived a related result under the condition (translated to our framework) that $\kappa > \max(0, 1-\delta)$, and so almost $\sqrt{n}$-consistency of $\hat{\gamma}$ was needed in case δ were just above $\frac{1}{2}$. This assumption was unavoidable in RH's framework, but exploiting our particular orthogonalization, we manage to avoid this requirement in the present setting. This relaxation is not trivial, because the theory for estimating parametrically (hence obtaining $\sqrt{n}$-consistent estimates) long- and short-memory parameters simultaneously in a multivariate setting is only fully developed in the stationary case. More importantly, γ necessarily needs to be estimated from a sort of residual (like OLS residuals), and so it is unclear in which sense one can base parametric estimates of γ on these residuals.

Remark 6. Note that in the semiparametric case we need some extra requirements (given in (22)) on the convergence rates of the estimates of the orders apart from $\kappa > 0$. This is due to the nonunique differencing applied to the observables in the test statistics, an issue that also arises in Hualde and Robinson (2005).

6. MONTE CARLO EVIDENCE

To offer some evidence of the finite-sample behavior of these test procedures, we present a small Monte Carlo experiment. There are two parts in our study, the first comparing the performance of semiparametric and parametric versions of our test in the simple bivariate situation where the error input process w_t is white noise, and the second focusing on the semiparametric case with correlated w_t and with three observables whose orders could possibly differ. In the first part of the study, we generated a univariate process x_t of lengths $n = 64$, 128, 256, 512, 1,024, as in (5) for the different values of $\delta = 0.3, 0.6, 1, 1.4$, and y_t as in (4) (for the same lengths as x_t) with $\beta = 1$ and γ taking four different values for each corresponding δ, which are

$$\gamma = \delta, \delta - 0.2, \delta - 0.4, \delta - 0.6,$$

except for $\delta = 0.3$, where $\gamma = 0.3, 0.2, 0.1, 0$, the first value representing in all cases the situation of absence of cointegration. The error input process w_t was generated as a mean-zero bivariate Gaussian white noise with a covariance structure leading to a white noise u_t with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with $\rho = 0.5$, noting that in view of the Monte Carlo results of Hualde and Robinson (2004), the tests are expected to behave in a better (worse) way as $|\rho|$ decreases (increases), being relatively unaffected by the sign of ρ. The parametric test statistic was computed following these steps.

1. Estimate δ from the raw series x_t as in Beran (1995), fixing the optimizing interval $[\delta - 1, \delta + 1]$.
2. Compute the NB estimate for β (see, e.g., Robinson and Marinucci, 2001), choosing bandwidths $m = 25, 40, 65, 120, 220$, for $n = 64, 128, 256, 512, 1{,}024$, respectively. Note that the OLS residuals are adequate to estimate γ under H_1 if $\delta > 0.5$, but if $\delta < 0.5$, the OLS estimate is in general inconsistent.
3. Estimate γ with the NB residuals by the same procedure as in step 1, optimizing over the interval $[\gamma - 1, \gamma + 1]$. Note that both intervals are infeasible but in practice their length could be adequate.
4. Compute $\hat{\omega}_{y.x}$ using corresponding estimates of components of the spectral density matrix (at frequency zero) calculated as in (16), noting (10) and (11), and estimate η by $\tilde{\eta}$ (see (18)).
5. Compute the feasible test statistic $\Xi_n(\hat{\omega}_{y.x}, \hat{\gamma}, \tilde{\eta})$.

We compare the behavior of Ξ_n with two semiparametric versions of the test, one where the GLS weighting referred to before is evaluated at all Fourier frequencies, the other weighting only frequency zero. Given that in the present setting u_t is a white noise process, both test statistics are expected to enjoy similar behavior, and this is corroborated by the results of our experiment. To calculate the semiparametric statistics, we use the following steps.

1.′ Estimate δ and γ by $\hat{\delta}$ and $\hat{\gamma}$, by the Robinson (1995b) versions of the log-periodogram of Geweke and Porter-Hudak (1983) (with bandwidths m given in step 2), without trimming or pooling applied to the series $\bar{x}_t$, $\bar{\hat{v}}_t$, where $\bar{x}_t = x_t 1(\delta < 1) + \Delta x_t 1(\delta \geq 1)$, $\bar{\hat{v}}_t = \hat{v}_t 1(\delta < 1) + \Delta \hat{v}_t 1(\delta \geq 1)$, denoting by $\hat{v}_t$ the NB residuals and adding back one to the estimates of the orders when the corresponding differenced series are employed.

2.′ Compute the unweighted estimate of $f(\lambda)$,

$$\hat{f}(\lambda_j) = \frac{1}{2m+1} \sum_{k=j-m}^{j+m} I_{z(\hat{\delta})}(\lambda_k).$$

Note that we used here the same bandwidth as for the estimates of the orders and m will be the corresponding bandwidths used for the semiparametric statistics.

3.′ Compute the following slightly modified versions of $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, which exploit the bivariate framework and add power. The only modification affects $\hat{\tau}_m$, $\hat{\tau}_m^0$ in (19), because instead of these statistics we compute

$$\tilde{\tau}_m = \sum_{j=0}^{m} s_j \frac{\operatorname{Re}\left\{ I_{y(\hat{\gamma})x}(\lambda_j) - \dfrac{\tilde{f}_{12}(\lambda_j)}{\hat{f}_{22}(\lambda_j)} I_{x(\hat{\gamma})x}(\lambda_j) \right\}}{\hat{f}_{11}(\lambda_j) - \dfrac{\hat{f}_{12}(\lambda_j)\hat{f}_{21}(\lambda_j)}{\hat{f}_{22}(\lambda_j)}},$$

$$\tilde{\tau}_m^0 = \sum_{j=0}^{m} s_j \frac{\operatorname{Re}\left\{ I_{y(\hat{\gamma})x}(\lambda_j) - \dfrac{\tilde{f}_{12}(0)}{\hat{f}_{22}(0)} I_{x(\hat{\gamma})x}(\lambda_j) \right\}}{\hat{f}_{11}(0) - \dfrac{\hat{f}_{12}(0)\hat{f}_{21}(0)}{\hat{f}_{22}(0)}},$$

where $\hat{f}_{ij}$ is the (i,j)th element of $\hat{f}$ and

$$\tilde{f}_{12}(\lambda_j) = \frac{1}{2m+1} \sum_{k=j-m}^{j+m} I_{y(\hat{\gamma})x(\hat{\gamma})}(\lambda_k),$$

which diverges under H_1, this being the source of additional power.

