Working Paper nº 08/06

Distribution-free Tests of Fractional Cointegration

Javier Hualde
Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Carlos Velasco
Departamento de Economía
Universidad Carlos III de Madrid
ABSTRACT

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 GLS-type of corrections which 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 which 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 analysed by means of a Monte Carlo experiment.

Javier Hualde       Carlos Velasco
Facultad de CC. Económicas y Empresariales  Departamento de Economía
Universidad de Navarra       Universidad Carlos III de Madrid
jhualde@unav.es       carlos.velasco@uc3m.es
Distribution-free tests of fractional cointegration *

Javier Hualde†
Universidad de Navarra
Tel: +34948425625
Fax: +34948425626

Carlos Velasco‡
Universidad Carlos III de Madrid
Tel: +34916249646
Fax: +34916249875

May 31, 2006

Abstract

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 GLS-type of corrections which control for the short run correlation of the weak dependent components of the fractionally integrated processes. We emphasize corrections based on nonparametric modeling of the innovations’ autocorrelation, relaxing important conditions which 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 analysed by means of a Monte Carlo experiment.

KEYWORDS: Fractional cointegration, Mixed normal asymptotics, Prewhitening, Semiparametric model.

JEL Classification: C12, C13, C22

†Research supported by the Spanish Ministerio de Educación y Ciencia through a contract Juan de la Cierva and Ref. no. SEJ2005-07657/ECON.

‡Research supported by the Spanish Ministerio de Educación y Ciencia, Ref. no. SEJ2004-04583/ECON.

*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, and the Unit Root and Cointegration Testing Conference at the University of Algarve, Faro, Portugal, and seminar participants at the Universidad de Navarra for helpful comments.
Distribution-free tests of fractional cointegration

Javier Hualde
Department of Quantitative Methods
School of Economics and Business Administration
Universidad de Navarra
Edificio Bibliotecas (Entrada Este)
31080 Pamplona (Navarra), Spain

email: jhualde@unav.es
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 analysing 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, 2004, Marmol and Velasco, 2004). All these problems have been tackled satisfactorily in the standard cointegrating set-up 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, while 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 in the precise specification of this idea. For example Robinson and Yajima (2002), RY hereinafter, presents several definitions already proposed in the literature (Johansen, 1996, Flóres and Szafarz, 1996, Robinson and Marinucci, 2003), and offer a new one. While 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 above, 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 of 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, while their cointegration testing procedure is able to determine the cointegrating rank (see also Chen and Hurvich, 2003, for a related approach). Breitung and Hassler (2002) proposed a test which extends Johansen (1988), and allows to determine the cointegrating rank of a vector of fractionally integrated processes. Hassler and Breitung (2004) test for the null of no cointegration by applying a modified 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 Marmol and Velasco (2004), MV hereinafter, which checks for the absence of cointegration comparing the ordinary least squares (OLS) and a generalized least squares (GLS) type of 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 hypothesis, 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 (2001, 2004, 2005), Robinson and Hualde (2003), RH hereinafter, which 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, deriving a null distribution free test, hence avoiding the null non standard limit distribution of the MV 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, 2004). 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 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, and 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 limited, we believe that given the previously mentioned improvements and extensions in some directions over existing works in the literature, our test procedure 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 which 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 behaviour 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

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

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

$$\delta_0 \geq \max_{i \in [1,l]} \delta_i > 0, \quad \min_{i \in [1,l]} \delta_i \geq 0;$$

the fractional difference operator $\Delta^{-\alpha}$ is defined in terms of $\Delta = 1 - L$, where $L$ is the lag operator, and the formal series, for any real $\alpha \neq -1, -2, \ldots,$

$$(1 - z)^{-\alpha} = \sum_{j=0}^{\infty} a_j (\alpha) z^j, \quad a_j (\alpha) = \frac{\Gamma(j + \alpha)}{\Gamma^2(\alpha) \Gamma(j + 1)};$$

where $\Gamma$ is the gamma function and $\Gamma(0)/\Gamma(0) = 1$; $u_t$ 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 $\delta_i > 0.5$, because the weights $a_j(\delta_i)$ are not square summable in this case, but 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 $i$-th 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 while cointegration involving this type of processes is possible in the strict sense, it does not have empirical appeal. For similar reasons we set $\max_i \delta_i > 0$, 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 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 Flöres 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 $\alpha^t z_t \sim I(\gamma)$ with $\gamma < \delta_0$, and at least a non-zero scalar component of $\alpha$ multiplies one component of $z_t$ with integration order equal to $\delta_0$.

Obviously this definition implies that for $z_t$ being cointegrated is needed that at least one observable apart from the first one have integration order $\delta_0$. Note that this definition could miss some cointegrating relations (see e.g. RY for a good example) 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, so they could be taken out from the model and interpret the new vector of observables without these variables in terms of Definition 1. Note that the test statistics which 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 below for a definition), a necessary and sufficient condition for the existence of cointegration among the elements of $z_t$ is that $f(0)$ is singular. If, on the contrary, $f(0)$ is full rank, we say that $z_t$ is spuriously related, because any non-trivial linear combination with a non-zero 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 in order 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 not cointegration in the model. These ideas are nicely captured by the following structure.

First, we partition the basic vectors as $z_t = (y_t, x_t)'$, $u_t = (u_{yt}, u_{xt})'$, with $x_t, u_{xt}$ being $l \times 1$, 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 $\zeta_t$, $$\zeta_t(c) = \Delta^c \{\zeta_t 1(t > 0)\},$$ suppose there exists a weakly dependent covariance stationary scalar process $v_{yt}$, a real number $\gamma$ such that $0 \leq \gamma \leq \delta_0$, and a $l \times 1$ 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 (2001) and RH, which for this case is

$$y_t = \beta' x_t + v_{yt} (-\gamma), \tag{4}$$
$$x_t = \Delta_t^{-1} (\overline{\delta}) \{u_{xt} 1(t > 0)\}, \tag{5}$$

where $\Delta_t (\overline{\delta}) = \text{diag} \{\Delta^{\delta_1}, \ldots, \Delta^{\delta_l}\}$. 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, so

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

Note that assuming that $u_{yt}$ is a covariance stationary $I(0)$ process and $\gamma < \delta_0$, under (3), $u_{yt}$ is only asymptotically stationary due to 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 $\delta_0$. However, based on the following definition we could easily obtain the covariance stationary approximation of $u_{yt}$. 

\[ \text{(Continued on page 7)} \]
Definition 2 Given $\alpha < 1/2$, let $\xi_t$ be a covariance stationary $I(0)$ process, and

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

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

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

(7)

Note that setting the difference equation

$$\Delta^\alpha \zeta_t = \xi_t, $$

both $\zeta_t, \tilde{\zeta}_t,$ are solutions of (8) (given certain initial conditions in the case of $\zeta_t$), but while $\zeta_t$ 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 $\alpha < 1/2,$ and the process (7) is well-defined. For $\alpha \geq 1/2,$ (8) does not have a stationary solution. The covariance stationary approximation of $u_{yt}$ is given by

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

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

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

and also let $\beta = (\beta_1', \beta_2')'$, where $\beta_1$ and $\beta_2$ are $\tilde{t}_1 \times 1$ and $\tilde{t}_2 \times 1$ vectors corresponding to components of $x_t$ with integration orders equal to $\delta_0$ or smaller than $\delta_0$ respectively, with $\tilde{t}_1 + \tilde{t}_2 = l$. 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 $x_t,$ $\beta$ is the fundamental vector (cf. Park, Ouliaris and Choi, 1988)

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

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

$$f_{xx}^{-1}(0) f_{xy}(0) = \left( \beta_1', 0_{\tilde{t}_2}' \right)' ,$$

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

3 The test procedure in the white noise case

We find 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 $\delta$. This condition is certainly restrictive, but it is also introduced by Breitung and Hassler (2002), Hassler and Breitung (2004), 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 \( w_t = (y_{yt}, u_{xt})' \) is independent and identically distributed (IID) with zero mean and nonsingular covariance matrix

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

where we assume that \( \Omega_{xx} \) is also positive definite. Furthermore, the IID condition of \( w_t \) will be taken as known, 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 in order to grasp the intuition behind our test methodology.

Using (3) we find that \((\tilde{u}_{yt}, u_{xt}^2)' \) has spectral density matrix

\[
f (\lambda) = \frac{1}{2\pi} \left( \begin{array}{c}
\omega_{yy} |h (\lambda)|^2 + 2\beta' \omega_{xy} \text{Re} \{ h (\lambda) \} + \beta' \Omega_{xx} \beta \\
\omega_{xy} h (-\lambda) + \Omega_{xx} \beta
\end{array} \right), \quad (9)
\]

where \( h (\lambda) = (1 - e^{i\lambda})^{-\gamma} \). Then, when \( \gamma = \delta, (u_{yt}, u_{xt})' = (\tilde{u}_{yt}, u_{xt}^2)' \) is a white noise sequence with nonsingular constant spectral density matrix, \( f (\lambda) = f (0) \), which does not depend on \( \gamma \) or \( \delta \) 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 \( \delta \) and \( \gamma \) can be useful to derive hypothesis tests of the null of no cointegration, and although the values of the nuisance parameters \( \delta \) and \( \beta \) are in general unknown, these could be estimated from data, and from these estimates, in turn, we may be able to estimate \( \gamma \) 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 non standard asymptotic properties of usual statistics, such as OLS coefficients. Note that the estimation of \( \gamma \) from residuals is inherent to our approach, so our test, although nonstandard, could be regarded as a “residual based regression test”.

