Diagnostic Testing for Cointegration

P.M. Robinson*

London School of Economics

May 14, 2007

Abstract

We develop a sequence of tests for specifying the cointegrating rank of, possibly fractional, multiple time series. Memory parameters of observables are treated as unknown, as are those of possible cointegrating errors. The individual test statistics have standard null asymptotics, and are related to Hausman specification test statistics: when the memory parameter is common to several series, an estimate of this parameter based on the assumption of no cointegration achieves an efficiency improvement over estimates based on individual series, whereas if the series are cointegrated the former estimate is generally inconsistent. However, a computationally simpler but asymptotically equivalent approach, which avoids explicit computation of the "efficient" estimate, is instead pursued here. Two versions of it are initially proposed, followed by one that robustifies to possible inequality between memory parameters of observables. Throughout, a semiparametric approach is pursued, modelling serial dependence only at frequencies near the origin, with the goal of validity under

*Corresponding author: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK; tel. +44-20-7955-7516; fax: +44-20-7955-6592. Email address: p.m.robinson@lse.ac.uk

broad circumstances and computational convenience. The main development is in terms of stationary series, but an extension to nonstationary ones is also described. The algorithm for estimating cointegrating rank entails carrying out such tests based on potentially all subsets of two or more of the series, though outcomes of previous tests may render some or all subsequent ones unnecessary. A Monte Carlo study of finite sample performance is included.

JEL Classification: C32.

Keywords: Fractional cointegration; Diagnostic testing; Specification testing; Cointegrating rank; Semiparametric estimation.

1. INTRODUCTION

The potential for detecting cointegration in economic and financial time series has expanded with a wider realization that the phenomenon is not restricted to the "unit root setting" of $I(1)$ observable series, where cointegrating errors are $I(0)$ (or to its familiar extensions to $I(2)$ series, or $I(1)$ series with deterministic trends). Fractional processes provide a significant mathematical extension of these. We say that a $p \times 1$ vector z_t of jointly dependent $I(\delta)$ processes, for some positive, real integration order δ that need not be an integer, is cointegrated if there exists a linear combination of elements of z_t that is $I(\gamma)$, for some real $\gamma \in [0, \delta)$. Indeed, z_t can even be stationary, in which case $\delta < \frac{1}{2}$. As in the traditional unit root setting, there could be up to $p-1$ cointegrating relations, however we permit these to have real-valued and possibly different integration orders.

It is useful to estimate cointegrating relations; these can be used to test hypotheses of interest (such as PPP) and to improve the precision of forecasts, for example. However, important initial questions are the existence of cointegration, and the cointegrating rank, that is the number of cointegrating relations. These have been addressed in the unit root setting, e.g. by Johansen (1988, 1991, 1995), Phillips and Ouliaris (1988, 1990). Here it has been usual to take for granted the $I(1)$ assumption on z_t, albeit with the presumption of pre-testing. Procedures based on known integration orders can be invalidated if they are mis-specified. With sufficiently long series, it may thus seem more satisfactory to adopt an agnostic approach, by estimating integration orders from the data, though estimates of course cannot be taken as synonymous with true values, and it seems desirable to investigate the existence of cointegration in the presence of unknown δ (and γ). There has been some, rather limited, theoretical investigation of this problem (see e.g. Robinson and Yajima, 2002).

The present paper studies a conceptually and computationally simple diagnostic

statistic, based on the Hausman specification testing idea, for testing the null hypothesis of no cointegration. A version of it was previously proposed (in bivariate series) by Marinucci and Robinson (2001), but though they applied it empirically its theoretical properties were only heuristically discussed. Here we not only provide a more formal treatment of its asymptotic null distribution, but also discuss its consistency properties, and its robustness to departures from mainstream assumptions; this in particular leads us to propose a robustified version. We propose an algorithm for estimating cointegrating rank in series of dimension 3 or more, which uses one of our tests in a sequential manner on each step. To prevent the discussion becoming too complicated we do not incorporate deterministic trends, indeed for partly expository purposes the main discussion is in terms of stationary observables, though we later describe an extension to nonstationary series.

The following section describes a stationary setting in which cointegration, with some rank, may or may not occur. Section 3 defines two basic test statistics, which assume integration orders of all observables are equal. Their asymptotic null distributions, and the consistency properties of one of them, are presented in Section 4. Section 5 introduces and justifies a robustified statistic which avoids the assumption of a equality of all integration orders, and includes an asymptotic local power comparison. Section 6 presents the algorithm for estimating cointegrating rank. Section 7 describes an extension to nonstationary series. A Monte Carlo study of finite-sample performance is reported in Section 8. Section 9 offers some brief final comments.

2. FRACTIONAL COINTEGRATION

We assume initially an observable p-dimensional column vector z_t, $t = 0, \pm 1, ...$, of jointly stationary series that all have the same, unknown, integration order $\delta \in (0, \frac{1}{2})$. By this we mean, denoting by $f_z(\lambda)$, $\lambda \in (-\pi, \pi]$, the spectral density matrix of z_t,

that

$$f_z(\lambda) \sim G_0 \lambda^{-2\delta}, \quad \text{as } \lambda \to 0+, \tag{2.1}$$

where "$\sim$" means throughout that the ratio of real parts, and of imaginary parts, of corresponding elements on the left- and right-hand side, tends to 1, and the $p \times p$ real matrix G_0 is positive semi-definite with positive diagonal elements.

We say that z_t is cointegrated with rank $r < p$ if there exist r linearly independent $p \times 1$ vectors β_i such that for $i = 1, ..., r$, $u_{it} = \beta_i' z_t$ has spectral density $f_{u_i}(\lambda)$ satisfying

$$f_{u_i}(\lambda) \sim \omega_{ii} \lambda^{-2\gamma_i}, \quad \text{as } \lambda \to 0+, \tag{2.2}$$

where

$$0 \leq \gamma_1, \gamma_2, ..., \gamma_r < \delta; \ \omega_{ii} > 0, \quad i = 1, ..., r. \tag{2.3}$$

For the convenience of a simple notation for all relevant integration orders, we also introduce

$$\gamma_i = \delta, \quad i = r+1, ..., p. \tag{2.4}$$

We can embed the cointegration within a non-singular system of degree p, writing

$$B z_t = u_t, \tag{2.5}$$

where $u_t = (u_{1t}, ..., u_{pt})'$ with $u_{it} = z_{it}$ for $i > r$, and

$$B = \begin{bmatrix} B_1 & B_2 \\ 0 & I_{p-r} \end{bmatrix} \tag{2.6}$$

with 0 a vector of zeroes, defining I_k to be the $k \times k$ identity matrix, and with B_1, B_2 being $r \times r$ and $r \times (p-r)$ matrices respectively, such that β_i' is the i-th row of (B_1, B_2), and it is assumed that elements of z_t have been arranged such that B_1 is non-singular. The β_i are not uniquely defined, but we are not concerned with their estimation, or with estimation of the γ_i, while δ will be estimated purely for the

purpose of testing for the existence of cointegration or for the value of r. It is possible to justify (2.5) from an additive representation of the z_{it} with unobserved components having different integration orders.

The elements of u_t can in general be cross-correlated and coherent at all frequencies. We assume the spectral density matrix $f_u(\lambda)$ of u_t satisfies

$$f_u(\lambda) \sim \Phi(\lambda e^{i\theta}; \gamma)^{-1} \Omega \Phi(\lambda e^{-i\theta}; \gamma)^{-1}, \quad \text{as } \lambda \to 0+, \tag{2.7}$$

where Φ is a $p \times p$ diagonal matrix such that for a scalar a and $p \times 1$ vector $b = (b_1, ..., b_p)'$,

$$\Phi(a; b) = diag \left\{ a^{b_1}, ..., a^{b_p} \right\}, \tag{2.8}$$

while θ is real, γ has ith-element γ_i, and Ω is a $p \times p$ real positive definite matrix.