Results of the proportion of rejections over 10,000 replications when comparing the values of the statistic with the $\alpha = 0.01$, 0.05, 0.10 nominal critical values of the χ_1^2 distribution are reported in Tables 1–4 for the differ-

Table 1. Proportion of rejections of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, bivariate case, $\delta = 0.3$

	γ	0.3			0.2			0.1			0		
	n/α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Ξ_n	64	0.015	0.059	0.113	0.030	0.087	0.149	0.073	0.162	0.245	0.141	0.263	0.357
	128	0.012	0.056	0.108	0.032	0.109	0.184	0.098	0.245	0.355	0.190	0.373	0.484
	256	0.011	0.052	0.103	0.065	0.191	0.290	0.242	0.470	0.595	0.347	0.552	0.652
	512	0.012	0.055	0.104	0.166	0.362	0.487	0.594	0.804	0.879	0.614	0.767	0.826
	1,024	0.009	0.053	0.101	0.396	0.642	0.754	0.931	0.981	0.993	0.843	0.907	0.932
$\hat{\Upsilon}_m$	64	0.036	0.051	0.068	0.090	0.118	0.141	0.191	0.237	0.269	0.331	0.382	0.417
	128	0.037	0.060	0.083	0.128	0.173	0.208	0.304	0.371	0.412	0.542	0.607	0.648
	256	0.043	0.076	0.111	0.196	0.257	0.304	0.490	0.567	0.611	0.790	0.839	0.863
	512	0.030	0.063	0.099	0.223	0.306	0.364	0.653	0.730	0.769	0.943	0.960	0.969
	1,024	0.029	0.066	0.109	0.295	0.398	0.465	0.856	0.899	0.920	0.996	0.998	0.999
$\hat{\Upsilon}_m^0$	64	0.039	0.056	0.072	0.097	0.130	0.154	0.207	0.257	0.290	0.352	0.410	0.446
	128	0.046	0.073	0.099	0.152	0.204	0.238	0.344	0.416	0.464	0.591	0.660	0.696
	256	0.058	0.100	0.140	0.232	0.301	0.354	0.548	0.623	0.666	0.839	0.879	0.901
	512	0.044	0.085	0.133	0.276	0.366	0.425	0.721	0.788	0.820	0.964	0.977	0.982
	1,024	0.044	0.096	0.151	0.375	0.477	0.542	0.900	0.934	0.951	0.998	0.999	0.999

Note: Proportion of rejections over 10,000 replications of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, when compared with the critical value of a χ_1^2 distribution with nominal size α. For $\gamma = \delta$ this is simulated size and for $\gamma < \delta$ simulated power.

TABLE 2. Proportion of rejections of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, bivariate case, $\delta = 0.6$

	γ	0.6			0.4			0.2			0		
	n/α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Ξ_n	64	0.035	0.098	0.159	0.111	0.205	0.284	0.379	0.488	0.564	0.691	0.754	0.789
	128	0.028	0.085	0.142	0.120	0.274	0.380	0.468	0.591	0.660	0.855	0.886	0.905
	256	0.022	0.070	0.124	0.246	0.469	0.594	0.565	0.661	0.714	0.956	0.966	0.971
	512	0.018	0.067	0.118	0.550	0.766	0.846	0.643	0.719	0.765	0.996	0.997	0.997
	1,024	0.015	0.061	0.113	0.889	0.964	0.981	0.757	0.809	0.835	1.00	1.00	1.00
$\hat{\Upsilon}_m$	64	0.051	0.088	0.130	0.198	0.242	0.286	0.540	0.589	0.623	0.844	0.874	0.887
	128	0.050	0.092	0.137	0.277	0.338	0.390	0.780	0.816	0.837	0.981	0.986	0.988
	256	0.051	0.100	0.149	0.407	0.477	0.523	0.949	0.961	0.968	1.00	1.00	1.00
	512	0.034	0.082	0.132	0.496	0.575	0.625	0.997	0.998	0.999	1.00	1.00	1.00
	1,024	0.028	0.078	0.133	0.649	0.728	0.769	1.00	1.00	1.00	1.00	1.00	1.00
$\hat{\Upsilon}_m^0$	64	0.053	0.092	0.136	0.205	0.252	0.293	0.555	0.607	0.638	0.859	0.885	0.900
	128	0.056	0.103	0.150	0.296	0.360	0.410	0.801	0.833	0.855	0.985	0.989	0.991
	256	0.057	0.112	0.163	0.431	0.502	0.551	0.957	0.968	0.974	1.00	1.00	1.00
	512	0.039	0.093	0.143	0.526	0.605	0.654	0.998	0.999	0.999	1.00	1.00	1.00
	1,024	0.033	0.088	0.147	0.684	0.755	0.792	1.00	1.00	1.00	1.00	1.00	1.00

Note: Proportion of rejections over 10,000 replications of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, when compared with the critical value of a χ_1^2 distribution with nominal size α. For $\gamma = \delta$ this is simulated size and for $\gamma < \delta$ simulated power.