We define the projection vector

\[
\eta = f_{xx}^{-1} (0) f_{xy} (0),
\]

noting that

\[
\eta = \Omega_{xx}^{-1} \omega_{xy} + \beta, \quad \text{under } H_0,
\]

\[
= \beta, \quad \text{under } H_1.
\]

Letting \( c, h \) be any possible value or estimate of the parameters \( \gamma, \eta \), 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

\[
\omega_{y,x} = 2\pi f_{yy} (0) (1 - \rho_{y,x}^2), \quad (11)
\]

where \( \rho_{y,x}^2 \) is the squared coefficient of multiple correlation between \( y_{yt} \) and \( u_{xt} \) given by

\[
\rho_{y,x}^2 = \frac{f_{yx} (0) f_{xx}^{-1} (0) f_{xy} (0)}{f_{yy} (0)} . \quad (12)
\]

More importantly, under \( H_0 \),

\[
E (x_t v_t (\gamma, \eta)) = 0, \quad (13)
\]
for all \( t \), whereas, noting that for any \( l \times 1 \) vector \( \zeta \),

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

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 (13) using consistent estimates of \( \gamma \) and \( \eta \) under the null. Under the alternative, however, inconsistent estimation of \( \eta \) guarantee that (13) fails and the residuals \( v_t \) are no longer \( I (0) \).

Thus, setting

\[
\tau_n (c, h) = \sum_{t=1}^{n} x_{t-1} v_t (c, h),
\]

(we explain later in this section why \( x_{t-1} \) replaces the “more natural” \( x_t \) in (14)), it can be shown that under additional regularity conditions (to be detailed in the next section),

\[
n^{-1/2} \tau_n (\gamma, \eta) \rightarrow_d N (0, \omega_{y,x} E (\tilde{\tilde{x}}_t z_t')),
\]

if \( \delta < 1/2 \), where

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

By contrast, denoting by “\( \Rightarrow \)” convergence in the Skorohod topology on the appropriate metric space,

\[
n^{-\delta} \tau_n (\gamma, \eta) \Rightarrow \int_0^1 W_x (r; \delta) dW_{y,x} (r),
\]

when \( \delta > 1/2 \). 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) \( W (r) = (W_y (r), W^*_x (r))^l \) 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 (16) is a mixed normal distribution, so that in view of this result and also (15), it is expected that an appropriately normalized statistic based on \( \tau_n (\gamma, \eta) \) has a \( \chi^2 \) limiting distribution irrespective of whether \( \delta < 1/2 \) or \( \delta > 1/2 \). In fact, defining for \( b \neq 0 \) the statistic

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

it is straightforward to show that

\[
\Xi_n (\omega_{y,x}, \gamma, \eta) \rightarrow_d \chi^2 \text{ under } H_0.
\]

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 \( \eta \) 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 (2001) 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 $\Xi_n$ compared with a $\chi^2_1$ distribution, once that consistent estimates of the unknown $\gamma$, $\eta$ and $\omega_{y,x}$ are found. As it will be seen in the next section, under correlated innovations we should replace the basic OLS-type fluctuations $\tau_n$ 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 since, in general, both the elements of $f(0)$ and $\gamma$ are unknown. However, given an estimate $\hat{\delta}$ of $\delta$, we can easily estimate the elements of $f(0)$ by

$$\bar{f}_{yy}(0) = \frac{1}{2\pi n} \sum_{t=1}^{n} y_y^2(\hat{\delta}), \quad \bar{f}_{xy}(0) = \frac{1}{2\pi n} \sum_{t=1}^{n} x_t(\hat{\delta})y_t(\hat{\delta}), \quad \bar{f}_{xx}(0) = \frac{1}{2\pi n} \sum_{t=1}^{n} x_t(\hat{\delta})x_t^*(\hat{\delta}),$$ (18)

and then from (10), (11), (12), obtain easily corresponding estimates of $\eta$ and $\omega_{y,x}$ respectively. $\hat{\delta}$ could be recovered from levels $y_t$ or either $x_t$, or from (asymptotically) stationary increments $\Delta y_t$ or $\Delta x_t$, with a rate of convergence,

$$\hat{\delta} = \delta + O_p\left(n^{-\kappa}\right), \quad \kappa > 0.$$ (19)

Most analysis 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 (19), 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 $\eta$ based on (18) are $\sqrt{n}$-consistent due to their parametric nature.

To obtain consistent estimates of $\gamma$ we can use the OLS or NB residuals

$$\hat{v}_t = y_t - x_t^*\hat{\beta},$$

to get

$$\hat{\gamma} = \gamma + O_p\left(n^{-\kappa}\right), \quad \kappa > 0,$$ (20)

under both hypotheses. If $H_0$ is true, so $\gamma = \delta$, since $\hat{\beta}$ is inconsistent for $\beta$, $\hat{v}_t$ is a linear combination (with stochastic coefficients) of $I(\delta)$ processes in a non-cointegrating direction, so that $\hat{\gamma}$ is expected to be a consistent estimate of $\delta$. 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 $\delta < 1/2$, but the NB suffices), 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 (19) is sufficient to show that under $H_0$

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

10
so that our feasible test statistics share the same (first order) asymptotic properties as the infeasible ones,

$$\Xi_n(\hat{\omega}_{y,x}, \hat{\gamma}, \hat{\eta}) \sim_d \chi^2_d$$ under $H_0$,

for both $\delta < 1/2$ and $\delta > 1/2$. When $\delta < 1/2$, the effect of the estimation of $\eta$ is negligible thanks to the utilization of the regressor $x_{t-1}$ in $\tau_n$ instead of other alternatives, such as $x_t$. However in a general parametric framework a similar strategy seems to be not possible, see Remark 4 in Section 5 below. On the contrary, when $\delta > 1/2$ it is indistinctive to use $x_t$ or $x_{t-1}$ in (14).

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 Robinson and Marinucci’s (2001) results, we can obtain the following sharp rates for $\tau_n(\gamma, \zeta)$ under $H_1$:

$$n^{-\max\{1/2, \delta\}} \tau_n(\gamma, \zeta) = O_p\left(n^{1/2}\right), \quad \delta < 1/2,$$

$$= O_p\left(n^{1-\delta}\right), \quad \delta > 1/2, \quad 2\delta - \gamma < 1,$$

$$= O_p\left(n^{1-\delta} \log n\right), \quad \delta > 1/2, \quad 2\delta - \gamma = 1,$$

$$= O_p\left(n^{\delta-\gamma}\right), \quad \delta > 1/2, \quad 2\delta - \gamma > 1,$$

all diverging with $n$, in contrast with (15), (16) under the null. Thus, the key is to employ an estimate of $\eta$ 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 $\eta$ under $H_0$ which diverge under the alternative. For our simple model, this could be

$$\tilde{\eta} = \left(\sum_{t=1}^n x_t(\hat{\delta})x_t'(\hat{\delta})\right)^{-1/2} \sum_{t=1}^n x_t(\hat{\gamma})v_t(\hat{\gamma}),$$

for which the following sharp rates can be derived:

$$\tilde{\eta} = O_p\left(1\right), \quad \delta - \gamma < 1/2,$$

$$= O_p\left(\log n\right), \quad \delta - \gamma = 1/2,$$

$$= O_p\left(n^{2(\delta-\gamma)-1}\right), \quad \delta - \gamma > 1/2,$$

so that, in case $\delta - \gamma \geq 1/2$, the divergence rate of the feasible $\tau_n$ 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 $\tilde{\eta}$ or $\hat{\eta}$, as both are $\sqrt{n}$-consistent estimates of $\eta$.

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 $H_1$, $\tilde{\rho}_{y,x}^2 - 1 = O_p\left(n^{-1/2}\right)$, 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

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

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 IID 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 which 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 $\tau_n$ by those of alternative statistics that preserve a similar orthogonality property to that achieved by $v_t (\gamma, \eta)$ with $x_t$, cf. (13), 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 which 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, since it is widely assumed that nonstationary processes have a very important role in economics, but, undoubtedly, 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 below for an explanation of this point).

Thus, a semiparametric approach which, while 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 which are valid for any nonnegative value of $\delta$, $i = 0, ..., l$, excluding 1/2, so basically no a priori knowledge of the type discussed before is needed in order to apply the following techniques.

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

**Assumption 1** For \( 0 \leq \gamma \leq \delta_0 \), there exists a \( 1 \times 1 \) vector \( \beta \neq 0 \) such that (3) holds and the process \( w_t = (v_{yt}, u_{yt})', \ t = 0, \pm 1, \ldots \) 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

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

(ii) \( A(e^{i\lambda}) \) is differentiable in \( \lambda \) with derivative in Lip(\( \varepsilon \)), \( \varepsilon > 1/2; \)

(iii) The \( \varepsilon_t \) are IID vectors with mean zero, positive definite covariance matrix \( \Omega \), 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 easily for stationary and invertible ARMA processes. Under \( H_0 \),

\[
  u_t = C(L)w_t, \quad C(L) = \begin{pmatrix}
    1 & \beta' \Delta \delta \Delta_{t-1} (\beta) \\
    0_t & I_t
  \end{pmatrix},
\]

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 \( i \in [1, l] \), this dependence disappears and \( C(L) \) reduces to

\[
  C(L) = C = \begin{pmatrix}
    1 & \beta' \\
    0_t & I_t
  \end{pmatrix}.
\]