The parameter θ introduces the possibility of phase shift. The property (2.7) occurs when, for example, u_t is generated from an underlying $p \times 1$ vector of jointly dependent $I(0)$ processes e_t by

$$u_t = diag \left\{ \Delta^{-\gamma_1}, ..., \Delta^{-\gamma_p} \right\} e_t, \tag{2.9}$$

where Δ is the first difference operator. In that case, denoting by $f_e(\lambda)$ the spectral density matrix of e_t, we have precisely

$$f_u(\lambda) = \Phi(1 - e^{i\lambda}; \gamma)^{-1} f_e(\lambda) \Phi(1 - e^{-i\lambda}; \gamma)^{-1}. \tag{2.10}$$

From

$$1 - e^{i\lambda} \sim \lambda e^{-i\pi/2} \quad \text{as} \quad \lambda \to 0+, \tag{2.11}$$

we deduce (2.7) with $\Omega = f_e(0)$. In this case $\theta = -\pi/2$. This is assumed by Robinson and Yajima (2002), for example. Another possibility is $\theta = 0$, in which case quadrature spectra of u_t are zero at zero frequency. The value of θ has no effect on diagonal elements of $f_u(\lambda)$ (power spectra) but it does affect off-diagonal elements (cross-spectra) when integration orders vary. We adopt an approach that is valid

for all θ and also does not require estimating θ. It follows from (2.7) that if z_t is cointegrated of rank r then (2.1) holds with G_0 of rank $p-r$, in particular

$$G_0 = \begin{bmatrix} B_1^{-1}B_2 \\ -I_{p-r} \end{bmatrix} \Omega_{22} \begin{bmatrix} B_1^{-1}B_2 \\ -I_{p-r} \end{bmatrix}', \tag{2.12}$$

in which Ω_{22} is the lower right $(p-r)\times(p-r)$ sub-matrix of Ω, which has full rank; the error in the approximation in (2.1) is $O\left(\lambda^{-\max_i \gamma_i - \delta}\right)$. The more detailed structure indicated by (2.7) will be relevant to power considerations.

3. DIAGNOSTIC STATISTICS FOR TESTING NON-COINTEGRATION

Our approach to testing starts from the specification test of Hausman (1978): a parameter estimate that is relatively efficient under the null hypothesis of correct specification, but inconsistent under the alternative of incorrect specification, is compared to one that is relatively inefficient under the null, but consistent in both circumstances. The main parameter of interest that arises under both the null hypothesis of no cointegration, and the alternative, is the integration order δ of the observables, see (2.1). We thus seek suitable estimates of δ.

Local Whittle estimation provides a common approach under both hypotheses (see e.g. Künsch, 1987, Robinson, 1995, Lobato, 1999, Shimotsu, 2007). For a general vector sequence $v_1, ..., v_n$ define the discrete Fourier transform

$$w_v(\lambda) = \frac{1}{(2\pi n)^{\frac{1}{2}}} \sum_{t=1}^{n} v_t e^{it\lambda} \tag{3.1}$$

and the periodogram matrix

$$I_v(\lambda) = w_v(\lambda) w_v'(-\lambda). \tag{3.2}$$

Given observations $z_1, ..., z_n$ we can estimate δ and G_0 in (2.1) by

$$(\hat{\delta}, \hat{G}) = \arg\min_{d,G} \sum_{j=1}^{m} \left[\log\det\left\{ G\lambda_j^{-2d} \right\} + tr\{ I_z(\lambda_j) G^{-1} \lambda_j^{2d} \} \right], \tag{3.3}$$

where $\lambda_j = 2\pi j/n$, m is a bandwidth number satisfying

$$p < m < \frac{n}{2} \tag{3.4}$$

and increasing slowly with n, the minimization with respect to G is over the space of positive definite matrices, and the minimization with respect to d is over a closed sub-interval U of $(0, \frac{1}{2})$. (The latter choice reflects a supposition that $\delta > 0$, as is relevant in the present setting, though the method of estimating δ to be described is asymptoticaly valid also for $\delta \in (-\frac{1}{2}, 0]$, when U is chosen to include it.) Then $\hat{\delta}$ is readily seen to satisfy

$$\hat{\delta} = \arg\min_{d \in U} S(d), \tag{3.5}$$

where

$$S(d) = \log\det\left\{\hat{G}(d)\right\} - \frac{2pd}{m}\sum_{j=1}^{m}\log\lambda_j, \tag{3.6}$$

$$\hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\left\{I_z(\lambda_j)\right\}\lambda_j^{2d}. \tag{3.7}$$

The real part operator is needed in (3.7) because $I_z(\lambda_j)$ can have complex-valued off-diagonal elements. It is justified by the fact that the summands in (3.3) are automatically all real-valued, and so the real part operator could have been initially, if redundantly, applied there to $I_z(\lambda_j)$. We take $\hat{\delta}$ to be our "efficient" estimate.

An "inefficient" estimate of δ is defined in terms of scalar local Whittle estimates, based on the observations on individual z_{it}. Denoting by $I_{z_i}(\lambda)$ the i-th diagonal element of $I_z(\lambda)$, we introduce

$$\tilde{\delta}_{(i)} = \arg\min_{d \in U}\left\{\log\hat{g}_{ii}(d) - \frac{2d}{m}\sum_{j=1}^{m}\log\lambda_j\right\}, \tag{3.8}$$

where

$$\hat{g}_{ii}(d) = \frac{1}{m}\sum_{j=1}^{m} I_{z_i}(\lambda_j)\lambda_j^{2d}. \tag{3.9}$$

Then our "inefficient" estimate of δ is

$$\tilde{\delta} = \sum_{i=1}^{p} a_i \tilde{\delta}_{(i)}, \tag{3.10}$$

where the a_i are arbitrarily chosen weights satisfying

$$\sum_{i=1}^{p} a_i = 1. \tag{3.11}$$

For example, we might take

$$a_i \equiv 1/p, \tag{3.12}$$

so the arithmetic mean of the $\tilde{\delta}_{(i)}$ is used, or

$$a_j = 1, \quad a_i = 0, \quad i \neq j, \text{ some } j, \tag{3.13}$$

so that $\tilde{\delta} = \tilde{\delta}_{(j)}$. In the latter case only $\tilde{\delta}_{(j)}$ need be computed, but in practice all the $\tilde{\delta}_{(i)}$ are useful in pre-testing the hypothesis of a common memory parameter in z_t (see Robinson and Yajima, 2002).

From asymptotic theory of Robinson (1995), Lobato (1999) we expect that under the null hypothesis of non-cointegration $m^{\frac{1}{2}}(\hat{\delta} - \tilde{\delta})$ will converge in distribution under suitable conditions to a normal variate with zero mean. We stress a computationally simpler approach that will also lend itself to robustification. The implicitly-defined estimate $\hat{\delta}$ can be approximated by a single Newton step based on (3.6) and starting from the root-m-consistent estimate $\tilde{\delta}$, in the sense that the limit distribution of $m^{\frac{1}{2}}(\hat{\delta} - \delta)$ (and thence of $m^{\frac{1}{2}}(\hat{\delta} - \tilde{\delta})$) is achieved. In particular, the approximate estimate can be defined as $\bar{\delta} = \tilde{\delta} + \left\{(\partial^2/\partial d^2)\, S(\widetilde{\delta})\right\}^{-1} s(\tilde{\delta})$, where

$$s(\tilde{\delta}) = \frac{1}{2}\frac{\partial S(\tilde{\delta})}{\partial d} = tr\left\{\hat{G}(\tilde{\delta})^{-1}\hat{H}(\tilde{\delta})\right\} \tag{3.14}$$

with

$$\hat{H}(d) = \frac{1}{m}\sum_{j=1}^{m} \nu_j \,\mathrm{Re}\left\{I_z(\lambda_j)\right\} \lambda_j^{2d}, \tag{3.15}$$

$$\nu_j = \log j - \frac{1}{m}\sum_{i=1}^{m} \log i. \tag{3.16}$$

We thus use a scaled $s(\tilde{\delta})$ as test statistic. We also introduce a modified version: define

$$\hat{G}^*(d) = \frac{1}{m}\sum_{j=1}^{m} I_z(\lambda_j)\lambda_j^{2d}, \quad \hat{H}^*(d) = \frac{1}{m}\sum_{j=1}^{m} \nu_j I_z(\lambda_j)\lambda_j^{2d}, \tag{3.17}$$

and consider

$$s^*(\tilde{\delta}) = tr\left\{\hat{G}^*(\tilde{\delta})^{-1}\hat{H}^*(\tilde{\delta})\right\}. \tag{3.18}$$

Though both $\hat{G}^*(d)$ and $\hat{H}^*(d)$ have complex-valued off-diagonal elements, $s^*(\tilde{\delta})$ is always real-valued since both are Hermitian. Though $s(\tilde{\delta})$ and $s^*(\tilde{\delta})$ are not numerically equivalent, they turn out to have the same null limit distribution, while $s^*(\tilde{\delta})$ has advantages discussed below.