Table 3. Proportion of rejections of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, bivariate case, $\delta = 1$

	γ	1			0.8			0.6			0.4		
	n/α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Ξ_n	64	0.052	0.118	0.188	0.099	0.175	0.245	0.374	0.463	0.528	0.752	0.795	0.826
	128	0.037	0.094	0.159	0.071	0.169	0.253	0.402	0.501	0.576	0.885	0.908	0.921
	256	0.023	0.077	0.137	0.077	0.206	0.301	0.447	0.549	0.616	0.970	0.977	0.981
	512	0.019	0.069	0.124	0.121	0.287	0.403	0.520	0.612	0.670	0.998	0.999	0.999
	1,024	0.014	0.061	0.114	0.210	0.418	0.543	0.633	0.712	0.747	1.00	1.00	1.00
$\hat{\Upsilon}_m$	64	0.048	0.103	0.152	0.179	0.245	0.304	0.498	0.548	0.586	0.842	0.866	0.879
	128	0.042	0.094	0.147	0.181	0.245	0.304	0.663	0.707	0.735	0.971	0.980	0.984
	256	0.038	0.089	0.138	0.204	0.271	0.327	0.840	0.869	0.885	0.999	0.999	0.999
	512	0.029	0.076	0.129	0.181	0.255	0.316	0.953	0.965	0.971	1.00	1.00	1.00
	1,024	0.025	0.075	0.127	0.176	0.260	0.331	0.995	0.997	0.998	1.00	1.00	1.00
$\hat{\Upsilon}_m^0$	64	0.049	0.102	0.152	0.182	0.247	0.307	0.500	0.554	0.593	0.849	0.870	0.884
	128	0.043	0.095	0.146	0.184	0.249	0.311	0.668	0.713	0.739	0.973	0.982	0.985
	256	0.038	0.090	0.139	0.208	0.276	0.333	0.844	0.872	0.888	0.999	0.999	0.999
	512	0.029	0.077	0.131	0.183	0.260	0.322	0.954	0.965	0.971	1.00	1.00	1.00
	1,024	0.024	0.074	0.128	0.178	0.265	0.336	0.995	0.997	0.998	1.00	1.00	1.00

Note: Proportion of rejections over 10,000 replications of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, when compared with the critical value of a χ_1^2 distribution with nominal size α. For $\gamma = \delta$ this is simulated size and for $\gamma < \delta$ simulated power.

TABLE 4. Proportion of rejections of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, bivariate case, $\delta = 1.4$

	γ	1.4			1.2			1			0.8		
	n/α	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Ξ_n	64	0.046	0.106	0.176	0.070	0.118	0.174	0.288	0.350	0.409	0.660	0.697	0.729
	128	0.030	0.087	0.150	0.037	0.086	0.149	0.287	0.350	0.404	0.779	0.805	0.825
	256	0.020	0.072	0.134	0.026	0.085	0.140	0.309	0.370	0.423	0.889	0.907	0.916
	512	0.018	0.065	0.125	0.030	0.099	0.163	0.346	0.412	0.465	0.970	0.975	0.979
	1,024	0.014	0.057	0.109	0.052	0.138	0.208	0.407	0.471	0.512	0.995	0.996	0.996
$\hat{\Upsilon}_m$	64	0.052	0.113	0.164	0.080	0.124	0.174	0.278	0.325	0.368	0.771	0.798	0.816
	128	0.041	0.092	0.142	0.084	0.126	0.174	0.404	0.452	0.490	0.898	0.913	0.924
	256	0.031	0.078	0.131	0.091	0.145	0.196	0.556	0.606	0.642	0.973	0.978	0.980
	512	0.024	0.074	0.121	0.079	0.135	0.190	0.677	0.721	0.746	0.996	0.997	0.998
	1,024	0.022	0.071	0.123	0.075	0.138	0.197	0.780	0.813	0.835	1.00	1.00	1.00
$\hat{\Upsilon}_m^0$	64	0.053	0.113	0.165	0.080	0.124	0.175	0.278	0.327	0.370	0.772	0.799	0.818
	128	0.041	0.091	0.141	0.084	0.126	0.176	0.404	0.451	0.491	0.898	0.914	0.924
	256	0.030	0.079	0.131	0.092	0.145	0.199	0.554	0.607	0.640	0.973	0.977	0.980
	512	0.024	0.073	0.121	0.080	0.137	0.192	0.675	0.722	0.748	0.995	0.996	0.997
	1,024	0.022	0.070	0.122	0.075	0.139	0.199	0.779	0.814	0.835	1.00	1.00	1.00

Note: Proportion of rejections over 10,000 replications of Ξ_n, $\hat{\Upsilon}_m$, $\hat{\Upsilon}_m^0$, when compared with the critical value of a χ_1^2 distribution with nominal size α. For $\gamma = \delta$ this is simulated size and for $\gamma < \delta$ simulated power.

ent values of δ. Overall all the three tests are oversized, but in almost all cases sizes react appropriately as n increases. As expected, in terms of size, the parametric test performs best, followed by $\hat{\Upsilon}_m$, with $\hat{\Upsilon}_m^0$ being worst, although for cases $\delta = 1, 1.4$, its behavior is very similar to that of $\hat{\Upsilon}_m$. For the parametric test, sizes are best for $\delta = 0.3$, and it also shows a better performance for the mean-reverting case ($\delta = 0.6$) than for the non-mean-reverting ones ($\delta = 1, 1.4$), the case $\delta = 1$ being worst. The mean-reverting and (asymptotically) stationary cases also favor the semiparametric statistics in terms of size when n is small, but here especially the proportions of rejections corresponding to $\hat{\Upsilon}_m^0$ do not show a very clear convergence pattern to the nominal sizes as n increases, unlike in the $\delta = 1, 1.4$ cases, where it appears to be a monotonic convergence toward the nominal values. Clearly, the parametric test is most powerful for $\delta = 0.6$, with a very good performance relative to other values of δ for which the reduction of the order of the observables under H_1 is just 0.2 (where indeed the increase of the proportion of rejections as n increases could be very slow). This perhaps indicates that the jump from nonstationary observables to (asymptotically) stationary cointegrating errors (which does not appear for other combinations of δ, $\gamma = \delta - 0.2$) is important. In terms of power, the semiparametric statistics are comparable to the parametric one (although note that the proportions of rejections are not size corrected). Similarly to the parametric test, the semiparametric tests have also problems detecting the alternative when $\delta = 1, 1.4$ (but not when $\delta = 0.6$) and $\gamma = \delta - 0.2$, the proportion of rejections being higher here than for the parametric test when n is small, although increasing at a slower rate as n increases. In almost all cases the proportions of rejection react appropriately as n and the cointegrating gap $(\delta - \gamma)$ increase.