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 in order 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 short 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 which 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

\[
  w_a(\lambda) = \frac{1}{(2\pi n)^{1/2}} \sum_{t=1}^{n} a_t \exp(i\lambda t), \quad I_{ab}(\lambda) = w_a(\lambda) w_b^*(-\lambda), \quad I_a(\lambda) = I_{aa}(\lambda),
\]

and

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

13
where \( \zeta = (1, 0')' \), given any real function \( g(\lambda) \), \( \lambda \in [-\pi, \pi] \), we consider statistics based on

\[
\hat{\tau}_m (c, \tilde{d}, g) = \sum_{j=0}^{m} s_j \Re w_x (-\lambda_j) g(\lambda_j) w_{z(c, \tilde{d})} (\lambda_j), \quad \hat{\tau}_{m0} (c, \tilde{d}, g) = \sum_{j=0}^{m} s_j \Re w_x (-\lambda_j) g(0) w_{z(c, \tilde{d})} (\lambda_j),
\]

(22)

where \( \lambda_j = 2\pi j/n \) are the Fourier frequencies, \( \tilde{d} = (d_1, ..., d_t)' \), \( z_t (c, \tilde{d}) = (y_t (c), x_{1t} (d_1), ..., x_{kt} (d_k)) \), \( m \) 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

\[
g(\lambda) = \hat{\rho}(\lambda) = \zeta' \hat{f}(\lambda)^{-1},
\]

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

Denoting by \([a]\) the integer part of \( a \), note that due to the symmetry properties of the Fourier transform, we have that, for example

\[
\hat{\tau}_{[m/2]} (c, \tilde{d}, \tilde{\rho}) = \sum_{j=1}^{m} w_x (-\lambda_j) \hat{\rho}(\lambda_j) w_{z(c, \tilde{d})} (\lambda_j),
\]

which, due to the presence of all the Fourier frequencies, could be referred as a full band statistic. When \( m/n \to 0 \) as \( n \to \infty \), \( \hat{\tau}_m (c, \tilde{d}, g) \) only considers frequencies on a shrinking neighbourhood around zero, being for this reason usually referred as a NB statistic. Some of the results below will also apply to cases where \( m/n \to K < 1/2 \) as \( n \to \infty \), but these do not have much intuitive appeal and are hardly stressed in the literature. \( \hat{\tau}_m^0 (c, \tilde{d}, \tilde{\rho}) \) is related to what Hualde and Robinson (2005) denoted as “zero-frequency” statistic, because the weighting factor \( \tilde{\rho} \) 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 below, 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 propose now our semiparametric test statistics. Defining, \( \hat{q}(\lambda) = \zeta' \hat{f}(\lambda)^{-1} \zeta \),

\[
\hat{b}_m = \sum_{j=0}^{m} s_j \Re \hat{q}(\lambda_j) I_x (\lambda_j), \quad \hat{b}_{m0} = \hat{q}(0) \sum_{j=0}^{m} s_j I_x (\lambda_j),
\]

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

\[
\hat{\Upsilon}_m (c, \tilde{d}, g) = \hat{\tau}_m^0 (c, \tilde{d}, g) \hat{b}_m^{-1} \hat{\tau}_m (c, \tilde{d}, g), \quad \hat{\Upsilon}_m^0 (c, \tilde{d}, g) = \hat{\tau}_m^0 (c, \tilde{d}, g) \left( \hat{b}_m^0 \right)^{-1} \hat{\tau}_m^0 (c, \tilde{d}, g),
\]

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

Alternatively, considering the known function \( f(\lambda; h) \), \( h \in \mathbb{R}^k \), where for a \( k \times 1 \) vector \( \theta \) of unknown parameters \( f(\lambda, \theta) = f(\lambda) \), it is straightforward to design a parametric version of our test statistics. In order to develop this extension, it is important to take into account important aspects which 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 \( H_0 \). While 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 \( \delta_0 \), which makes 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 $\delta_0$, $u_{yt}$ 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 which exploits all the information contained in $f$ and could be the basis of our test procedure is

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

where $p (\lambda; h) = \zeta^j f^{-1} (\lambda; h)$. The feasibility of the test depends now on estimates of $\gamma$ and $\theta$ 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 $\gamma$, calculating the residual vector $z_t (\hat{\gamma})$ it is possible to identify the parametric model driving $u_t$, and, on building on the results of RH, it is simple to show that parametric estimates of $\theta$ 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 is fully developed in the multivariate framework (see e.g. Dunsmuir and Hannan, 1976, Dunsmuir, 1979). Here, methods which 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 Dickey and Startz (1998) and Hassler and Breitung (2004).

5  Asymptotic properties of cointegration tests

In order 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

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

$$
|\hat{\gamma}| + \left\| \hat{\delta} \right\| \leq K,
$$

and $\kappa > 0$ such that

$$
\hat{\gamma} = \gamma + O_p (n^{-\kappa}), \quad \hat{\delta} = \delta + O_p (n^{-\kappa}),
$$

where, as $n \to \infty$

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

There are several important remarks related to this assumption. First, (24) is not restrictive if our semiparametric estimates are optimizers of corresponding loss functions over compact sets. Next, the likability of (25), (26) 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 \( \gamma \) should be based, are a linear combination (with stochastic coefficients) of fractionally integrated processes with dominant order \( \delta_0 \). The presence of these stochastic coefficients complicates matters substantially, and while a very detailed analysis goes beyond the scope of the present paper, we offer a brief justification of why (25) holds for a particular estimate of \( \gamma \), the Gaussian semiparametric, proposed by Künsch (1987) and analysed by Robinson (1995a). First, suppose that all observables share the same integration order \( \delta_0 \), and for simplicity let \( 1/2 < \delta_0 < 3/2 \). Denote by \( \hat{b} \) the OLS or NB estimate of \( \beta \). Under \( H_0 \), it can be shown that \( \hat{b} \) converges weakly, so \( \hat{b} = O_p(1) \) and \( \|\hat{b}\| > 0 \) with probability tending to one. Define \( \tilde{b} = (1, \hat{\beta})^t \| (1, \hat{\beta}) \| \), so that \( \|\tilde{b}\| = 1 \), and \( \tilde{c}_t(\tilde{b}) = \tilde{b}' z_t \). Clearly, under Assumption 1, the spectral density of \( \Delta z_t \) behaves like

\[
\hat{f}_{\Delta z}(\lambda) \sim G_{\Delta z} \lambda^{-2(\delta_0-1)} \quad \text{as} \quad \lambda \to 0,
\]

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 (2004), 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

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

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 \( \Delta z_t \) around frequency 0. Thus, in the most favourable 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 \( \delta_0 \) typically are not bounded in probability, invalidating in principle the estimation of \( \gamma \) based on the residuals \( y_t - \hat{\beta}' x_t \). For the sake of an easy exposition, collect those components of \( \hat{\beta} \) in a \( l^* \times 1 \) vector \( \hat{\beta}^* \), where \( l^* \in [1, l - 1] \), and correspondingly let \( x_t^* \) be the vector of components of \( x_t \) with integration order strictly smaller than \( \delta_0 \). To describe the procedure we propose in this situation let also \( \hat{\beta} \) be a NB estimate. Then, it can be easily shown that if \( \delta_0 < 1/2 \) (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 \( \gamma \) grow slower than that used in the NB estimation of \( \beta \), the contribution of \( \hat{\beta}^* x_t^* \) to the spectral density of the residuals \( y_t - \hat{\beta}' x_t \) can be made negligible, so the terms with \( \delta_0 \) memory dominate. Note that this strategy is only valid if \( \delta_0 < 1/2 \), but if we suspect that for example \( 1/2 < \delta_0 < 3/2 \), the same reasoning applies if we estimate \( \beta \) from the NB regression of \( \Delta y_t \) on \( \Delta x_t \), obtaining \( \hat{\beta} \), and then estimating \( \gamma \) from \( y_t - \hat{\beta}' x_t \). Higher \( \delta_0 \)'s could be treated by estimating \( \beta \) 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 (25) is well known for an estimate of \( \gamma \) 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 \( \delta_0 < 1/2 \) (so there is the so-called stationary cointegration), but NB ones suffice. When \( \delta_i < \delta_0 \) for some \( i \in \{1, \ldots, l\} \), if \( \gamma < \min_i \delta_i \), our estimate of \( \beta \) 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 \( \gamma \) from \( y_t - \hat{\beta}' x_t \) will lead
to a consistent estimate of $\gamma$ under $H_1$ and the test will gain power. If $\gamma < \delta_i$ 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 $\gamma$ is not going to be affected if we restrict the rate of growth of the bandwidth employed in the estimation of $\gamma$ accordingly.

The conditions for $\hat{\sigma}$ are satisfied for standard semiparametric estimates of $\varphi$, based on the corresponding components of $z_t$. Finally, (26) reflects a trade off between the rate of growth of $m$ and the smoothness of $f$ through the positive relation between this smoothness and $\kappa$. Note however that even if $\kappa$ is very small, (26) 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 $\delta_0 > 1/2$, almost $\sqrt{n}$-consistent estimates of the orders might be necessary, 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 $\hat{\tau}_m(c, \theta, g)$ and $\hat{\tau}_m^0(c, \theta, g)$ respectively.

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

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

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

where, as $n \to \infty$

$$n^{-\varkappa} m^{1 - \max \{\min \{\delta_1, \ldots, \delta_l, 1\}, 1/2\}} \to 0, \quad (27)$$

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

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

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

for which (27) 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 $\delta$ consistent under both hypotheses, or alternatively on residuals $z_t(\hat{\gamma}, \hat{\delta})$, 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 in order 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 1/2, for which we introduce the following condition.