We introduce

$$\begin{aligned} X &= ms(\tilde{\delta})^2 \Big/ \left\{p^2 tr(\hat{R}A\hat{R}A) - p\right\}, & (3.19)\\ X^* &= ms^*(\tilde{\delta})^2 \Big/ \left\{p^2 tr(\hat{R}^*A\hat{R}^*A) - p\right\}, & (3.20) \end{aligned}$$

where

$$\begin{aligned} \hat{R} &= \hat{D}^{-\frac{1}{2}}\hat{G}(\tilde{\delta})\hat{D}^{-\frac{1}{2}}, \quad \hat{R}^* = \hat{D}^{-\frac{1}{2}}\hat{G}^*(\tilde{\delta})\hat{D}^{-\frac{1}{2}},\\ \hat{D} &= diag\left\{\hat{g}_{11}, ..., \hat{g}_{pp}\right\}, \; A = diag\left\{a_1, ..., a_p\right\}, \end{aligned}$$

where $\hat{g}_{ii}$ is the i-th diagonal element of $\hat{G}(\tilde{\delta})$ (and of $\hat{G}^*(\tilde{\delta})$). Though $\hat{R}^*$ has complex-valued off-diagonal elements it is Hermitian, and thus $tr\left\{\hat{R}^*A\hat{R}^*A\right\}$ is real. Moreover, by the same property, writing $A^{\frac{1}{2}}\hat{R}^*A^{\frac{1}{2}} = U + iV$ for real matrices U and V, we have $U = U'$ and $V = -V'$, and then

$$\begin{aligned} tr\left\{\hat{R}^*A\hat{R}^*A\right\} &= tr\left\{(U+iV)(U-iV)'\right\}\\ &= tr(U^2) + tr(VV') + i\left\{tr(VU) - tr(UV')\right\}\\ &\geq tr(U^2) = tr\left\{\hat{R}A\hat{R}A\right\}, \end{aligned} \tag{3.21}$$

since VV' is positive semi-definite and $tr(VU) = tr(U'V') = tr(UV')$. Moreover, since $\hat{R}$ has unit diagonal elements

$$tr(U^2) \geq tr\{A^2\} = \sum_{i=1}^{p} a_i^2. \tag{3.22}$$

It follows from the Cauchy inequality and (3.1) that the denominators in (3.9) and (3.20) are guaranteed non-negative and ordered:

$$p^2 tr\left\{\hat{R}^* A \hat{R}^* A\right\} - p \geq p^2 tr\left\{\hat{R} A \hat{R} A\right\} - p \geq p^2 \sum_{i=1}^{p} a_i^2 - p \geq 0. \tag{3.23}$$

Under (3.12) the denominators reduce to $tr(\hat{R}^2) - p$ and $tr(\hat{R}^{*2}) - p$ respectively, where $tr(\hat{R}^2) \leq tr(\hat{R}^{*2}) \leq p^2$, and under (3.13) both are $p(p-1)$, which is data-free.

4. ASYMPTOTIC NULL DISTRIBUTION AND CONSISTENCY

We now establish the asymptotic distributions of X and X^* under the non-cointegrated null hypothesis, introducing regularity conditions which are similar to ones of Robinson (1995), Lobato (1999), Shimotsu (2007). These conditions are capable of some modification and extension but their use allows us to apply some basic results. We assume that z_t has representation

$$z_t = Ez_t + \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad \sum_{j=0}^{\infty} \|C_j\|^2 < \infty, \tag{4.1}$$

where the C_j are $p \times p$ matrices, $\|.\|$ denotes Euclidean norm, and the matrix function

$$C(\lambda) = \sum_{j=0}^{\infty} C_j e^{-ij\lambda} \tag{4.2}$$

is differentiable in a neighbourhood of $\lambda = 0$ and satisfies the conditions

$$\lambda^{\delta} C(\lambda) = Q' + O(\lambda^{\beta}), \quad \text{as } \lambda \to 0+, \quad \text{some } \beta \in (0, 2], \tag{4.3}$$

$$\frac{dC(\lambda)}{d\lambda} = O(\lambda^{-\delta-1}), \quad \text{as } \lambda \to 0+, \tag{4.4}$$

where Q is a $p \times p$ full rank matrix such that $Q'Q = G_0$, "O" applies here to each element of a matrix, and the ε_t are $p \times 1$ vectors satisfying

$$E\left(\varepsilon_t \left| \mathcal{F}_{t-1} \right.\right) = 0, \quad E\left(\varepsilon_t \varepsilon_t' \left| \mathcal{F}_{t-1} \right.\right) = I_p, \quad a.s., \tag{4.5}$$

and all third and fourth conditional (on $\mathcal{F}_{t-1}$) moments and cross-moments of elements of ε_t are a.s. constant, where $\mathcal{F}_t$ is the σ-field of events generated by ε_s, $s \leq t$. These conditions imply that z_t is a non-cointegrated $I(\delta)$ vector, satisfying (2.1) with positive definite G_0. We assume that δ is an interior point of U. We assume that

$$\frac{(\log n)^2 m^{1+2\beta}}{n^{2\beta}} + \frac{(\log n)^8}{m} \to 0, \quad \text{as } n \to \infty. \tag{4.6}$$

Theorem 1 *Under the assumptions in the previous paragraph,*

$$X, X^* \to_d \chi^2_1, \quad as\ n \to \infty. \tag{4.7}$$

Proof. The mean value theorem gives

$$s^*(\tilde{\delta}) = s^*(\delta) + \frac{ds^*(\bar{\delta})}{dx}(\tilde{\delta} - \delta), \tag{4.8}$$

where $\left|\bar{\delta} - \delta\right| \leq \left|\tilde{\delta} - \delta\right|$. Denote by $\hat{G}_R$ ($\hat{G}_I$) and $\hat{H}_R$ ($\hat{H}_I$) the real (imaginary) parts of $\hat{G}^*(\delta)$ and $\hat{H}^*(\delta)$. We have

$$\hat{G}^*(\delta)^{-1} = \left(I_p - i\hat{G}_R^{-1}\hat{G}_I\right)\left(\hat{G}_R + \hat{G}_I\hat{G}_R^{-1}\hat{G}_I\right)^{-1}. \tag{4.9}$$

Then since $\hat{G}^*(\delta)$ and $\hat{H}^*(\delta)$ are Hermitian it follows that

$$s^*(\delta) = tr\left\{\left(\hat{G}_R + \hat{G}_I\hat{G}_R^{-1}\hat{G}_I\right)^{-1}\left(\hat{H}_R + \hat{H}_I\hat{G}_R^{-1}\hat{G}_I\right)\right\}, \tag{4.10}$$

imaginary parts cancelling. It follows from arguments routinely extending those used to establish (4.8) of Robinson (1995) (see also Lobato (1999), Appendix C) that

$$\hat{G}^*(\delta) = m^{-1}\sum_{j=1}^{m} Q' I_\varepsilon(\lambda_j) Q + o_p(m^{-\frac{1}{2}}), \tag{4.11}$$

$$\hat{H}^*(\delta) = m^{-1}\sum_{j=1}^{m} \nu_j Q' I_\varepsilon(\lambda_j) Q + o_p(m^{-\frac{1}{2}}). \tag{4.12}$$

The imaginary parts of the leading terms on the right of (4.11) and (4.12) are easily seen to be $O_p(m^{-\frac{1}{2}})$, so that $\hat{G}_I = O_p(m^{-\frac{1}{2}})$, $\hat{H}_I = O_p(m^{-\frac{1}{2}})$, and thus

$$s^*(\delta) = s(\delta) + o_p(m^{-\frac{1}{2}}). \tag{4.13}$$

Applying (4.11), (4.12) again we have $\hat{G}_R = G_0 + o_p(m^{-\frac{1}{2}})$ and then

$$\begin{aligned} m^{\frac{1}{2}} s(\delta) &= m^{-\frac{1}{2}} tr\left[G_0^{-1} \sum_{j=1}^{m} Q' \nu_j \operatorname{Re}\left\{ I_\varepsilon(\lambda_j) \right\} Q \right] + o_p(1) \\ &= m^{-\frac{1}{2}} tr\left\{ \sum_{j=1}^{m} \nu_j I_\varepsilon(\lambda_j) \right\} + o_p(1). \end{aligned} \tag{4.14}$$