In the second part of the experiment, we analyze the behavior of the semiparametric statistic $\hat{\Upsilon}_m$, in a multivariate framework (with three observables), with possibly different integration orders. We generated ε_t (see Assumption 1) as a trivariate zero-mean Gaussian white noise with covariance matrix

$$\Omega = \begin{pmatrix} 1.5 & -0.75 & -0.75 \\ -0.75 & 1 & 0.25 \\ -0.75 & 0.25 & 1 \end{pmatrix},$$

noting that if $A(L) = I_3$ in Assumption 1, this covariance structure leads to a white noise u_t with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.25 \\ 0.5 & 0.25 & 1 \end{pmatrix},$$

and so the scenario is similar to that described in the first part of the experiment. However, we introduced further short-memory structure to our design by setting $A^{-1}(L) = I_3 - \Phi L$, $\Phi = diag(0.5, 0.5, 0.5)$ and generating w_t accordingly. Then, denoting $x_t = (x_{1t}, x_{2t})'$, $u_{xt} = (u_{x1,t}, u_{x2,t})'$, x_{1t}, x_{2t} are generated from the input processes $u_{x1,t}$, $u_{x2,t}$, as fractionally integrated processes of orders δ_1, δ_2, respectively, where $\delta_1 = 1.4, 1$ and $\delta_2 = \delta_1, \delta_1 - 0.2, \delta_1 - 0.4$. Finally, y_t was generated as in (4) with $\beta = (1,1)'$ and

$$\gamma = \delta_1, \delta_1 - 0.2, \delta_1 - 0.4.$$

Note that in the present setting, especially when $\delta_2 < \delta_1$, the covariance structure of u_t is very distant from that of the white noise situation, and so the use of $\hat{\Upsilon}_m$ instead of $\hat{\Upsilon}_m^0$ seems more appropriate, although both statistics are, at least to first-order properties, asymptotically equivalent. To compute the statistic and the estimates of the nuisance parameters, we employed the set of bandwidths $m = 12, 20, 31, 60, 110$, for $n = 64, 128, 256, 512, 1{,}024$, respectively. Note that these bandwidths are approximately half of the bandwidths used in the first part of the experiment, and they were chosen on the observation that when short-memory structure is present, smaller bandwidths than in the white noise situation are warranted. The orders δ_0, δ_1, δ_2 are estimated by $\hat{\delta}_0$, $\hat{\delta}_1$, $\hat{\delta}_2$, which are calculated by the same procedure described in the first part of the experiment, from the series y_t, x_{1t}, x_{2t}, respectively. Also

$$\hat{f}(\lambda_j) = \frac{1}{m+1} \sum_{k=j-m/2}^{j+m/2} I_{z(\hat{\underline{\delta}})}(\lambda_k),$$

where $\hat{\underline{\delta}} = (\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2)'$. Here, note that on the estimation of f we chose bandwidth $m/2$, because, especially when $\delta_2 < \delta_1$, the estimation of f at a particular frequency rapidly gets distorted when incorporating information from frequencies that are relatively far from this particular frequency. As in any semiparametric procedure, the choice of bandwidth is fundamental, and although a more extensive Monte Carlo experiment checking the sensitivity of the test to variations of all bandwidths (and indeed of the short-memory parameters) involved is interesting, our proposed bandwidths give general hints to practitioners on possible choices that behave relatively well.

As mentioned in Section 5, the main issue here is to estimate γ. Following the strategy described there, we compute the NB estimate from the regression of Δy_t on Δx_t and estimate γ from residuals $y_t - \hat{\beta}' x_t$ by the method described in the first part of the experiment, finally obtaining $\hat{\Upsilon}_m$. Our results are presented in Tables 5 and 6. As in the bivariate case, when n is small our test is clearly oversized (especially for $\delta_1 = 1.4$), although as n increases size reacts in the appropriate direction, and so they are very close to the nominal ones when $n = 1{,}024$ (the statistic being in some cases undersized for large n). For small n, a decrease in δ_2 implies a decrease in size, this effect disappearing as n

TABLE 5. Proportion of rejections of $\hat{\Upsilon}_m$, trivariate case, $\delta_1 = 1$

		$\gamma = 1$			$\gamma = 0.8$			$\gamma = 0.6$		
	δ_2	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
64	1	0.132	0.183	0.222	0.427	0.500	0.546	0.704	0.762	0.795
	0.8	0.071	0.106	0.140	0.292	0.362	0.415	0.598	0.669	0.706
	0.6	0.048	0.078	0.109	0.213	0.269	0.317	0.532	0.610	0.650
128	1	0.089	0.135	0.171	0.497	0.577	0.623	0.843	0.882	0.900
	0.8	0.038	0.067	0.096	0.313	0.395	0.450	0.744	0.804	0.832
	0.6	0.029	0.055	0.081	0.236	0.312	0.364	0.696	0.762	0.794
256	1	0.071	0.112	0.153	0.601	0.678	0.724	0.941	0.961	0.970
	0.8	0.026	0.054	0.086	0.415	0.503	0.560	0.884	0.917	0.932
	0.6	0.021	0.044	0.072	0.310	0.400	0.457	0.868	0.905	0.923
512	1	0.039	0.074	0.110	0.720	0.798	0.838	0.990	0.995	0.997
	0.8	0.016	0.046	0.083	0.504	0.610	0.668	0.974	0.986	0.990
	0.6	0.011	0.036	0.073	0.379	0.494	0.564	0.967	0.982	0.986
1,024	1	0.019	0.048	0.080	0.849	0.906	0.933	1.00	1.00	1.00
	0.8	0.009	0.040	0.084	0.642	0.744	0.800	0.998	0.999	0.999
	0.6	0.010	0.047	0.094	0.488	0.629	0.703	0.998	0.999	0.999

Note: Proportion of rejections over 10,000 replications of $\hat{\Upsilon}_m(\hat{\gamma},\hat{\hat{\delta}},\hat{p})$ when compared with the critical value of a χ^2_2 distribution with nominal size α. For $\gamma = \delta_1$ this is simulated size and for $\gamma < \delta_1$ simulated power.