**Assumption 5** $\delta_i \neq 1/2$ for all $i = 0, \ldots, l$.

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

$$\delta_1 \geq \delta_2 \geq \ldots \geq \delta_l \geq 0.$$
Thus, we set
\[
\tilde{\sigma} = \left( \tilde{\sigma}_1, \tilde{\sigma}_2 \right),
\]
where \( \tilde{\sigma}_1 = (\delta_1, \ldots, \delta_{l_1})' \), \( \tilde{\sigma}_2 = (\delta_{l_1+1}, \ldots, \delta_l)' \), with \( \delta_i > 1/2, i = 1, \ldots, l_1 \) and \( \delta_i < 1/2, i = l_1 + 1, \ldots, l \), where \( l_1 = l \) indicates that all integration orders of the \( x \)'s are strictly larger than \( 1/2 \), \( l_1 = 0 \) meaning that all the orders are smaller than \( 1/2 \).

Finally, we impose some conditions on the bandwidth \( m \).

**Assumption 6** If \( l_1 < l \), for \( \epsilon \) in Assumption 1,
\[
m_{\max, i \in [l_1+1, l]} \delta_i^{-1/2} \log^{1/2} n + m^{3+2\epsilon}/n^{2+2\epsilon} \to 0, \quad \text{as } n \to \infty.
\]

**Assumption 7** Assumption 6 holds and if \( l_1 > 0 \)
\[
m/m_{\min, i \in [1, l_1]} \delta_i \to 0, \quad \text{as } n \to \infty.
\]

We introduce some additional notation. Let \( D (L) = C (L) A (L) \), \( W (r) \) be the \( p \times 1 \) Bm with covariance matrix \( \Omega \),
\[
\tilde{W} (r) = 2\pi \zeta + D (1)^{-1} \Omega^{-1} W (r),
\]
and define the \( l_1 \times 1 \) column vector
\[
\tilde{W} (r, \tilde{\sigma}_1) = \left( 0, i_j \right) \frac{1}{\Gamma (\delta_j)} \int_0^r (r-s)^{\delta_j-1} D (1) dW (s) \right)_{j=1, \ldots, l_1},
\]
where \( i_j \) is a \( l \times 1 \) vector of zeroes except 1 in the \( j \)th position; denoting by \( f_{ij} (\lambda) \) the \((i,j)\)-th element of \( f (\lambda) \), let
\[
\mathcal{J}_{22} (0) = \left[ f_{ij} (0) \right]_{i=j=1, \ldots, l_1},
\]
finally set
\[
\Lambda (n) = \text{diag} \left( n^{-\delta_1}, \ldots, n^{-\delta_{l_1}}, n^{-1/2} m_{\delta_{l_1+1}}, \ldots, n^{-1/2} m_{\delta_l} \right).
\]
We next present the null limiting distribution of the statistic \( \hat{\tau}_m (\tilde{\gamma}, \tilde{\sigma}, \tilde{\rho}) \).

**Theorem 1** Under Assumptions 1, 2, 3, 5, 6 and \( H_0 \), as \( n \to \infty \),
\[
\Lambda (n) \hat{\tau}_m (\tilde{\gamma}, \tilde{\sigma}, \tilde{\rho}) \Rightarrow \left( \int_0^1 \frac{\tilde{W} (r, \tilde{\sigma}_1) dW (r)}{Z} \right), \quad \text{(28)}
\]
where \( Z \) is a \((l-l_1) \times 1\) vector of random variables normally distributed with \( E (Z) = 0 \) and
\[
\text{Var} (Z) = \frac{1}{2} \mathcal{J}_{22} (0),
\]
which for all \( r \) is independent from \( \tilde{W} (r, \tilde{\sigma}_1) \) and \( W (r) \); under Assumptions 1, 2, 4, 5, 7 and \( H_0 \), an identical result to (28) applies for \( \hat{\tau}_m^0 (\tilde{\gamma}, \tilde{\sigma}, \tilde{\rho}) \).

The proof of Theorem 1 is given in the Appendix. Denoting by \( \hat{\mathcal{Y}}_m^* \) any of the \( \hat{\mathcal{Y}}_m (\tilde{\gamma}, \tilde{\sigma}, \tilde{\rho}) \), \( \hat{\mathcal{Y}}_m^0 (\tilde{\gamma}, \tilde{\sigma}, \tilde{\rho}) \), we have
Corollary 1 Under the conditions of Theorem 1 and $H_0$, as $n \to \infty$,
\[ \hat{\gamma}_n \to_d \chi^2_1. \]

Remark 1. The distribution on the right of (28) is mixed normal because $\hat{W}(r; \delta)$ and $\hat{W}(r)$ 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 substantially our statistic by replacing the process $z_t(\hat{\gamma}, \delta)$ 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 (22) with $g(\lambda) = \hat{p}(\lambda)$ would not be correct. This problem is avoided by considering $z_t(\hat{\gamma}, \delta)$ instead, noting that this modification should not imply any loss of power, because under $H_1, y_t(\hat{\gamma})$ is underdifferenced, so it is the leading component in $z_t(\hat{\gamma}, \delta)$.

Remark 3. The “zero-frequency” statistic has a direct interpretation relative to that proposed for the white noise situation. Clearly
\[ \bar{\gamma}_{n/2}^0(\gamma, \delta, \hat{\theta}) = \frac{1}{\omega_{y,x}} \sum_{t=1}^{n} x_t \left( \gamma_t(\gamma) - \hat{f}_{yx}(0) \hat{f}_{xx}^{-1}(0) \Delta \delta(x_t) \right), \]
although, in view of our assumptions, this statistic enjoys nice properties only when $\delta_i \geq 1, i = 0, ..., l$, since otherwise the incorrect treatment of the short-memory components by weighting only at frequency zero distorts heavily its asymptotic behaviour.

For the analysis of the parametric test we set

Assumption 8 (i) Assumption 1 holds;

(ii) $f(\lambda; h)$ satisfies the technical smoothness conditions imposed in Assumption 2 of RH;

(iii) There exist estimates $\hat{\gamma}, \hat{\theta}$ of $\gamma, \theta$, respectively, for which Assumption 2 (for $\hat{\gamma}$, without the need of (26)) holds and
\[ \hat{\theta} = \theta + O_p\left(n^{-\frac{1}{2}}\right); \]

(iv) $\delta_i = \delta_0$ for all $i \in [1, l]$, where $\delta_0 > \frac{1}{2}$.

Thus, we have the following results, whose proof is given in the Appendix.

Theorem 2 Under Assumption 8 and $H_0$ an equivalent result to (28) holds for $\hat{\gamma}_n(\hat{\gamma}, \hat{\theta})$.

Corollary 2 Under the conditions of Theorem 2 and $H_0$, as $n \to \infty$,
\[ \hat{\gamma}_n \to_d \chi^2_1. \]

Remark 4. Theorem 2 uncovers another important limitation of the parametric setting, because the result is only given for the $\delta_0 > 1/2$ case, so one should know a priori the (purely) nonstationary condition of the observables. When $\delta_0 < 1/2$ it can be shown that under our conditions $n^{-1/2} \hat{\gamma}_n(\gamma, \theta)$ is asymptotically normal, but even if $\hat{\theta}$ is $\sqrt{n}$-consistent the asymptotic distribution of the properly normalized statistic $\hat{\gamma}_n(\gamma, \hat{\theta})$ 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 \( n^{-1/2} \hat{\gamma} \) is asymptotically normal, but the asymptotic variance of the normalized statistic is not the same as in the case where \( \theta \) is known, and depends on the particular form of \( \hat{\theta} \). Dealing appropriately with the \( \delta < 1/2 \) case was the precise reason why \( x_{t-1} \) replaces (the more natural) \( x_t \) in (14), 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, 2001).

Remark 5. Theorem 2 mainly uses (in a different framework) results from RH, the main distinguishing feature being now that the requirement on the estimate of the order \( \delta_0 \) (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 \( \kappa > \max(0, 1 - \delta_0) \), so almost \( \sqrt{\kappa} \)-consistency of \( \hat{\gamma} \) was needed in case \( \delta_0 \) were just above 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, since the theory for estimating parametrically (hence obtaining \( \sqrt{\kappa} \)-consistent estimates) long and short-memory parameters simultaneously in a multivariate setting is only fully developed in the stationary case. More importantly, \( \gamma \) needs necessarily be estimated from a sort of residuals (like OLS ones), so it is unclear in which sense one can base parametric estimates of \( \gamma \) on these residuals.

Remark 6. Note that in the semiparametric case we need some extra requirements (given in (26)) on the convergence rates of the estimates of the orders apart from \( \kappa > 0 \). This is due to the non-unique differencing applied to the observables in the test statistics, an issue which also arises in Hualde and Robinson (2005).

6 Monte Carlo evidence

In order to offer some evidence of the finite sample behaviour of these test procedures, we present a small Monte Carlo experiment. There are two parts to 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 \( n = 64, 128, 256, 512, 1024 \), 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 \( \gamma \) taking four different values for each corresponding \( \delta \), 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 $\rho$. The parametric test statistic was computed following these steps.

1. Estimate $\delta$ from the raw series $x_t$ as in Beran (1995), fixing the optimizing interval $[\delta - 1, \delta + 1]$.

2. Compute the NB estimate for $\beta$ (see e.g. Robinson and Marinucci, 2001), choosing bandwidths $m = 25, 40, 65, 120, 220$, for $n = 64, 128, 256, 512, 1024$, respectively. Note that the OLS residuals are adequate to estimate $\gamma$ under $H_1$ if $\delta > 0.5$, but if $\delta < 0.5$, the OLS estimate is in general inconsistent.

3. Estimate $\gamma$ with the NB residuals by the same procedure as in 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 (18), noting (11), (12), and estimate $\eta$ by $\tilde{\eta}$ (see (21)).

5. Compute the feasible test statistic $\Xi_n(\hat{\omega}_{y,x}, \hat{\gamma}, \tilde{\eta})$.

We compare the behaviour of $\Xi_n$ with two semiparametric versions of the test, one where the GLS weighting referred before is evaluated at all Fourier frequencies, the other weighting only frequency 0. Given that in the present setting $u_t$ is a white noise process, both test statistics are expected to enjoy a similar behaviour, and this is corroborated by the results of our experiment. In order to calculate the semiparametric statistics, we use the following steps.