In addition, denoting by q_i the i-th column of Q, it is straightforward to show that

$$\begin{aligned} \tilde{\delta}_{(i)} - \delta &= -\frac{1}{2} \sum_{j=1}^{m} \nu_j I_{z_i}(\lambda_j) \lambda_j^{2\delta} / \sum_{j=1}^{m} I_{z_i}(\lambda_j) \lambda_j^{2\delta} + o_p(m^{-\frac{1}{2}}) \\ &= -\frac{1}{2m} \sum_{j=1}^{m} \nu_j \frac{q_i' I_\varepsilon(\lambda_j) q_i}{q_i' q_i} + o_p(m^{-\frac{1}{2}}), \quad i = 1, ..., p, \end{aligned} \tag{4.15}$$

proceeding much as in Robinson (1995), the only difference being that in the present case the scalar sequences z_{it} each depend on the vector white noise sequence ε_t. Also, for scalar argument x,

$$\frac{ds(x)}{dx} = tr\left\{ \hat{G}(x)^{-1} \left(\frac{d\hat{H}(x)}{dx} - \frac{d\hat{G}(x)}{dx} \hat{G}(x)^{-1} \hat{H}(x) \right) \right\}. \tag{4.16}$$

Now (4.15) implies that $\bar{\delta} = \delta + O_p(m^{-\frac{1}{2}})$, whence arguments like those in Robinson (1995) give

$$\hat{G}(\bar{\delta}) \rightarrow_p G_0, \quad \hat{H}(\bar{\delta}) \rightarrow_p 0,$$

and also

$$\frac{d\hat{H}(\bar{\delta})}{dx} = \frac{2}{m} \sum_{j=1}^{m} \nu_j (\log \lambda_j) \operatorname{Re}\left\{ I_z(\lambda_j) \right\} \lambda_j^{2\bar{\delta}} = 2G_0 + o_p(1). \tag{4.17}$$

It follows that

$$\frac{ds(\bar{\delta})}{dx} \rightarrow_p 2p. \tag{4.18}$$

Similarly

$$\frac{ds^*(\bar{\delta})}{dx} \rightarrow_p 2p. \tag{4.19}$$

Thus, using (4.14), (4.15), (4.18) and (4.19), it is straightforwardly shown that both $m^{\frac{1}{2}}s(\tilde{\delta})$ and $m^{\frac{1}{2}}s^*(\tilde{\delta})$ differ by $o_p(1)$ from

$$m^{-\frac{1}{2}}tr\left\{\sum_{j=1}^{m}\nu_j I_\varepsilon(\lambda_j)\left(I_p - pQAD^{-1}Q'\right)\right\}, \tag{4.20}$$

where $D = diag\{g_{11}, ..., g_{pp}\}$, with g_{ij} the (i,j)-th element of G_0. Then after a martingale approximation, as in Robinson (1995), we deduce that (4.20) converges in distribution to a $N\left(0, tr\left\{\left(I_p - pQAD^{-1}Q'\right)^2\right\}\right)$ variate, where (since diagonal matrices commute) the variance equals

$$tr\left\{I_p - 2pA + p^2RARA\right\} = p\left(ptr\left\{RARA\right\} - 1\right), \tag{4.21}$$

where $R = D^{-\frac{1}{2}}G_0D^{-\frac{1}{2}}$. The proof is completed by noting that $\hat{G}(\tilde{\delta}) \rightarrow_p G_0$, $\hat{G}^*(\tilde{\delta}) \rightarrow_p G_0$ imply $\hat{R} \rightarrow_p R$, $\hat{R}^* \rightarrow_p R$. ■

We now consider the consistency of the test implied by Theorem 1 for X^*, against fixed cointegrated alternatives. We adopt for convenience the same assumptions as for Theorem 1 above, with the exception that z_t is given by (2.5) with u_t having a representation of form (4.1), and (2.7) holds with (2.3), (2.4), (4.3), (4.5). Define also $\omega = diag\left\{\omega_{11}, ..., \omega_{pp}\right\}$, where ω_{ij} is the (i,j)th element of Ω, and write $\Upsilon = \omega^{-\frac{1}{2}}\Omega\omega^{-\frac{1}{2}}$. Also define the $p \times p$ matrices E and F with (i,j)-th elements given in

$$E = \left(\frac{1}{2\delta - \gamma_i - \gamma_j + 1}\right), \quad F = \left(\frac{2\delta - \gamma_i - \gamma_j}{(2\delta - \gamma_i - \gamma_j + 1)^2}\right), \tag{4.22}$$

and write $e = diag\left\{(2\delta - 2\gamma_1 + 1)^{-1}, ..., (2\delta - 2\gamma_p + 1)^{-1}\right\}$. Denote by $\circ$ the Hadamard product operator.

Theorem 2 *Under the assumptions in the previous paragraph, as* $n \to \infty$

$$m^{-1}X^* \to_p \frac{\left[tr\left\{(E \circ \Upsilon)^{-1}(F \circ \Upsilon)\right\}\right]^2}{p^2 tr\left\{\left[e^{-\frac{1}{2}}(E \circ \Upsilon)e^{-\frac{1}{2}}A\right]^2\right\} - p}. \tag{4.23}$$

Proof. From (2.5),

$$s^*(\delta) = tr\left\{\left(\sum_{j=1}^m \lambda_j^{2\delta} I_u(\lambda_j)\right)^{-1} \sum_{j=1}^m \nu_j \lambda_j^{2\delta} I_u(\lambda_j)\right\}. \tag{4.24}$$

By an approximation analogous to that used in (2.8) of Robinson (1994a)

$$\Phi(\lambda_m;\gamma)\frac{1}{m}\sum_{j=1}^m \nu_j^s \left(\frac{\delta}{m}\right)^{2\delta}\left\{I_u(\lambda_j) - \Phi\left(\lambda_j e^{i\theta};\gamma\right)^{-1}\Omega\Phi\left(\lambda_j e^{-i\theta};\gamma\right)^{-1}\right\}\Phi(\lambda_m;\gamma) = o_p(1), \tag{4.25}$$

for $s = 0, 1$. Then because $\Phi(.;.)$ is diagonal and $\lambda_j e^{i\theta} = j \times (2\pi/n) \times e^{i\theta}$, we have also $\Phi\left(\lambda_j e^{i\theta};\gamma\right) = \Phi(j;\gamma)\Phi\left(\frac{2\pi}{n};\gamma\right)\Phi\left(e^{i\theta};\gamma\right)$, where all factors are non-singular, and thence there is cancellation of the last two factors, to give

$$\begin{aligned} s^*(\delta) &= tr\left\{\left(\sum_{j=1}^m (\frac{j}{m})^{2\delta}\Phi(\frac{j}{m};\gamma)^{-1}\Upsilon\Phi(\frac{j}{m};\gamma)^{-1}\right)^{-1}\sum_{j=1}^m \nu_j\left(\frac{j}{m}\right)^{2\delta}\Phi(\frac{j}{m};\gamma)^{-1}\Upsilon\Phi(\frac{j}{m};\gamma)^{-1}\right\} \\ &\quad + o_p(1). \end{aligned} \tag{4.26}$$

Using the approximations

$$m^{-1-a}\sum_{j=1}^m j^a \sim \frac{1}{a+1}, \quad m^{-1-a}\sum_{j=1}^m \nu_j j^a \sim \frac{a}{(a+1)^2}, \tag{4.27}$$

valid for all $a > -1$, we deduce that as $m \to \infty$

$$m^{-1}\sum_{j=1}^m \left(\frac{j}{m}\right)^{2\delta}\Phi\left(\frac{j}{m};\gamma\right)^{-1}\Upsilon\Phi\left(\frac{j}{m};\gamma\right)^{-1} \to E \circ \Upsilon, \tag{4.28}$$

$$m^{-1}\sum_{j=1}^m \nu_j\left(\frac{j}{m}\right)^{2\delta}\Phi\left(\frac{j}{m};\gamma\right)^{-1}\Upsilon\Phi\left(\frac{j}{m};\gamma\right)^{-1} \to F \circ \Upsilon, \tag{4.29}$$

from which it follows that

$$s^*(\delta) \to_p tr\left\{(E \circ \Upsilon)^{-1}(F \circ \Upsilon)\right\}. \tag{4.30}$$