Table 6. Proportion of rejections of $\hat{\Upsilon}_m$, trivariate case, $\delta_1 = 1.4$

		$\gamma = 1.4$			$\gamma = 1.2$			$\gamma = 1$		
	δ_2	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
64	1.4	0.168	0.226	0.271	0.464	0.534	0.579	0.738	0.789	0.816
	1.2	0.122	0.176	0.219	0.341	0.405	0.450	0.644	0.704	0.740
	1	0.101	0.149	0.190	0.270	0.337	0.387	0.585	0.650	0.692
128	1.4	0.113	0.163	0.204	0.509	0.583	0.627	0.850	0.889	0.906
	1.2	0.073	0.118	0.157	0.339	0.420	0.467	0.750	0.805	0.833
	1	0.070	0.114	0.151	0.276	0.353	0.405	0.704	0.764	0.796
256	1.4	0.085	0.135	0.179	0.589	0.661	0.705	0.927	0.950	0.960
	1.2	0.058	0.100	0.146	0.397	0.478	0.534	0.863	0.899	0.917
	1	0.048	0.089	0.130	0.326	0.408	0.468	0.836	0.878	0.899
512	1.4	0.046	0.086	0.124	0.650	0.727	0.769	0.963	0.973	0.980
	1.2	0.034	0.068	0.108	0.434	0.526	0.583	0.924	0.949	0.960
	1	0.031	0.067	0.103	0.369	0.465	0.523	0.910	0.939	0.952
1,024	1.4	0.023	0.053	0.085	0.741	0.807	0.843	0.979	0.986	0.989
	1.2	0.017	0.047	0.079	0.497	0.599	0.654	0.957	0.970	0.978
	1	0.016	0.048	0.083	0.416	0.522	0.586	0.953	0.968	0.975

Note: Proportion of rejections over 10,000 replications of $\hat{\Upsilon}_m(\hat{\gamma},\hat{\hat{\delta}},\hat{p})$ when compared with the critical value of a χ^2_2 distribution with nominal size α. For $\gamma = \delta_1$ this is simulated size and for $\gamma < \delta_1$ simulated power.

increases. In terms of power, the test behaves in a very similar way for both the $\delta_1 = 1$ and $\delta_1 = 1.4$ cases. Power decreases with δ_2 and reacts appropriately as n increases. The test is able to detect the alternative $\gamma = \delta_1 - 0.2$ now, although, especially when $\delta_2 = \delta_1 - 0.4$, the increase in power as n increases is slow. Overall, we find that the results in this second part of the experiment, which describe a more realistic situation than the first one, are certainly encouraging, noting that for simplicity we neither applied the provision made in Section 5 about the rate of growth of the bandwidth used in the estimation of γ (in comparison to that used in the estimation of β) nor used sophisticated estimates of the nuisance parameters. In fact, estimation procedures of these parameters using bias-reducing devices are readily available (see Hualde and Robinson, 2004), and using them might lead to even better finite-sample results.

REFERENCES

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society, Series B* 57, 659–672.

Breitung, J. & U. Hassler (2002) Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics* 110, 167–185.

Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Chen, W.W. & C.M. Hurvich (2003) Semiparametric estimation of multivariate fractional cointegration. *Journal of the American Statistical Association* 463, 629–642.

Chen, W.W. & C.M. Hurvich (2006) Semiparametric estimation of fractional cointegrating subspaces. *Annals of Statistics* 34, 2939–2979.

Dueker, M. & R. Startz (1998) Maximum-likelihood estimation of fractional cointegration with an application to U.S.and Canadian bond rates. *Review of Economics and Statistics* 80, 420–426.

Dunsmuir, W. (1979) A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. *Annals of Statistics* 7, 490–506.

Dunsmuir, W. & E.J. Hannan (1976) Vector linear time series models. *Advances in Applied Probability* 8, 339–364.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.

Flôres, R.G. & A. Szafarz (1996) An enlarged definition of cointegration. *Economics Letters* 50, 193–195.

Geweke, J. & S. Porter-Hudak (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.

Gil-Alaña, L.A. (2003) Multivariate tests of fractionally integrated hypothesis. *South African Statistical Journal* 37, 1–28.

Hassler, U. & J. Breitung (2006) A residual-based LM-type test against fractional cointegration. *Econometric Theory* 22, 1091–1111.

Hassler, U., F. Marmol, & C. Velasco (2006) Residual log-periodogram inference for long-run relationships. *Journal of Econometrics* 130, 165–207.

Hualde, J. & P.M. Robinson (2004) Semiparametric Estimation of Fractional Cointegration. Preprint, Universidad de Navarra.

Hualde, J. & P.M. Robinson (2005) Semiparametric Inference in Multivariate Fractionally Cointegrated Systems. Preprint, Universidad de Navarra.

Hualde, J. & P.M. Robinson (2007) Root-n-consistent estimation of weak fractional cointegration. *Journal of Econometrics* 140, 450–484.

Johansen, S. (1988) Statistical analysis of co-integration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Kim, C.S. & P.C.B. Phillips (2000) Fully Modified Estimation of Fractional Co-integration Models. Preprint, Yale University.

Künsch, H.R. (1987) Statistical aspects of self-similar processes. In Y.A. Prohorov & V.V. Sazonov (eds.), *Proceedings of the 1st World Congress of the Bernoulli Society*, pp. 67–74. VNU Science Press.

Kurtz, T.G. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.

Lasak, K. (2005) Likelihood Based Testing for Fractional Cointegration. Preprint, Universitat Autònoma de Barcelona.