1. Estimate $\delta$ and $\gamma$ by $\hat{\delta}$ and $\hat{\gamma}$, by Robinson’s (1995b) versions of the log-periodogram of Geweke and Porter-Hudak (1983) (with bandwidths $m$ given in 2.), without trimming or pooling applied to the series $\bar{\sigma}_t$, $\bar{\delta}_t$, where $\bar{\sigma}_t = x_t1(\delta < 1) + \Delta x_t1(\delta \geq 1)$, $\bar{\delta}_t = \hat{\delta}1(\delta < 1) + \Delta \hat{\delta}1(\delta \geq 1)$, denoting by $\hat{\delta}$ 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 $f(\lambda)$,

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

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

3. Compute the following slightly modified versions of $\hat{\Gamma}_m$, $\hat{\Gamma}_m^0$, which exploit the bivariate framework and add power. The only modification affects $\hat{\tau}_m$, $\hat{\tau}_m^0$ in (22), because instead of this statistic we compute

$$\hat{\tau}_m = \sum_{j=0}^{m} \frac{8_j}{\tilde{f}_{11}(\lambda_j)} \left\{ I_{y(\gamma)x}(\lambda_j) - \frac{\tilde{f}_{12}(\lambda_j)}{\tilde{f}_{22}(\lambda_j)} I_{x(\gamma)x}(\lambda_j) \right\}, \quad \hat{\tau}_m^0 = \sum_{j=0}^{m} \frac{8_j}{\tilde{f}_{11}(0)} \left\{ I_{y(\gamma)x}(\lambda_j) - \frac{\tilde{f}_{12}(0)}{\tilde{f}_{22}(0)} I_{x(\gamma)x}(\lambda_j) \right\},$$

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

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

which diverges under $H_1$, being this the source of additional power.
Results of the proportion of rejections over 10,000 replications when comparing the values of the statistic with the $\alpha = 0.01, 0.05, 0.1$ nominal critical values of the $\chi_1^2$ distribution are reported in Tables 1-4 for the different values of $\delta$. 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 $\hat{\gamma}_m$, with $\hat{\gamma}_m^\circ$ being worst, although for cases $\delta = 1, 1.4$, its behaviour is very similar to that of $\hat{\gamma}_m$. For the parametric test, sizes are best for $\delta = 0.3$, showing also 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 favour the semiparametric statistics in terms of size when $n$ is small, but here, especially the proportions of rejections corresponding to $\hat{\gamma}_m^\circ$ 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 towards the nominal values. Clearly, the parametric test is most powerful for $\delta = 0.6$, with a very good performance relative to other values of $\delta$ 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 $\delta, \gamma = \delta - 0.2$) is important. In terms of power, the semiparametric statistics are comparable to the parametric one (although note that the proportion of rejections are not size corrected). Similarly to the parametric test, the semiparametric ones have also problems to detect the alternative when $\delta = 1, 1.4$ (but not when $\delta = 0.6$) and $\gamma = \delta - 0.2$, the proportion of rejections being here higher 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 analyse the behaviour of the semiparametric statistic $\hat{\gamma}_m$, in a multivariate framework (with three observables), with possibly different integration orders. We generated $\varepsilon_t$ (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

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

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 $A^{-1}(L) = I_3 - \Phi L, \Phi = \text{diag} (0.5, 0.5, 0.5)$ and generate $w_t$ accordingly. Then, denoting $x_t = (x_{1t}, x_{2t})', u_{xt} = (u_{x_1,t}, u_{x_2,t})', x_{1t}, x_{2t}$ are generated from the input processes $u_{x_1,t}, u_{x_2,t}$, as fractionally integrated processes of orders $\delta_1, \delta_2$, respectively, where $\delta_1 = 1.4, 1$, and $\delta_2 = \delta_1 - 0.2, \delta_1 - 0.4$. 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, so the use of $\hat{\gamma}_m$ instead of $\hat{\gamma}_m^\circ$ seems more appropriate, although both statistics are, at least to first order properties, asymptotically equivalent. In
order to compute the statistic and the estimates of the nuisance parameters, we employed the set of bandwidths $m = 12, 20, 31, 60, 110$, for $n = 64, 128, 256, 512, 1024$, respectively. Note that these bandwidths are approximately half of the ones 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 $\delta_0, \delta_1, \delta_2$ 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

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

where $\hat{\delta} = (\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_2)'$. 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 gets rapidly distorted when incorporating information from frequencies which are relatively far from this particular frequency.

As in any semiparametric procedure, the choice of bandwidth is fundamental, and while a more extensive Monte Carlo 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 which behave relatively well.

As mentioned in Section 5 the main issue here is to estimate $\gamma$. Following the strategy described there, we compute the NB estimate from the regression of $\Delta y_t$ on $\Delta x_t$, and estimate $\gamma$ from residuals $y_t - \hat{\beta}' x_t$ by the method described in the first part of the experiment, and finally obtain $\tilde{Y}_m$. Our results are presented in Tables 5, 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, so they are very close to the nominal ones when $n = 1024$ (the statistic being in some cases undersized for large $n$). For small $n$, a decrease in $\delta_2$ implies a decrease in size, this effect disappearing as $n$ increases. In terms of power, the test behaves in a very similar way for both $\delta_1 = 1$ and $\delta_1 = 1.4$ cases. Power decreases with $\delta_2$ and reacts appropriately as $n$ increases. Now we can notice that the test is able to detect the alternative $\gamma = \delta_1 - 0.2$, 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 described a more realistic situation than the first one, are certainly encouraging, noting that for simplicity we did not apply the provision made in Section 5 about the rate of growth of the bandwidth used in the estimation of $\gamma$ (in comparison to that used in the estimation of $\beta$), and also that we did not use 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 could lead to even better finite sample results.

APPENDIX
A Proof of Theorem 1

The proof follows on showing

\[
\Lambda (n) \tilde{\tau}_m (\gamma, \delta, p) \Rightarrow \left( \int_0^1 \tilde{W} (r; \tilde{\vartheta}_1) \, d\tilde{W} (r) \right),
\]

(29)

\[
\Lambda (n) \left( \tilde{\tau}_m (\tilde{\gamma}, \tilde{\delta}, \tilde{p}) - \tilde{\tau}_m (\gamma, \delta, p) \right) = o_p (1).
\]

(30)

First, (30) follows directly from Hualde and Robinson (2005). Next, the proof of (29) 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

\[
\Lambda (n) \tilde{\tau}_m (\gamma, \delta, p) = \Lambda (n) \sum_{t=2}^{n} \left( x_{t-1}^{(1)} p (0) D (1) \varepsilon_t , e_t' \sum_{v=1}^{t-1} \varepsilon_{t-v} \varepsilon_v , ..., e_t' \sum_{v=1}^{t-1} \varepsilon_{t-v} \varepsilon_v \right)' + o_p (1),
\]

where for \( k \in [l_1 + 1, l] \)

\[
c_t^k = \frac{1}{n^{l}} e_t' \sum_{j=1}^{m} \rho^k (\lambda_j) \cos (t \lambda_j),
\]

and

\[
\rho^k (\lambda) = \left( 1 - e^{-i \lambda} \right)^{-(\delta_k / 2)} \Omega^{(1)} D' \left( e^{i \lambda} \right) p' (\lambda) i_k D (e^{-i \lambda}) \Omega^{(2)}
\]

\[
+ \left( 1 - e^{i \lambda} \right)^{-(\delta_k / 2)} \Omega^{(1)} D' \left( e^{i \lambda} \right) i_k p (-\lambda) D (e^{i \lambda}) \Omega^{(2)}.
\]

Then, (29) 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

\[
c_1 (r) = \left( \frac{1}{\eta^{1/2} \lambda_{1} \left[ n \right]}, ..., \frac{1}{\eta^{1/2} \lambda_{l_1} \left[ n \right]} \right) \cdot
\]

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

\[
c_3 (r) = \left( \frac{\lambda^{d_{1}}}{m^{l/2}} \sum_{t=1}^{[n^{l/2}]} e_t' \sum_{v=1}^{t-1} \varepsilon_{t-v} \varepsilon_v , ..., \frac{\lambda^{d_{l}}}{m^{l/2}} \sum_{t=1}^{[n^{l/2}]} e_t' \sum_{v=1}^{t-1} \varepsilon_{t-v} \varepsilon_v \right)'.
\]

Thus, (29) 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.