Now for $s = 0$,

$$\begin{aligned} &\left\| \Phi(\lambda_m;\gamma) \frac{1}{m} \sum_{j=1}^{m} \nu_j^s \left\{ \left(\frac{j}{m}\right)^{2\tilde{\delta}} - \left(\frac{j}{m}\right)^{2\delta} \right\} I_u(\lambda_j) \Phi(\lambda_m;\gamma) \right\| \\ \leq \;& 8 \left|\tilde{\delta} - \delta\right| (\log m)^2 tr \left\{ \Phi(\lambda_m;\gamma) \frac{1}{m} \sum_{j=1}^{m} \left(\frac{j}{m}\right)^{2\delta} I_u(\lambda_j) \Phi(\lambda_m;\gamma) \right\}, \end{aligned} \tag{4.31}$$

using an inequality like that near the bottom of p.133 of Robinson (1994a). Now since $\hat{\delta}$ is $m^{\frac{1}{2}}$-consistent for δ, and the trace is $O_p(1)$ from the above arguments, it follows that (4.31) $= o_p(1)$. Routine arguments then give $s^*(\tilde{\delta}) - s(\delta) \to_p 0$. The arguments above imply that $\hat{R}^* \to_p e^{-\frac{1}{2}}(E \circ \Upsilon)e^{-\frac{1}{2}}$, to complete the proof. ■

Since the denominator of (4.23) is always finite, the test is consistent when the numerator is non-zero. Whether this is the case appears in general to depend on Υ, as well as γ. Take $p = 2$ for example. Write $\zeta = \delta - \gamma_1$ and ρ for the off-diagonal element of Υ. Then

$$tr\left\{(E \circ \Upsilon)^{-1}(F \circ \Upsilon)\right\} = 2\zeta \frac{(2\zeta+1)^{-2} - \rho^2(\zeta+1)^{-3}}{(2\zeta+1)^{-1} - \rho^2(\zeta+1)^{-2}}. \tag{4.32}$$

The denominator of (4.32) is finite (and nonzero for $|\rho| \leq 1$ and $\zeta \in (0, \frac{1}{2})$), and the numerator can be zero only when $\rho^2 = (\zeta+1)^3/(2\zeta+1)^2$. The right side of this is decreasing in ζ and thus we can say, for example, that whenever $\rho^2 \leq 27/32$, that is, $|\rho| \leq 0.918$, (4.32) is non-zero for any $\zeta \in (0, \frac{1}{2})$. Of course power will be poor when (4.32) is close to zero.

The right hand side of (4.23) is desirably free of θ, and also of B. It would be possible to extend Theorem 2 to allow variation in integration orders and elements of z_t. In the simplest case, the upper-triangular B^{-1} is also block-diagonal, with blocks

corresponding to the varying z_t integration orders. This is the situation studied by Robinson and Yajima (2002), where z_t was partitioned into subsets with common integration orders, and cointegration studied only within subsets. We again achieve a limit of $m^{-1}X^*$ which is also free of B and (possibly varying) phase parameters, yet depends on all the integration orders of z_t, u_t.

5. A ROBUSTIFIED STATISTIC

There is always a difficulty drawing conclusions from a test rejection, because tests can have power against unanticipated departures from the null hypothesis. We would like to take rejection as evidence of cointegration. The statistics X, X^* take for granted that all elements of z_t have the same integration order, δ. Suppose, however, that z_{it} is an $I(\delta_i)$ process, $0 \leq \delta_i < \frac{1}{2}$, $i = 1, ..., p$, where not all the δ_i are equal, so that

$$f_z(\lambda) \sim \Phi\left(\lambda e^{i\theta}; \xi\right)^{-1} \Omega \Phi\left(\lambda e^{-i\theta}; \xi\right)^{-1}, \tag{5.1}$$

where $\xi = (\delta_1, ..., \delta_p)'$. Then z_t is not cointegrated, but under otherwise similar conditions to those of Theorem 2, we deduce that

$$m^{-1}X^* \rightarrow_p \frac{\left[tr\left\{\left(E^\dagger \circ \Upsilon\right)^{-1}\left(F^\dagger \circ \Upsilon\right)\right\}\right]^2}{p^2 tr\left\{\left[e^{\dagger - \frac{1}{2}}\left(E^\dagger \circ \Upsilon\right)e^{\dagger - \frac{1}{2}}A\right]^2\right\} - p}, \tag{5.2}$$

where

$$E^\dagger = \left(\frac{1}{2\delta - \delta_i - \delta_j + 1}\right), \quad F^\dagger = \left(\frac{2\delta - \delta_i - \delta_j}{(2\delta - \delta_i - \delta_j + 1)^2}\right), \tag{5.3}$$

with now $\delta = \sum_{i=1}^p a_i \delta_i$, denoting the probability limit of $\tilde{\delta}$, and $e^\dagger = diag\{(2\delta - 2\delta_1 + 1)^{-1}, ..., (2\delta - 2\delta_p + 1)^{-1}\}$. If not all δ_i are equal the right side of (5.2) can be non-zero. By some alternative definitions (see below) it is possible that cointegration can exist without all δ_i being equal, so long as at least 2 are. But equally it may

not exist in these circumstances, and if $p = 2$ it cannot exist unless $\delta_1 = \delta_2$. It is anticipated that a pre-test of equality of the δ_i would be carried out; tests that do not presume the existence or non-existence of cointegration were introduced by Robinson and Yajima (2002), Hualde (2004). But non-rejection of equality may not be sufficiently reassuring. Thus we propose a test that is robust to inequalities in the δ_i.

Define

$$\hat{G}^{**}(x) = \frac{1}{m}\sum_{j=1}^{m} \Phi(\lambda_j; x) I_z(\lambda_j) \Phi(\lambda_j; x), \quad \hat{H}^{**}(x) = \frac{1}{m}\sum_{j=1}^{m} \nu_j \Phi(\lambda_j; x) I_z(\lambda_j) \Phi(\lambda_j; x) \tag{5.4}$$

for $x = (x_1, ..., x_p)'$. Denote $\tilde{\xi} = \left(\tilde{\delta}_{(1)}, ..., \tilde{\delta}_{(p)}\right)'$, with the $\tilde{\delta}_{(i)}$ defined as in Section 3. Finally, denote by $\hat{g}_{ii}^{**}$ the i-th diagonal element of $\hat{G}^{**}(\tilde{\xi})$, and introduce

$$s^{**}(x) = tr\left\{\hat{G}^{**}(x)^{-1}\hat{H}^{**}(x)\right\}, \tag{5.5}$$

$$\hat{D}^{**} = diag\left\{\hat{g}_{11}^{**}, ..., \hat{g}_{pp}^{**}\right\}, \quad \hat{R}^{**} = \hat{D}^{**-\frac{1}{2}}\hat{G}^{**}(\tilde{\xi})\hat{D}^{**-\frac{1}{2}}, \tag{5.6}$$

$$X^{**} = ms^{**}(\tilde{\xi})^2 \Big/ \left(tr\left\{\hat{R}^{**2}\right\} - p\right). \tag{5.7}$$

We impose the same conditions as for Theorem 1, except that (4.3) and (4.4) are replaced respectively by

$$\Phi\left(\lambda e^{i\theta}; \xi\right) C(\lambda) = Q' + O(\lambda^{\beta}), \quad \text{as } \lambda \to 0+, \quad \text{some } \beta \in (0, 2], \tag{5.8}$$

$$\Phi\left(\lambda e^{i\theta}; \xi\right) \frac{d}{d\lambda} C(\lambda) = O(\lambda^{-1}), \quad \text{as } \lambda \to 0+. \tag{5.9}$$