Lobato, I.G. (1999) A semiparametric two-step estimation on a multivariate long memory model. *Journal of Econometrics* 90, 129–153.

Marinucci, D. & P.M. Robinson (1999) Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* 80, 111–122.

Marinucci, D. & P.M. Robinson (2000) Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* 86, 103–120.

Marinucci, D. & P.M. Robinson (2001) Semiparametric fractional cointegration analysis. *Journal of Econometrics* 105, 225–247.

Marmol, F. & C. Velasco (2004) Consistent testing of cointegrating relationships. *Econometrica* 72, 1809–1844.

Park, J.Y., S. Ouliaris, & B. Choi (1988) Spurious Regressions and Tests for Cointegration. CAE Working paper 88–07, Cornell University.

Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420–1437.

Robinson, P.M. (1995a) Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* 23, 1630–1661.

Robinson, P.M. (1995b) Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23, 1048–1072.

Robinson, P.M. (2005a) The distance between rival nonstationary fractional processes. *Journal of Econometrics* 128, 283–300.

Robinson, P.M. (2005b) Diagnostic Testing for Cointegration. Preprint, London School of Economics.

Robinson, P.M. & J. Hualde (2003) Cointegration in fractional systems with unknown integration orders. *Econometrica* 71, 1727–1766.

Robinson, P.M. & F. Iacone (2005) Cointegration in fractional systems with deterministic trends. *Journal of Econometrics* 129, 263–298.

Robinson, P.M. & D. Marinucci (2001) Narrow-band analysis of nonstationary processes. *Annals of Statistics* 29, 947–986.

Robinson, P.M. & D. Marinucci (2003) Semiparametric frequency-domain analysis of fractional cointegration. In P.M. Robinson (ed.), *Time Series with Long Memory*, pp. 334–373. Oxford University Press.

Robinson, P.M. & Y. Yajima (2002) Determination of cointegrating rank in fractional systems. *Journal of Econometrics* 106, 217–241.

Velasco, C. (2003) Gaussian semiparametric estimation of fractional cointegration. *Journal of Time Series Analysis* 24, 345–378.

Velasco, C. (2004) The Periodogram of Fractional Processes. Preprint, Universidad Carlos III de Madrid.

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

APPENDIX

Proof of Theorem 1. The proof follows on showing that

$$\Lambda(n)\hat{\tau}_m(\gamma,\delta,p) \Rightarrow \begin{pmatrix} \int_0^1 \hat{W}(r;\bar{\delta}_1)\, d\overline{W}(r) \\ Z \end{pmatrix}, \tag{A.1}$$

$$\Lambda(n)(\hat{\tau}_m(\hat{\gamma},\hat{\delta},\hat{p}) - \hat{\tau}_m(\gamma,\delta,p)) = o_p(1). \tag{A.2}$$

First, (A.2) follows directly from Propositions 2 and 3 of Hualde and Robinson (2005). Next, the proof of (A.1) follows from the application of the steps in the proof of Proposition 1 of Hualde and Robinson (2005) to a somewhat different framework. Denoting by $x_t^{(1)}$ the $l_1 \times 1$ vector of the first l_1 components of x_t (the purely nonstationary ones), it can be easily shown that

$$\Lambda(n)\hat{\tau}_m(\gamma,\delta,p) = \Lambda(n)\sum_{t=2}^{n}\left(x_{t-1}^{(1)\prime}p(0)D(1)\varepsilon_t, \varepsilon_t' \sum_{v=1}^{t-1} c_{t-v}^{l_1+1}\varepsilon_v, \ldots, \varepsilon_t' \sum_{v=1}^{t-1} c_{t-v}^{l}\varepsilon_v\right)' + o_p(1),$$

where for $k \in \{l_1+1,\ldots,l\}$

$$c_t^k = \frac{1}{\pi n}\varepsilon_t' \sum_{j=1}^{m} \rho^k(\lambda_j)\cos(t\lambda_j)$$

and

$$\rho^k(\lambda) = (1-e^{-i\lambda})^{-\delta_k}\Omega^{1/2}D'(e^{i\lambda})p'(\lambda)i_k' D(e^{-i\lambda})\Omega^{1/2}$$
$$+ (1-e^{i\lambda})^{-\delta_k}\Omega^{1/2}D'(e^{i\lambda})i_k p(-\lambda)D(e^{-i\lambda})\Omega^{1/2}.$$

Then, (A.1) holds as in Hualde and Robinson (2005) by analyzing the joint convergence of the vector $c(r) = (c_1'(r), c_2(r), c_3'(r))'$, where

$$c_1(r) = \left(\frac{1}{n^{\delta_1-1/2}}x_{1,[nr]}, \ldots, \frac{1}{n^{\delta_{l_1}-1/2}}x_{l_1,[nr]}\right)',$$

$$c_2(r) = \frac{1}{n^{1/2}}\sum_{t=1}^{[nr]} p(0)D(1)\varepsilon_t,$$

$$c_3(r) = \left(\frac{\lambda_m^{\delta_{l_1+1}}}{m^{1/2}}\sum_{t=2}^{[nr]}\varepsilon_t'\sum_{v=1}^{t-1}c_{t-v}^{l_1+1}\varepsilon_v, \ldots, \frac{\lambda_m^{\delta_l}}{m^{1/2}}\sum_{t=2}^{[nr]}\varepsilon_t'\sum_{v=1}^{t-1}c_{t-v}^{l}\varepsilon_v\right)'.$$

Thus, (A.1) follows by Marinucci and Robinson (2000), Brown (1971), and Kurtz and Protter (1991), the independence between the components being due to the result that the processes $c_1(r)$, $c_2(r)$, $c_3(r)$ are uncorrelated. ■