B Proof of Theorem 2

The proof follows on showing that

\[
n^{-\delta_0} \tilde{\tau}_n (\gamma, \theta) \Rightarrow \int_0^1 \tilde{W} (r; \tilde{\vartheta}_0) \, d\tilde{W} (r),
\]

(31)
where
\[
\tilde{W}(r; \delta_0) = (0_t, I_t) D_1 (1) \frac{1}{1 (\delta_0)} \int_{0}^{r} (r - s)^{\delta_0 - 1} dW(s), \quad \tilde{W}(r) = 2\pi \zeta' \left( D (1)^{-1} \right)' \Omega^{-1} W(r),
\]
and
\[
\tilde{\gamma}_n(\tilde{\gamma}, \tilde{\theta}) - \tilde{\gamma}_n(\gamma, \theta) = a_p(n^\delta). \tag{32}
\]
First, noting that under \( H_0 \)
\[
\tilde{\gamma}_n(\gamma, \theta) = \sum_{j=1}^{n} w_x(-\lambda_j) p(\lambda_j; \theta) w_u(\lambda_j),
\]
(31) follows from a trivial multivariate extension of Propositions 1-3 of RH, the only significative difference being now 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 (32), we only show
\[
\tilde{\gamma}_n(\tilde{\gamma}, \tilde{\theta}) - \tilde{\gamma}_n(\gamma, \theta) = a_p(n^\delta), \tag{33}
\]
the rest of the proof following directly by Propositions 7 and 10 of RH. The result in (33) corresponds to Proposition 9 in RH, but our present situation is more delicate because we just require that \( \tilde{\gamma} \) satisfies (25) 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 (33) is
\[
\frac{1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \sum_{t=2}^{n} \sum_{m=1}^{t-1} a_m(\delta - \tilde{\gamma}) u_{t-m} e^{it\lambda_j} \sum_{s=1}^{n} x_s' e^{-is\lambda_j}
= \frac{R-1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \sum_{t=2}^{n} \sum_{m=1}^{t-1} a_m^{(r)}(0) u_{t-m} e^{it\lambda_j} \sum_{s=1}^{n} x_s' e^{-is\lambda_j}
\]
\[
+ \frac{1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \sum_{t=2}^{n} \sum_{m=1}^{t-1} a_m^{(r)}(\delta - \tilde{\gamma}) u_{t-m} e^{it\lambda_j} \sum_{s=1}^{n} x_s' e^{-is\lambda_j}, \tag{34}
\]
where \( p(\lambda) = p(\lambda; \theta), a_s^{(r)}(c) = d'^r a_s(c) / dc^r, \) and \( |\gamma - \delta| \leq |\gamma - \delta| \). First, as in RH, the second term in (35) can be shown to be of smaller order for \( R \) large enough. Next, we show that the \( r \)-th term in (34) is \( O_p(n^{-\kappa + \delta + \epsilon}) \) for any \( \epsilon > 0 \). First
\[
E \left( \frac{1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \sum_{t=2}^{n} \sum_{m=1}^{t-1} a_m^{(r)}(0) u_{t-m} e^{it\lambda_j} \sum_{s=1}^{n} x_s' e^{-is\lambda_j} \right)
= \frac{1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \int_{-\pi}^{\pi} \sum_{s=1}^{n} a_s^{(r)}(0) e^{is\lambda_j} D_{n-s}(\lambda_j - \mu) \sum_{t=0}^{n-1} a_t(\delta) e^{-it\lambda_j} D_{n-t}(\mu - \lambda_j) f(\mu) \xi d\mu,
\]
\[
\text{where } D_t(\lambda) = \sum_{k=1}^{t} e^{ik\lambda} \text{ is the Dirichlet kernel. Noting that for any } \lambda, p(\lambda) f(\lambda) \xi \text{ is identically zero, by periodicity, (36) can be written as }
\]
\[
\frac{1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \int_{-\pi}^{\pi} \sum_{s=1}^{n} a_s^{(r)}(0) e^{is\lambda_j} D_{n-s}(-\mu) \sum_{t=0}^{n-1} a_t(\delta) e^{-it\lambda_j} D_{n-t}(\mu) [f(\mu + \lambda_j) - f(\lambda_j)] \xi d\mu,
\]
\[25\]
which, by summation by parts, is
\[
\frac{1}{2\pi i} \sum_{j=1}^{n} \mu (\lambda_j) \int_{-\pi}^{\pi} \sum_{s=1}^{n-1} a_s^{(r)} (0) e^{is\lambda_j} D_{n-s}(-\mu) \left\{ a_{n-1} D_t (\mu) \left[ f (\mu + \lambda_j) - f (\lambda_j) \right] \xi \sum_{t=0}^{n-1} e^{-it\lambda_j} d\mu \right. \\
- \left. \left[ f (\mu + \lambda_j) - f (\lambda_j) \right] \xi \sum_{t=0}^{n-2} (a_{t+1} D_{n-t-1}(\mu) - a_t D_{n-t-1}(\mu)) \sum_{h=0}^{t} e^{-ih\lambda_j} d\mu \right\}. 
\]
(37)

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 (37) is bounded in norm by
\[
K |a_{n-1}| \sum_{s=1}^{n-1} |a_s^{(r)} (0)| \int_{-\pi}^{\pi} |\mu||D_{n-s}(-\mu)| d\mu, 
\]
(38)

where throughout \(K\) denotes a generic finite positive constant. Noting that for \(0 < \lambda < \pi,\)
\[
|D_t (\lambda)| < K \min \left\{ |\lambda|^{-1}, t \right\},
\]
(39)
(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, (38) is bounded by
\[
Kn^{\delta-1} \sum_{s=1}^{n-1} \log^{r-1} \frac{s}{n} = O \left( n^{\delta-1} \log^{r} s \right). 
\]

Regarding the second term in (37), noting that
\[
a_{t+1} D_{n-t-1}(\mu) - a_t D_{n-t-1}(\mu) = (a_{t+1} - a_t) D_{n-t-1}(\mu) - e^{i(\mu-t)} \mu a_t,
\]
(40)
the contribution of the first term on the right of (40) to the second term of (37) is 0 for \(\beta = 1,\) as in this case \(a_{t+1} = a_t, \) \(t = 0, ..., n - 2.\) For \(\beta \neq 1,\) this contribution is bounded in modulus by
\[
K n^{-1} \sum_{j=1}^{n} \left\{ \int_{-\pi}^{\pi} \left| \sum_{s=1}^{n-1} a_s^{(r)} (0) e^{is\lambda_j} D_{n-s}(-\mu) \right|^2 \| f (\mu + \lambda_j) - f (\lambda_j) \| d\mu \right\}^{\frac{1}{2}} \\
\times \left\{ \int_{-\pi}^{\pi} \left| \sum_{t=0}^{n-2} (a_{t+1} - a_t) D_{n-t-1}(\mu) (D_t(-\lambda_j) + 1) \right|^2 \| f (\mu + \lambda_j) - f (\lambda_j) \| d\mu \right\}^{\frac{1}{2}}. 
\]
(41)

The term in the first braces is bounded uniformly in \(j\) by
\[
K \int_{-\pi}^{\pi} |\mu| \left| \sum_{s=1}^{n-1} \sum_{t=1}^{n-1-n} a_s^{(r)} (0) a_t^{(r)} (0) e^{i(s-t)\lambda_j} D_{n-s}(-\mu) D_{n-t}(\mu) \right| d\mu \\
\leq K \int_{-\pi}^{\pi} |\mu| \left( \sum_{s=1}^{n-1} \left| a_s^{(r)} (0) \right|^2 \right) |D_{n-s}(-\mu)|^2 d\mu \leq K \log n,
\]
because by Zygmund (1977)
\[
\int_{-\pi}^{\pi} |D_n (\mu)| d\mu = O (\log n). 
\]

26
Next, the term in the second braces is bounded by

\[
K \int -\pi^{\infty} \left| \frac{n-2}{n} \sum_{t=0}^{n-2} \sum_{s=0}^{n-s} \left( a_{t+1}(\delta) - a_{t}(\delta) \right) D_{n-t-1}(\mu) (D_{t}(\lambda) + 1) \right. \\
\times \left. \left( a_{s+1}(\delta) - a_{s}(\delta) \right) D_{n-s-1}(-\mu) (D_{s}(\lambda) + 1) \right) d\mu \\
= O \left( j^{-2} n^2 \log n \left( \sum_{t=1}^{n} \frac{1}{t^{d-2}} \right)^2 \right)
\]

by Lemma C.1 of RH and (39), which is \( O (j^{-2} n^2 \log n (\delta < 1) + j^{-2} n^{2\delta} \log n (\delta > 1)) \), implying that (41) is \( O (\log^2 n (\delta < 1) + n^{\delta-1/2} \log n (\delta > 1)) \). Finally, the contribution of the second term on the right of (40) to the second term of (37) is bounded in modulus by

\[
Kn^{-1} \sum_{j=1}^{n} \left\{ \int_{-\pi}^{\pi} |\mu|^2 \left| \sum_{s=1}^{n-s} a_s^{(r)}(0) e^{i\lambda s} D_{n-s}(-\mu) \right|^2 d\mu \right\}^{1/2} \\
\int_{-\pi}^{\pi} e^{-i(s-t)^t} d\mu = 2\pi, \quad s = t; \quad = 0, \text{ otherwise,}
\]

the second is bounded by \( K \sum_{j=1}^{n} a_j^2(\delta) |D_{t}(\lambda)|^2 \), so that (42) is bounded by \( Kn^{-1} \log^2 n \sum_{j=1}^{n} \left\{ n^{2\delta+1} j^{-2} \right\}^{1/2} \), which is \( O (n^{\delta-1/2} \log^{1+1} n) \), implying that the left of (36) is \( O (n^{1/2} \log n (\delta < 1) + n^{\delta-1/2} \log^{1+1} n (\delta > 1)) \).

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

\[
Var \left( \frac{1}{2\pi n} \sum_{j=1}^{n} p(\lambda_j) \sum_{r=2}^{n} \sum_{j=1}^{n-r} a_j^{(r)}(0) u_{r-j} e^{i\lambda_j} \sum_{s=1}^{n} e^{-i\lambda_j} \right) = O (n^{2\delta+\epsilon})
\]

for any \( \epsilon > 0 \), which implies that the \( r \)-th term of (34) is \( O_p \left( n^{-(\kappa+\delta+\epsilon)} \right) \), for any \( \epsilon > 0 \), so that (33) holds for any \( \kappa > 0 \) on choosing \( \epsilon < \kappa \), to conclude the proof.