Theorem 3 *Under the assumptions in the previous paragraph,*

$$X^{**} \to_d \chi_1^2, \quad as\ n \to \infty. \tag{5.10}$$

Proof. We have

$$s^{**}(\tilde{\xi}) = s^{**}(\xi) + \frac{\partial s^{**}(\bar{\xi})'}{\partial x}(\tilde{\xi} - \xi), \tag{5.11}$$

where $\left\|\bar{\xi}-\xi\right\| \leq \left\|\tilde{\xi}-\xi\right\|$. Much as in the proof of Theorem 1 (cf (4.11), (4.12)), we have

$$\hat{G}^{**}(\xi) = \Phi\left(e^{i\theta};\xi\right) m^{-1} \sum_{j=1}^{m} Q' I_{\varepsilon}(\lambda_j) Q \Phi(e^{-i\theta};\xi) + o_p(m^{-1/2}), \tag{5.12}$$

$$\hat{H}^{**}(\xi) = \Phi\left(e^{i\theta};\xi\right) m^{-1} \sum_{j=1}^{m} \nu_j Q' I_{\varepsilon}(\lambda_j) Q \Phi(e^{-i\theta};\xi) + o_p(m^{-1/2}). \tag{5.13}$$

The factors $\Phi\left(e^{i\theta};\xi\right)$ cancel on inserting these approximations in $s^{**}(\xi)$, and we get

$$m^{\frac{1}{2}} s^{**}(\xi) = m^{-\frac{1}{2}} tr\left\{\sum_{j=1}^{m} \nu_j I_{\varepsilon}(\lambda_j)\right\} + o_p(1), \tag{5.14}$$

c.f. (4.14). Also, proceeding similarly to the proof of Theorem 1,

$$\frac{\partial s^{**}(\bar{\xi})}{\partial x'} \rightarrow_p (2, ..., 2). \tag{5.15}$$

Then applying (4.15) leads to

$$m^{\frac{1}{2}} s^{**}(\tilde{\xi}) = m^{-\frac{1}{2}} tr\left\{\sum_{j=1}^{m} \nu_j I_{\varepsilon}(\lambda_j)\left(I_p - QD^{-1}Q'\right)\right\}. \tag{5.16}$$

But the right side is just a special case of (4.20), so it converges to a $N\left(0, tr\left\{\left(I_p - QD^{-1}Q'\right)^2\right\}\right)$ variate. The limiting variance equals $tr(R^2) - p$, and from previous arguments $\hat{R}^{**} \rightarrow_p R$, to complete the proof. ■

While our statistics all have the same null limit distribution, their powers can differ, and there follows a derivation of local power properties. Consider the $p \times 1$ vector process u_t, with i-th element u_{it}, satisfying the same conditions as z_t did in Theorem 1. Introduce the triangular array $p \times 1$ vector process $u_t^{(m)}$, whose i-th element is $\Delta^{c_i m^{-\frac{1}{2}}} u_{it}$ for $i \leq r$, and u_{it} for $i > r$, where $c_i > 0$, $i = 1, ..., r$. Then $z_t^{(m)} = B^{-1} u_t^{(m)}$, with B as in (2.6), exhibits locally, rank r, cointegrated alternatives (of Pitman type, but of order $m^{-\frac{1}{2}}$), from the non-cointegration null. The approach taken here is analogous to one employed in a fractional context by Robinson (1994b), though there the topic was testing for integration order in a parametric setting, with departures

of order $n^{-\frac{1}{2}}$. Partly because of this reference, and to save the space required by a detailed proof, we only briefly sketchderivations. Considering X^*, we begin from (4.8), and then deduce

$$m^{\frac{1}{2}}s^*(\delta) = m^{\frac{1}{2}}tr\left\{\left(\sum_{j=1}^{m} I_{u^{(m)}}(\lambda_j)\lambda_j^{2\delta}\right)^{-1}\sum_{j=1}^{m}\nu_j I_{u^{(m)}}(\lambda_j)\lambda_j^{2\delta}\right\}, \tag{5.17}$$

where $I_{u^{(m)}}(.)$ is the periodogram of the $u_t^{(m)}$. Defining the $p \times 1$ vector process $v_t^{(m)}$ having i-th element $v_{it}^{(m)} = c_i m^{-\frac{1}{2}}(\log\Delta)\Delta^{c_i m^{-\frac{1}{2}}}u_{it}$ for $i \leq r$, and zero for $i > r$, an argument like that in the proof of Theorem 2 of Robinson (1994b) indicates that the effect of replacing the $u_t^{(m)}$ by the $u_t + v_t^{(m)}$ is negligible. Further, the effect of then replacing the $v_{it}^{(m)}$ by the $c_i m^{-\frac{1}{2}}(\log\Delta)u_{it}$ is negligible, and then, by frequency-domain approximation the discrete Fourier transform at frequency λ of the latter quantity can be replaced by that of u_{it} times $c_i m^{-\frac{1}{2}}\log(1-e^{i\lambda})$. Thence we can approximate $m^{\frac{1}{2}}s^*(\delta)$ by

$$m^{-\frac{1}{2}}tr\left\{G^{-1}\sum_{j=1}^{m}\nu_j I_u(\lambda_j)\lambda_j^{2\delta}\right\} + 2\sum_{j=1}^{m}c_i, \tag{5.18}$$

using also $m^{-1}\Sigma_{j=1}^{m}\nu_j\log(1-e^{i\lambda_j}) \sim m^{-1}\Sigma_{j=1}^{m}\nu_j^2 \sim 1$. From the proof of Theorem 1, $m^{\frac{1}{2}}s^*(\widetilde{\delta})$ is asymptotically distributed as $N\left(2\Sigma_{i=1}^{r}c_i, p^2tr(RARA) - p\right)$, and thence we deduce

$$X^* \rightarrow_d \chi_1^{'2}\left(4\left(\sum_{i=1}^{r}c_i\right)^2 / \left\{p^2tr(RARA) - p\right\}\right), \tag{5.19}$$

$\chi_1^{'2}(.)$ indicating a non-central χ_1^2 variate with non-centrality parameter in parentheses.

By a similar derivation to that of (5.19), but using the proof of Theorem 3,

$$X^{**} \rightarrow_d \chi_1^{'2}\left(4\left(\sum_{i=1}^{r}c_i\right)^2 / \left\{tr(R^2) - p\right\}\right). \tag{5.20}$$

Note also that (5.19) and (5.20) reflect an increase in power with increasing r.

6. ESTIMATING COINTEGRATING RANK

It is possible to use our tests in a sequential way in order to estimate cointegrating rank. Let us suppose that the full set of variables we wish to consider are the elements of the $P \times 1$ vector $x_t = (x_{1t}, ..., x_{Pt})'$, for $P \geq 3$. As in the previous section they need not all have the same integration order, so we will apply the X^{**} statistic; under the assumption of identical integration orders X or X^* might be used instead. The cointegrating rank of x_t is denoted R. We introduce new notation to allow z_t and p to refer to a subset of x_t and its dimension in a particular test, so these vary over our sequential procedure.

Consider the null hypothesis

$$H_{0p}(j_1, ..., j_p) : x_{j_1 t}, ..., x_{j_p t} \text{ are not cointegrated,} \tag{6.1}$$

for some integers p, $j_1, ..., j_p$, where these satisfy

$$1 \leq j_1 < j_2 < ... < j_p \leq P, \quad 2 \leq p \leq P. \tag{6.2}$$

Taking $z_t = (x_{j_1 t}, ..., x_{j_p t})'$, we reject $H_{0p}(j_1, ..., j_p)$ if X^{**} is significant at some prescribed level, applying the large sample approximation in Theorem 3.

For given k, $j_1, ..., j_k$, place the $\binom{P}{k}$ hypotheses $H_{0k}(j_1, ..., j_k)$ in (say) lexicographic order. (There is sensitivity to the ordering.) Then form an ordering of all $2^P - P - 1$ hypotheses (6.1) satisfying (6.2) such that for $k < \ell$ the ordered H_{0k} precede the ordered $H_{0\ell}$. The hypotheses are tested in this order, but some can be omitted, as we now describe.