Proof of Theorem 2. The proof follows on showing that

$$n^{-\delta}\tilde{\tau}_n(\gamma,\theta) \Rightarrow \int_0^1 \hat{W}(r;\delta)\, d\overline{W}(r), \tag{A.3}$$

where

$$\hat{W}(r;\delta) = (0_l, I_l)D(1)\,\frac{1}{\Gamma(\delta)}\int_0^r (r-s)^{\delta-1}\, dW(s),$$

$$\overline{W}(r) = 2\pi\zeta'(D(1)^{-1})'\Omega^{-1}W(r)$$

and

$$\tilde{\tau}_n(\hat{\gamma},\hat{\theta}) - \tilde{\tau}_n(\gamma,\theta) = o_p(n^{\delta}). \tag{A.4}$$

First, noting that under H_0

$$\tilde{\tau}_n(\gamma,\theta) = \sum_{j=1}^{n} w_x(-\lambda_j)p(\lambda_j;\theta)w_u(\lambda_j),$$

(A.3) follows from a trivial multivariate extension of Propositions 1–3 of RH, the only significant difference now being that the vector $w_x(-\lambda_j)$ replaces the discrete Fourier transform of a filtered scalar process x_t in RH, but this could be straightforwardly accounted for.

Regarding (A.4), we only show

$$\tilde{\tau}_n(\hat{\gamma},\theta) - \tilde{\tau}_n(\gamma,\theta) = o_p(n^{\delta}), \tag{A.5}$$

the rest of the proof following directly by Propositions 7 and 10 of RH. The result in (A.5) corresponds to Proposition 9 in RH, but our present situation is more delicate because we just require that $\hat{\gamma}$ satisfy (21) with $\kappa > 0$. Thus, our proof strategy is substantially different, and it is worth giving a detailed analysis.

The transpose of the left side of (A.5) is

$$\frac{1}{2\pi n}\sum_{j=1}^{n} p(\lambda_j)\sum_{t=2}^{n}\sum_{m=1}^{t-1} a_m(\delta-\hat{\gamma})u_{t-m}e^{it\lambda_j}\sum_{s=1}^{n} x_s' e^{-is\lambda_j}$$

$$= \sum_{r=1}^{R-1}\frac{(\delta-\hat{\gamma})^r}{2\pi n}\sum_{j=1}^{n} p(\lambda_j)\sum_{t=2}^{n}\sum_{m=1}^{t-1} a_m^{(r)}(0)u_{t-m}e^{it\lambda_j}\sum_{s=1}^{n} x_s' e^{-is\lambda_j} \tag{A.6}$$

$$+ \frac{(\delta-\hat{\gamma})^R}{2\pi n}\sum_{j=1}^{n} p(\lambda_j)\sum_{t=2}^{n}\sum_{m=1}^{t-1} a_m^{(r)}(\delta-\overline{\gamma})u_{t-m}e^{it\lambda_j}\sum_{s=1}^{n} x_s' e^{-is\lambda_j}, \tag{A.7}$$

where $p(\lambda) = p(\lambda;\theta)$, $a_s^{(r)}(c) = d^r a_s(c)/dc^r$, and $|\overline{\gamma}-\delta| \le |\hat{\gamma}-\delta|$. First, as in RH, the second term in (A.7) can be shown to be of smaller order for R large enough. Next, we show that the rth term in (A.6) is $O_p(n^{-\kappa r}n^{\delta+\epsilon})$ for any $\epsilon > 0$. First

$$E\left(\frac{1}{2\pi n}\sum_{j=1}^{n}p(\lambda_j)\sum_{t=2}^{n}\sum_{m=1}^{t-1}a_m^{(r)}(0)u_{t-m}e^{it\lambda_j}\sum_{s=1}^{n}x_s'e^{-is\lambda_j}\right)$$

$$= \frac{1}{2\pi n}\sum_{j=1}^{n}p(\lambda_j)\int_{-\pi}^{\pi}\sum_{s=1}^{n-1}a_s^{(r)}(0)e^{is\lambda_j}D_{n-s}(\lambda_j-\mu)$$

$$\times\sum_{t=0}^{n-1}a_t(\delta)e^{-it\lambda_j}D_{n-t}(\mu-\lambda_j)f(\mu)\xi d\mu, \quad \textbf{(A.8)}$$

where $D_t(\lambda) = \sum_{k=1}^{t}e^{ik\lambda}$ is the Dirichlet kernel. Noting that for any λ, $p(\lambda)f(\lambda)\xi$ is identically zero, by periodicity, (A.8) can be written as

$$\frac{1}{2\pi n}\sum_{j=1}^{n}p(\lambda_j)\int_{-\pi}^{\pi}\sum_{s=1}^{n-1}a_s^{(r)}(0)e^{is\lambda_j}D_{n-s}(-\mu)$$

$$\times\sum_{t=0}^{n-1}a_t(\delta)e^{-it\lambda_j}D_{n-t}(\mu)[f(\mu+\lambda_j)-f(\lambda_j)]\xi d\mu,$$

which, by summation by parts, is

$$\frac{1}{2\pi n}\sum_{j=1}^{n}p(\lambda_j)\int_{-\pi}^{\pi}\sum_{s=1}^{n-1}a_s^{(r)}(0)e^{is\lambda_j}D_{n-s}(-\mu)$$

$$\times\left\{a_{n-1}D_1(\mu)[f(\mu+\lambda_j)-f(\lambda_j)]\xi\sum_{t=0}^{n-1}e^{-it\lambda_j}\,d\mu-[f(\mu+\lambda_j)-f(\lambda_j)]\xi\right.$$

$$\left.\times\sum_{t=0}^{n-2}(a_{t+1}D_{n-t-1}(\mu)-a_tD_{n-t}(\mu))\sum_{h=0}^{t}e^{-ih\lambda_j}\,d\mu\right\}, \quad \textbf{(A.9)}$$

where $a_t = a_t(\delta)$. Because $\sum_{t=0}^{n-1}e^{-it\lambda_j} = n$, $j = 0$, mod n; $= 0$, otherwise, and f is boundedly differentiable, the contribution of the first term in braces in (A.9) is bounded in norm by

$$K|a_{n-1}|\sum_{s=1}^{n-1}|a_s^{(r)}(0)|\int_{-\pi}^{\pi}|\mu||D_{n-s}(-\mu)|\,d\mu, \quad \textbf{(A.10)}$$

where throughout K denotes a generic finite positive constant. Noting that for $0 < \lambda < \pi$,

$$|D_t(\lambda)| < K\min\{|\lambda|^{-1}, t\} \quad \textbf{(A.11)}$$

(see Zygmund, 1977), it can be easily shown that, uniformly in s,

$$\int_{-\pi}^{\pi}|\mu||D_{n-s}(-\mu)|\,d\mu = O(1),$$

so that by Lemma D.4 of RH, (A.10) is bounded by

$$Kn^{\delta-1}\sum_{s=1}^{n-1}\frac{\log^{r-1}s}{s} = O(n^{\delta-1}\log^r s).$$