References


### Table 1

<table>
<thead>
<tr>
<th>( \gamma \backslash n \alpha )</th>
<th>( .1 )</th>
<th>( .05 )</th>
<th>( .10 )</th>
<th>( .1 )</th>
<th>( .05 )</th>
<th>( .10 )</th>
<th>( .0 )</th>
<th>( .05 )</th>
<th>( .10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Xi_n )</td>
<td>64</td>
<td>.015</td>
<td>.059</td>
<td>.113</td>
<td>.030</td>
<td>.087</td>
<td>.149</td>
<td>.073</td>
<td>.162</td>
</tr>
<tr>
<td>128</td>
<td>.012</td>
<td>.056</td>
<td>.108</td>
<td>.032</td>
<td>.109</td>
<td>.184</td>
<td>.098</td>
<td>.245</td>
<td>.355</td>
</tr>
<tr>
<td>256</td>
<td>.011</td>
<td>.052</td>
<td>.103</td>
<td>.065</td>
<td>.191</td>
<td>.290</td>
<td>.242</td>
<td>.470</td>
<td>.595</td>
</tr>
<tr>
<td>512</td>
<td>.012</td>
<td>.055</td>
<td>.104</td>
<td>.166</td>
<td>.362</td>
<td>.487</td>
<td>.594</td>
<td>.804</td>
<td>.879</td>
</tr>
<tr>
<td>1024</td>
<td>.009</td>
<td>.053</td>
<td>.101</td>
<td>.396</td>
<td>.642</td>
<td>.754</td>
<td>.931</td>
<td>.981</td>
<td>.993</td>
</tr>
<tr>
<td>( \hat{T}_m )</td>
<td>64</td>
<td>.036</td>
<td>.051</td>
<td>.068</td>
<td>.090</td>
<td>.118</td>
<td>.141</td>
<td>.191</td>
<td>.237</td>
</tr>
<tr>
<td>128</td>
<td>.037</td>
<td>.060</td>
<td>.083</td>
<td>.128</td>
<td>.173</td>
<td>.208</td>
<td>.304</td>
<td>.371</td>
<td>.412</td>
</tr>
<tr>
<td>256</td>
<td>.043</td>
<td>.076</td>
<td>.111</td>
<td>.196</td>
<td>.257</td>
<td>.304</td>
<td>.490</td>
<td>.567</td>
<td>.611</td>
</tr>
<tr>
<td>512</td>
<td>.030</td>
<td>.063</td>
<td>.099</td>
<td>.223</td>
<td>.306</td>
<td>.364</td>
<td>.653</td>
<td>.730</td>
<td>.769</td>
</tr>
<tr>
<td>1024</td>
<td>.029</td>
<td>.066</td>
<td>.109</td>
<td>.295</td>
<td>.398</td>
<td>.465</td>
<td>.856</td>
<td>.899</td>
<td>.920</td>
</tr>
</tbody>
</table>

Proportion of rejections over 10000 replications of \( \Xi_n, \hat{T}_m \) when compared with the critical value of a \( \chi^2 \) distribution with nominal size \( \alpha \). For \( \gamma = \delta \) this is simulated size and for \( \gamma < \delta \) simulated power.

### Table 2

<table>
<thead>
<tr>
<th>( \gamma \backslash n \alpha )</th>
<th>( .6 )</th>
<th>( .4 )</th>
<th>( .2 )</th>
<th>( .0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( .1 )</td>
<td>( .05 )</td>
<td>( .10 )</td>
<td>( .0 )</td>
<td>( .05 )</td>
</tr>
<tr>
<td>( \Xi_n )</td>
<td>64</td>
<td>.035</td>
<td>.098</td>
<td>.159</td>
</tr>
<tr>
<td>128</td>
<td>.028</td>
<td>.085</td>
<td>.142</td>
<td>.120</td>
</tr>
<tr>
<td>256</td>
<td>.022</td>
<td>.070</td>
<td>.124</td>
<td>.246</td>
</tr>
<tr>
<td>1024</td>
<td>.016</td>
<td>.061</td>
<td>.113</td>
<td>.889</td>
</tr>
<tr>
<td>( \hat{T}_m )</td>
<td>64</td>
<td>.051</td>
<td>.088</td>
<td>.130</td>
</tr>
<tr>
<td>128</td>
<td>.050</td>
<td>.092</td>
<td>.137</td>
<td>.277</td>
</tr>
<tr>
<td>256</td>
<td>.051</td>
<td>.100</td>
<td>.149</td>
<td>.407</td>
</tr>
<tr>
<td>512</td>
<td>.034</td>
<td>.082</td>
<td>.132</td>
<td>.496</td>
</tr>
<tr>
<td>1024</td>
<td>.028</td>
<td>.078</td>
<td>.133</td>
<td>.649</td>
</tr>
<tr>
<td>( \hat{\theta}_m )</td>
<td>64</td>
<td>.053</td>
<td>.092</td>
<td>.136</td>
</tr>
<tr>
<td>128</td>
<td>.056</td>
<td>.103</td>
<td>.150</td>
<td>.296</td>
</tr>
<tr>
<td>256</td>
<td>.057</td>
<td>.112</td>
<td>.163</td>
<td>.431</td>
</tr>
<tr>
<td>512</td>
<td>.039</td>
<td>.093</td>
<td>.143</td>
<td>.526</td>
</tr>
<tr>
<td>1024</td>
<td>.033</td>
<td>.088</td>
<td>.147</td>
<td>.684</td>
</tr>
</tbody>
</table>

Proportion of rejections over 10000 replications of \( \Xi_n, \hat{T}_m \) when compared with the critical value of a \( \chi^2 \) distribution with nominal size \( \alpha \). For \( \gamma = \delta \) this is simulated size and for \( \gamma < \delta \) simulated power.
### Table 3

**Proportion of Rejections of $\Xi_n$, $\hat{T}_m$, $\hat{T}_m^0$, Bivariate Case, $\delta = 1$**

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$n \backslash \alpha$</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Xi_n$</td>
<td>64</td>
<td>.052</td>
<td>.118</td>
<td>.188</td>
<td>.099</td>
<td>.175</td>
<td>.245</td>
<td>.374</td>
<td>.463</td>
<td>.528</td>
<td>.752</td>
<td>.795</td>
<td>.826</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>.037</td>
<td>.094</td>
<td>.159</td>
<td>.071</td>
<td>.169</td>
<td>.253</td>
<td>.402</td>
<td>.501</td>
<td>.576</td>
<td>.885</td>
<td>.908</td>
<td>.921</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>.023</td>
<td>.077</td>
<td>.137</td>
<td>.077</td>
<td>.206</td>
<td>.301</td>
<td>.447</td>
<td>.549</td>
<td>.616</td>
<td>.970</td>
<td>.977</td>
<td>.981</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>.019</td>
<td>.069</td>
<td>.124</td>
<td>.121</td>
<td>.287</td>
<td>.403</td>
<td>.520</td>
<td>.612</td>
<td>.670</td>
<td>.998</td>
<td>.999</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>.014</td>
<td>.061</td>
<td>.114</td>
<td>.210</td>
<td>.418</td>
<td>.543</td>
<td>.633</td>
<td>.712</td>
<td>.747</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{T}_m$</td>
<td>64</td>
<td>.048</td>
<td>.103</td>
<td>.152</td>
<td>.179</td>
<td>.245</td>
<td>.304</td>
<td>.498</td>
<td>.548</td>
<td>.586</td>
<td>.842</td>
<td>.866</td>
<td>.879</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>.042</td>
<td>.094</td>
<td>.147</td>
<td>.181</td>
<td>.245</td>
<td>.304</td>
<td>.663</td>
<td>.707</td>
<td>.735</td>
<td>.971</td>
<td>.980</td>
<td>.984</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>.038</td>
<td>.089</td>
<td>.138</td>
<td>.204</td>
<td>.271</td>
<td>.327</td>
<td>.840</td>
<td>.869</td>
<td>.885</td>
<td>.999</td>
<td>.999</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>.029</td>
<td>.076</td>
<td>.129</td>
<td>.181</td>
<td>.255</td>
<td>.316</td>
<td>.953</td>
<td>.965</td>
<td>.971</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>.025</td>
<td>.075</td>
<td>.127</td>
<td>.176</td>
<td>.260</td>
<td>.331</td>
<td>.995</td>
<td>.997</td>
<td>.998</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{T}_m^0$</td>
<td>64</td>
<td>.049</td>
<td>.102</td>
<td>.152</td>
<td>.182</td>
<td>.247</td>
<td>.307</td>
<td>.500</td>
<td>.554</td>
<td>.593</td>
<td>.849</td>
<td>.870</td>
<td>.884</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>.043</td>
<td>.095</td>
<td>.146</td>
<td>.184</td>
<td>.249</td>
<td>.311</td>
<td>.668</td>
<td>.713</td>
<td>.739</td>
<td>.973</td>
<td>.982</td>
<td>.985</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>.038</td>
<td>.090</td>
<td>.139</td>
<td>.208</td>
<td>.276</td>
<td>.333</td>
<td>.844</td>
<td>.872</td>
<td>.888</td>
<td>.999</td>
<td>.999</td>
<td>.999</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>.029</td>
<td>.077</td>
<td>.131</td>
<td>.183</td>
<td>.260</td>
<td>.322</td>
<td>.954</td>
<td>.965</td>
<td>.971</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>.024</td>
<td>.074</td>
<td>.128</td>
<td>.178</td>
<td>.265</td>
<td>.336</td>
<td>.995</td>
<td>.997</td>
<td>.998</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Proportion of rejections over 10000 replications of $\Xi_n$, $\hat{T}_m$ when compared with the critical value of a $\chi^2_1$ distribution with nominal size $\alpha$. For $\gamma = \delta$ this is simulated size and for $\gamma < \delta$ simulated power.