For some k, $j_1, ..., j_k$, if $H_{0k}(j_1, ..., j_k)$ is the i-th hypothesis to be rejected, for some $i \geq 1$, define the set $S_i = \{j_1, ..., j_k\}$. Now suppose that q hypotheses have already been rejected when we consider whether to test $H_{0\ell}(j_1, ..., j_\ell)$, for some ℓ, $j_1, ..., j_\ell$. We do not test it if the set $S = \{j_1, ..., j_\ell\}$ satisfies either $S \subset (S_1 \cup ... \cup S_q)$ or $S_i \subset S$, some $i = 1, ..., q$. The reason is as follows. If $S \subset (S_1 \cup ... \cup S_q)$ then a

linear combination of the q cointegrating relations whose existence has already been "established" is a linear combination of $x_{j_1 t}, ..., x_{j_\ell t}$. If $S_i \subset S$ for some $i = 1, ..., q$ then the i-th cointegrating relation implies $x_{j_1 t}, ..., x_{j_\ell t}$ are also cointegrated (e.g., if necessary we can give zero weights to the $j_1, ..., j_\ell$ that are not in S_i). Thus, in some circumstances there will be "gaps" in the sequence of hypothesis tests, and in some cases termination with no need to consider further hypotheses. The estimate of the cointegrating rank R is the total number of rejections.

In view of the sequential nature of the procedure and the varying possible outcomes there is a difficulty in attaching probability statements to the event that R is correctly determined given the significance levels used in the individual tests. One approach, which applies Bonferroni's inequality, is to assign significance levels to the individual tests so as to approximately yield a desired size, α (e.g. $\alpha = 0.05$), for testing the non-cointegration hypothesis $H_0^* : R = 0$ against $H_1^* : R > 0$. Since we only have the opportunity to not reject H_0^* by carrying out all $2^P - P - 1$ tests, the usual Bonferroni argument and Theorem 3 suggests using a $\alpha/(2^P - P - 1)$ significance level based on the χ_1^2 distribution for each. Ignoring the approximation in null distribution, this actually corresponds to a significance level for testing H_0^* of less than α.

It may be helpful to illustrate how the algorithm operates for small values of P. Take $P = 3$. The ordered hypotheses are $H_{02}(1,2)$, $H_{02}(1,3)$, $H_{02}(2,3)$, $H_{03}(1,2,3)$. If $H_{02}(1,2)$ and $H_{02}(1,3)$ are rejected we estimate R to be 2 and stop. Otherwise we test $H_{02}(2,3)$. We only test $H_{03}(1,2,3)$ if all the H_{02} have not been rejected. "Gaps" are not possible when $P = 3$, but they are when $P = 4$. In this case, suppose $H_{02}(1,2)$ and $H_{03}(1,3)$ are rejected, but $H_{02}(1,4)$ is not rejected. Then we skip $H_{02}(2,3)$, but test $H_{02}(2,4)$, and if that is not rejected, $H_{02}(3,4)$. If either is rejected we stop, estimating R to be 2. If both are not rejected we skip $H_{03}(1,2,3)$, $H_{03}(1,2,4)$ and $H_{03}(1,3,4)$, but test $H_{03}(2,3,4)$, and then stop whatever the outcome. On the other hand if all 6 H_{02} are rejected then even if $H_{02}(1,2,3)$ is rejected we test $H_{03}(1,2,4)$

(as makes sense because non-cointegration of x_{1t}, x_{2t} has already been "established"). But if $H_{02}(1,2,4)$ is also rejected we estimate R to be 2 and stop; this is the maximum possible estimate we can conclude given non-rejection of all H_{02}.

Alternative sequential rules can be determined. The most obvious operates in the opposite direction, testing $H_{0P}(1,...,P)$ first, which is able to immediately determine non-cointegration. Various approaches for determining cointegrating rank have been developed, for both non-fractional (nonstationary) and fractional series. Our algorithm is in one sense laborious, but on the other hand desirably allows lack of knowledge of, and inequalities in, integration orders, and does not require estimation of any cointegrating relations or any user-chosen tuning numbers beyond the bandwidth m.

Though in Section 5 we motivated the allowance for variable integration orders from the perspective of test size, we can also extend the definition of cointegration, and cointegrating rank, to such circumstances, and indeed the algorithm just described is still suitable for estimating cointegrating rank. We adopt the set-up of Robinson and Yajima (2002).

Partition the $P \times 1$ vector x_t into $s > 1$ sub-vectors $x_t^{(i)}$ of dimension P_i, such that each element of $x_t^{(i)}$ has the same integration order $\delta^{(i)}$, $i = 1,...,s$, so $\sum_{i=1}^{s} P_i = P$. In the sense described in Section 2, suppose that $x_t^{(i)}$ has cointegrating rank $R_i < P_i$, where $R_i = 0$ when $x_t^{(i)}$ is not cointegrated. Then we say that x_t has cointegrating rank $R = \sum_{i=1}^{s} R_i$. It is possible that $P_i = 1$ for some i, in which case $R_i = 0$. It is also possible for the $x_t^{(i)}$ to be cross-correlated. Now a routine extension of the consistency discussion of Section 4 indicates that X^{**}, with $z_t = x_t$, can detect cointegration when at least one of the $x_t^{(i)}$ is cointegrated, subject to the caveat in the paragraph following the proof of Theorem 2. Furthermore, our algorithm for estimating cointegrating rank is just as relevant - its description made no reference to equality of integration orders. This is the case whether it is applied to the full

vector x_t, or separately to the $x_t^{(i)}$ for those i such that $P_i \geq 3$. The latter approach is arguably the less expensive except insofar as the partition of x_t is based on pre-testing of equalities between integration orders (as in Robinson and Yajima, 2002). If x_t is over-partitioned, such that there is actually equality between some $\delta^{(\ell)}$, there is potential for under-estimating R.

When there is variation in integration orders there are other definitions of cointegrating rank, as reviewed by Robinson and Yajima (2002). One phenomenon that our algorithm cannot detect is the extension of what has been called "polynomial cointegration" in the integer integration order literature (see Johansen, 1996, p.39). This occurs if a cointegrating error for $x_t^{(i)}$ is cointegrated with elements of x_t, and its investigation would require a more refined analysis.

7. EXTENSION TO NONSTATIONARY SERIES

Velasco (1999) showed that a modified local Whittle estimate for scalar series retains its $m^{\frac{1}{2}}$-consistency and asymptotic normality properties in the presence of quite general nonstationarity. (See also Shimotsu and Phillips, 2005). The nonstationarity is defined by partial summation of stationary fractional series, describing $I(\delta)$ processes for all $\delta > 0$ such that $\delta \neq \frac{1}{2}, \frac{3}{2}, ...$ The modifications consist of data tapering and "skipping" of Fourier frequencies. One anticipates that similar modifications of our statistics can provide tests for non-cointegration of non-stationary vector series.

We consider a sequence $h_t = h_{t,n}$, $t = 1, ..., n$. Following Velasco (1999) we say that $\{h_t\}$ is a taper of order $q \geq 1$ if it is symmetric about $[n/2]$, if (for simplicity) $N = n/q$ is an integer, and if

$$w_h(\lambda) = \frac{c_n(\lambda)}{n^{q-1}} \left\{ \frac{\sin(n\lambda/2q)}{\sin(\lambda/2)} \right\}^q, \tag{7.1}$$

where for all sufficiently large n, $c_n(\lambda)$ has modulus bounded and bounded away from

zero, and $q-1$ derivatives that are bounded, and also

$$\sum_{t=1}^{n} h_t^2 \sim Kn, \quad \text{as } n \to \infty, \quad K \in (0, \infty). \tag{7.2}$$

Velasco (1999) described examples of such sequences. For $q \geq 2$, h_t is roughly constant for central values of t but tapers to zero at the ends, with smoothness indexed by q. In case $q = 1$ we may take $h_t \equiv 1$, $t = 1, ..., n$, so there is no tapering. Velasco (1999) also noted that such a taper can eliminate polynomial trends of degree $q-1$ or less. Define also

$$\psi_n = \left(\sum_{t=1}^{n} h_t^2 \right)^{-2} {\sum_j}' \left\{ \sum_{t=1}^{n} h_t^2 \cos t\lambda_j \right\}^2, \tag{7.3}$$

where the primed sum is over $j = q, 2q, ..., n$.