Regarding the second term in (A.9), noting that

$$a_{t+1}D_{n-t-1}(\mu) - a_t D_{n-t}(\mu) = (a_{t+1} - a_t)D_{n-t-1}(\mu) - e^{i(n-t)\mu}a_t, \quad \textbf{(A.12)}$$

the contribution of the first term on the right of (A.12) to the second term of (A.9) is 0 for $\delta = 1$, as in this case $a_{t+1} = a_t$, $t = 0, \ldots, n-2$. For $\delta \neq 1$, this contribution is bounded in modulus by

$$Kn^{-1}\sum_{j=1}^{n}\left\{\int_{-\pi}^{\pi}\left|\sum_{s=1}^{n-1} a_s^{(r)}(0)e^{is\lambda_j}D_{n-s}(-\mu)\right|^2 \|f(\mu+\lambda_j) - f(\lambda_j)\|\,d\mu\right\}^{1/2}$$
$$\times\left\{\int_{-\pi}^{\pi}\left|\sum_{t=0}^{n-2}(a_{t+1}-a_t)D_{n-t-1}(\mu)(D_t(-\lambda_j)+1)\right|^2\|f(\mu+\lambda_j)-f(\lambda_j)\|\,d\mu\right\}^{1/2}. \quad \textbf{(A.13)}$$

The term in the first set of braces is bounded uniformly in j by

$$K\int_{-\pi}^{\pi}|\mu|\sum_{s=1}^{n-1}\sum_{t=1}^{n-1}a_s^{(r)}(0)a_t^{(r)}(0)e^{i(s-t)\lambda_j}D_{n-s}(-\mu)D_{n-t}(\mu)\,d\mu$$
$$\leq K\int_{-\pi}^{\pi}|\mu|\sum_{s=1}^{n-1}(a_s^{(r)}(0))^2|D_{n-s}(-\mu)|^2\,d\mu \leq K\log n,$$

because by Zygmund (1977)

$$\int_{-\pi}^{\pi}|D_n(\mu)|\,d\mu = O(\log n).$$

Next, the term in the second set of braces is bounded by

$$K\int_{-\pi}^{\pi}|\mu|\sum_{t=0}^{n-2}\sum_{s=0}^{n-2}(a_{t+1}-a_t)D_{n-t-1}(\mu)(D_t(-\lambda_j)+1)$$
$$\times (a_{s+1}-a_s)D_{n-s-1}(-\mu)(D_s(\lambda_j)+1)\,d\mu$$
$$= O\left(j^{-2}n^2\log n\left(\sum_{t=1}^{n}t^{\delta-2}\right)^2\right),$$

by Lemma C.1 of RH and (A.11), which is $O(j^{-2}n^2\log n 1(\delta < 1) + j^{-2}n^{2\delta}\log n 1(\delta > 1))$, implying that (A.13) is $O(\log^2 n 1(\delta < 1) + n^{\delta-1}\log^2 n 1(\delta > 1))$. Finally, the contribution of the second term on the right of (A.12) to the second term of (A.9) is bounded in modulus by

$$Kn^{-1}\sum_{j=1}^{n}\left\{\int_{-\pi}^{\pi}|\mu|^2\left|\sum_{s=1}^{n-1}a_s^{(r)}(0)e^{is\lambda_j}D_{n-s}(-\mu)\right|^2 d\mu\right.$$
$$\left.\times\int_{-\pi}^{\pi}\left|\sum_{t=0}^{n-2}e^{i(n-t)\mu}a_t(D_t(-\lambda_j)+1)\right|^2 d\mu\right\}^{1/2}. \quad \textbf{(A.14)}$$

The first integral inside the braces is $O(\log^{2r} n)$ by (A.11), whereas noting that

$$\int_{-\pi}^{\pi} e^{-i(s-t)\mu}\, d\mu = 2\pi, \qquad s = t; \qquad = 0, \quad \text{otherwise},$$

the second is bounded by $K\sum_{t=1}^{n} a_t^2 |D_t(\lambda_j)|^2$, so that (A.14) is bounded by $Kn^{-1}\log^r n \sum_{j=1}^{n}\{n^{2\delta+1} j^{-2}\}^{1/2}$, which is $O(n^{\delta-1/2}\log^{r+1} n)$, implying that the left of (A.8) is $O(n^{1/2}\log n 1(\delta < 1) + n^{\delta-1/2}\log^{r+1} n 1(\delta \geq 1))$.

Next, by straightforward calculations and application of Lemma C.2 of RH

$$\operatorname{Var}\left(\frac{1}{2\pi n}\sum_{j=1}^{n} p(\lambda_j)\sum_{t=2}^{n}\sum_{j=1}^{t-1} a_j^{(r)}(0) u_{t-j} e^{it\lambda_j}\sum_{s=1}^{n} x_s' e^{-is\lambda_j}\right) = O(n^{2\delta+\epsilon}),$$

for any $\epsilon > 0$, which implies that the rth term of (A.6) is $O_p(n^{-\kappa r+\delta+\epsilon})$, for any $\epsilon > 0$, so that (A.5) holds for any $\kappa > 0$ on choosing $\epsilon < \kappa$, to conclude the proof. ■