### Table 4

**Proportion of Rejections of $\Xi_n$, $\hat{T}_m$, $\hat{T}_m^0$, Bivariate Case, $\delta = 1.4$**

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$n \backslash \alpha$</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Xi_n$</td>
<td>64</td>
<td>.046</td>
<td>.106</td>
<td>.176</td>
<td>.070</td>
<td>.118</td>
<td>.174</td>
<td>.288</td>
<td>.350</td>
<td>.409</td>
<td>.660</td>
<td>.697</td>
<td>.729</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>.030</td>
<td>.087</td>
<td>.150</td>
<td>.037</td>
<td>.086</td>
<td>.149</td>
<td>.287</td>
<td>.350</td>
<td>.404</td>
<td>.779</td>
<td>.805</td>
<td>.825</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>.020</td>
<td>.072</td>
<td>.134</td>
<td>.026</td>
<td>.085</td>
<td>.140</td>
<td>.309</td>
<td>.370</td>
<td>.423</td>
<td>.889</td>
<td>.907</td>
<td>.916</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>.018</td>
<td>.056</td>
<td>.125</td>
<td>.030</td>
<td>.099</td>
<td>.163</td>
<td>.346</td>
<td>.412</td>
<td>.465</td>
<td>.970</td>
<td>.975</td>
<td>.979</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>.014</td>
<td>.057</td>
<td>.109</td>
<td>.052</td>
<td>.138</td>
<td>.208</td>
<td>.407</td>
<td>.471</td>
<td>.512</td>
<td>.995</td>
<td>.996</td>
<td>.996</td>
</tr>
<tr>
<td>$\hat{T}_m$</td>
<td>64</td>
<td>.052</td>
<td>.113</td>
<td>.164</td>
<td>.080</td>
<td>.124</td>
<td>.174</td>
<td>.278</td>
<td>.325</td>
<td>.368</td>
<td>.771</td>
<td>.798</td>
<td>.816</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>.041</td>
<td>.092</td>
<td>.142</td>
<td>.084</td>
<td>.126</td>
<td>.174</td>
<td>.404</td>
<td>.452</td>
<td>.490</td>
<td>.898</td>
<td>.913</td>
<td>.924</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>.031</td>
<td>.078</td>
<td>.131</td>
<td>.091</td>
<td>.145</td>
<td>.196</td>
<td>.556</td>
<td>.606</td>
<td>.642</td>
<td>.973</td>
<td>.978</td>
<td>.980</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>.024</td>
<td>.074</td>
<td>.121</td>
<td>.079</td>
<td>.135</td>
<td>.190</td>
<td>.677</td>
<td>.721</td>
<td>.746</td>
<td>.996</td>
<td>.997</td>
<td>.998</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>.022</td>
<td>.071</td>
<td>.123</td>
<td>.075</td>
<td>.138</td>
<td>.197</td>
<td>.780</td>
<td>.813</td>
<td>.835</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{T}_m^0$</td>
<td>64</td>
<td>.053</td>
<td>.113</td>
<td>.165</td>
<td>.080</td>
<td>.124</td>
<td>.175</td>
<td>.278</td>
<td>.327</td>
<td>.370</td>
<td>.772</td>
<td>.799</td>
<td>.818</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>.041</td>
<td>.091</td>
<td>.141</td>
<td>.084</td>
<td>.126</td>
<td>.176</td>
<td>.404</td>
<td>.451</td>
<td>.491</td>
<td>.898</td>
<td>.914</td>
<td>.924</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>.030</td>
<td>.079</td>
<td>.131</td>
<td>.092</td>
<td>.145</td>
<td>.199</td>
<td>.554</td>
<td>.607</td>
<td>.640</td>
<td>.973</td>
<td>.977</td>
<td>.980</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>.024</td>
<td>.073</td>
<td>.121</td>
<td>.080</td>
<td>.137</td>
<td>.192</td>
<td>.675</td>
<td>.722</td>
<td>.748</td>
<td>.995</td>
<td>.996</td>
<td>.997</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>.022</td>
<td>.070</td>
<td>.122</td>
<td>.075</td>
<td>.139</td>
<td>.199</td>
<td>.779</td>
<td>.814</td>
<td>.835</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Proportion of rejections over 10000 replications of $\Xi_n$, $\hat{T}_m$ when compared with the critical value of a $\chi^2_1$ distribution with nominal size $\alpha$. For $\gamma = \delta$ this is simulated size and for $\gamma < \delta$ simulated power.
## Table 5

**Proportion of Rejections of \( \hat{T}_n(\bar{Z}, \bar{\delta}, \bar{\rho}) \), Trivariate Case, \( \delta_1 = 1 \)**

<table>
<thead>
<tr>
<th>( \delta_2 )</th>
<th>( \gamma = 1 )</th>
<th>( \gamma = .8 )</th>
<th>( \gamma = .6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha = .01 )</td>
<td>( \alpha = .05 )</td>
<td>( \alpha = .10 )</td>
</tr>
<tr>
<td>1</td>
<td>.132</td>
<td>.183</td>
<td>.222</td>
</tr>
<tr>
<td>64 .8</td>
<td>.071</td>
<td>.106</td>
<td>.140</td>
</tr>
<tr>
<td>.6</td>
<td>.048</td>
<td>.078</td>
<td>.109</td>
</tr>
<tr>
<td>1</td>
<td>.089</td>
<td>.135</td>
<td>.171</td>
</tr>
<tr>
<td>128 .8</td>
<td>.038</td>
<td>.067</td>
<td>.096</td>
</tr>
<tr>
<td>.6</td>
<td>.029</td>
<td>.055</td>
<td>.081</td>
</tr>
<tr>
<td>1</td>
<td>.071</td>
<td>.112</td>
<td>.153</td>
</tr>
<tr>
<td>256 .8</td>
<td>.026</td>
<td>.054</td>
<td>.086</td>
</tr>
<tr>
<td>.6</td>
<td>.021</td>
<td>.044</td>
<td>.072</td>
</tr>
<tr>
<td>1</td>
<td>.039</td>
<td>.074</td>
<td>.110</td>
</tr>
<tr>
<td>512 .8</td>
<td>.016</td>
<td>.046</td>
<td>.083</td>
</tr>
<tr>
<td>.6</td>
<td>.011</td>
<td>.036</td>
<td>.073</td>
</tr>
<tr>
<td>1</td>
<td>.019</td>
<td>.048</td>
<td>.080</td>
</tr>
<tr>
<td>1024 .8</td>
<td>.009</td>
<td>.040</td>
<td>.084</td>
</tr>
<tr>
<td>.6</td>
<td>.010</td>
<td>.047</td>
<td>.091</td>
</tr>
</tbody>
</table>

Proportion of rejections over 100000 replications of \( \hat{T}_n(\bar{Z}, \bar{\delta}, \bar{\rho}) \) when compared with the critical value of a \( \chi^2_2 \) distribution with nominal size \( \alpha \). For \( \gamma = \delta_1 \) this is simulated size and for \( \gamma < \delta_1 \) simulated power.

## Table 6

**Proportion of Rejections of \( \hat{T}_n(\bar{Z}, \bar{\delta}, \bar{\rho}) \), Trivariate Case, \( \delta_1 = 1.4 \)**

<table>
<thead>
<tr>
<th>( \delta_2 )</th>
<th>( \gamma = 1.4 )</th>
<th>( \gamma = 1.2 )</th>
<th>( \gamma = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha = .01 )</td>
<td>( \alpha = .05 )</td>
<td>( \alpha = .10 )</td>
</tr>
<tr>
<td>1.4</td>
<td>.168</td>
<td>.226</td>
<td>.271</td>
</tr>
<tr>
<td>64 1.2</td>
<td>.122</td>
<td>.176</td>
<td>.219</td>
</tr>
<tr>
<td>1</td>
<td>.101</td>
<td>.149</td>
<td>.190</td>
</tr>
<tr>
<td>1.4</td>
<td>.113</td>
<td>.163</td>
<td>.204</td>
</tr>
<tr>
<td>128 1.2</td>
<td>.073</td>
<td>.118</td>
<td>.157</td>
</tr>
<tr>
<td>1</td>
<td>.070</td>
<td>.114</td>
<td>.151</td>
</tr>
<tr>
<td>1.4</td>
<td>.085</td>
<td>.135</td>
<td>.179</td>
</tr>
<tr>
<td>256 1.2</td>
<td>.058</td>
<td>.100</td>
<td>.146</td>
</tr>
<tr>
<td>1</td>
<td>.048</td>
<td>.089</td>
<td>.130</td>
</tr>
<tr>
<td>1.4</td>
<td>.046</td>
<td>.086</td>
<td>.124</td>
</tr>
<tr>
<td>512 1.2</td>
<td>.034</td>
<td>.068</td>
<td>.108</td>
</tr>
<tr>
<td>1</td>
<td>.031</td>
<td>.067</td>
<td>.103</td>
</tr>
<tr>
<td>1.4</td>
<td>.023</td>
<td>.053</td>
<td>.085</td>
</tr>
<tr>
<td>1024 1.2</td>
<td>.017</td>
<td>.047</td>
<td>.079</td>
</tr>
<tr>
<td>1</td>
<td>.016</td>
<td>.048</td>
<td>.083</td>
</tr>
</tbody>
</table>

Proportion of rejections over 100000 replications of \( \hat{T}_n(\bar{Z}, \bar{\delta}, \bar{\rho}) \) when compared with the critical value of a \( \chi^2_2 \) distribution with nominal size \( \alpha \). For \( \gamma = \delta_1 \) this is simulated size and for \( \gamma < \delta_1 \) simulated power.