We discuss an extension only of our robustified statistic X^{**}. Define the product $y_t = h_t z_t$, and thence

$$\hat{G}_h^{**}(x) = \frac{1}{m} {\sum_j}' \Phi(\lambda_j; x) I_y(\lambda_j) \Phi(\lambda_j; x), \tag{7.4}$$

$$\hat{H}_h^{**}(x) = \frac{1}{m} {\sum_j}' \nu_j \Phi(\lambda_j; x) I_y(\lambda_j) \Phi(\lambda_j; x), \tag{7.5}$$

$$s_h^{**}(x) = tr \left\{ \hat{G}_h^{**}(x)^{-1} \hat{H}_h^{**}(x) \right\}. \tag{7.6}$$

Denoting by y_{it} the i-th element of y_t, define

$$\tilde{\delta}_{(i)h} = \arg\min_{d \in U} \left\{ \log \hat{g}_{iih}(d) - \frac{2dq}{m} {\sum_j}'' \log \lambda_j \right\}, \tag{7.7}$$

where

$$\hat{g}_{iih}(d) = \frac{q}{m} {\sum_j}'' I_{y_i}(\lambda_j) \lambda_j^{2d}, \tag{7.8}$$

$\sum_j''$ is a sum over $j = q, 2q, ..., m$ and $U = [\nabla_1, \nabla_2]$ is now a compact interval on the positive real line. Writing $\tilde{\xi}_h = \left(\tilde{\delta}_{(1)h}, ..., \tilde{\delta}_{p(h)} \right)'$ and denoting by $\hat{g}_{iih}^{**}$ the i-th diagonal element of $\hat{G}_y^{**}(\tilde{\xi}_h)$, we define

$$\begin{aligned} \hat{D}_h^{**} &= diag \left\{ \hat{g}_{11h}^{**}, .., \hat{g}_{pph}^{**} \right\}, \quad \hat{R}_h^{**} = \hat{D}_h^{**-\frac{1}{2}} \hat{G}_h^{**}(\tilde{\xi}) \hat{D}_h^{**-\frac{1}{2}}, \\ X_h^{**} &= (m/q\psi_n) s_h^{**}(\tilde{\xi})^2 \Big/ \left(tr \left\{ \hat{R}_h^{**2} \right\} - p \right). \end{aligned} \tag{7.9}$$

We define a non-cointegrated, possibly nonstationary z_t as follows. Let v_t be a $p \times 1$ vector stationary process with zero mean such that

$$v_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} \|C_j\|^2 < \infty, \tag{7.10}$$

where the ε_t satisfy the conditions stated before Theorem 1. For $\delta_i > 0$, $i = 1, ..., p$, define $s_i = [\delta_i + \frac{1}{2}]$, $\kappa_i = \delta_i - s_i$, so that $\kappa_i = \delta_i$ for $0 < \delta_i < \frac{1}{2}$, $\kappa_i = \delta_i - 1$ for $\frac{1}{2} < \delta_i < \frac{3}{2}$, and so on. We suppose that $C(\lambda)$ defined as in (4.2) satisfies (5.8) and (5.9) with ξ replaced by $\kappa = (\kappa_1, ..., \kappa_p)'$ and $\beta \in (1, 2]$. Finally define

$$z_{it} = \Delta^{-s_i} v_{it}^*, \quad t \geq 1, \quad i = 1, ..., p, \tag{7.11}$$

where v_{it}^* is the i-th element of v_t, for $t \geq 1$, and zero for $t \leq 0$. Then we may call z_{it} an $I(\delta_i)$ process, for all $\delta_i > 0$ except $\delta_i = \frac{1}{2}, \frac{3}{2}, ...$ Finally we assume that for all i, $\delta_{0i} \in [\nabla_1, \nabla_2]$ where $\nabla_1 > 0$, and $\left[\nabla_2 + \frac{1}{2}\right] + 1 \leq q$.

Theorem 4 *Under the assumptions in the previous paragraph,*

$$X_h^{**} \rightarrow_d \chi_1^2, \quad \text{as } n \rightarrow \infty. \tag{7.12}$$

The proof combines ideas from Velasco (1999) and the proofs of Theorems 1 and 3 too straightforwardly to warrant discussion. By virtue of the full rank assumption on Q, z_t is not cointegrated irrespective of whether there is equality among any δ_i. We can define a cointegrated z_t by means of (2.5) with u_t generated similarly to v_t above, but such that v_{it} is $I(\gamma_i)$ with $0 \leq \gamma_i \leq \delta_i$ for all i, and $\gamma_i < \delta_i$ for some i; again cointegration requires equality of at least two δ_i. Since Theorem 4 covers integer values of the δ_i, γ_i, it would be possible to conduct a test in the often-assumed situation that observables are $I(1)$ and any cointegrating errors are $I(0)$. Procedures which use this information (or other prior information on integration orders, either

stationary or nonstationary ones) would be expected to perform much better when it is correct. The sequential algorithm described in the previous section can still be applied to estimate cointegrating rank in nonstationary environments.

8. MONTE CARLO STUDY OF FINITE-SAMPLE PROPERTIES

A Monte Carlo study was carried out to investigate finite-sample performance of the procedures. All the experiments reported employed 1000 replications.

Our first experiments compare size and power of the tests based on X, X^* and X^{**} (see Sections 3-5) when $p = 2$. With respect to size we employ bivariate non-cointegrated sequences

$$z_t = \Delta^{-0.35} e_t, \quad t = 1, 2, ..., n, \tag{8.1}$$

where the e_t are independent bivariate normal vectors, whose first and second elements have mean zero, have standard deviations 1 and $\sqrt{65}$ respectively, and correlation $8/\sqrt{65}$. Thus we are considering X^{**} in a setting in which the robustification is unnecessary. We generated series of lengths $n = 128$, 512 and 1024. In the first place, the estimates $\widetilde{\delta}_{(1)}$ and $\widetilde{\delta}_{(2)}$ (3.8) of $\delta = 0.35$ were computed, for $m = 10$, 20, 40 when $n = 128$; $m = 20$, 40, 80, 150 when $n = 512$; and $m = 80, 150$ and 300 when $n = 1024$. Then $\widetilde{\delta}$ (3.10) was computed in the equal-weights case (3.12). Table 1 reports empirical size of tests based on Theorem 1 with nominal sizes $\alpha = 0.01$ and 0.05, along with Monte Carlo mean-squared error (MSE) of $\widetilde{\delta}$.

(Table 1 about here)

Empirical sizes are clearly too small when $m = 10$ and $n = 128$, but in other cases they are not too bad, if mainly too small especially for X^{**}, though there is similarity across X, X^*. On the whole, sizes do not vary greatly over n, though there is sensitivity to m; m is a measure of effective degrees of freedom, so the approximation to asymptotic behaviour does not seem too bad in the circumstances. The MSE of $\widetilde{\delta}$

decays with increasing n, and also with increasing m, due in part to low bias resulting from the simple monotonically decaying spectrum in this experiment.

We next generated cointegrated series z_t, by (2.5), (2.9) with $\gamma_1 = 0.05$, $\gamma_2 = \delta = 0.35$, the e_t as in the previous experiment, and

$$B = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}. \tag{8.2}$$

The results, with otherwise the same specifications as before, are presented in Table 2. The most striking feature is the abysmal performance of X, except for the largest m when $n = 512$ and 1024, indeed the powers are actually mostly less than the empirical sizes reported in Table 1. The fact that we did not include a consistency proof for X does not imply it is not consistent, indeed powers do dramatically increase as m goes from 80 to 150. Looking at a single replicate, we found the denominators of X and X^* to be almost identical, whereas $s(\tilde{\delta}) = -0.0025$ and $s^*(\tilde{\delta}) = -0.1974$, this substantial difference being due to the imaginary parts of $G^*(\tilde{\delta})$ and $H^*(\widetilde{\delta})$ (which in the latter case are slightly larger than the real parts of the corresponding off-diagonal elements). In Table 2, X^{**} performs disappointingly against X^*, given that the derivations in Section 5 indicate comparable local power; we can only suggest that our departure from non-cointegration should not be interpreted as local. However, X^{**} mainly does substantially better than X, except for the largest m. The powers increase monotonically with m except for $n = 1024$ where they fall then rise. Even the powers of X^* are poor for small m with $n = 128$ and 512, but they increase quite rapidly with m, and are uniformly high with $n = 1024$. Its performance seems quite satisfactory, especially as one expects cointegration with gap $\delta - \gamma_1 = 0.3$ to be much harder to detect than, say, with $I(1)$ observables and $I(0)$ cointegrating errors, where $\delta - \gamma_1 = 1$.

(Table 2 about here)

We now go on to study performance of the algorithm for choosing r described in

Section 6, using X^{**}. We generated a trivariate system with cointegrating rank 1. Specifically, we used (2.5), (2.9), where the e_t are independent trivariate normal with zero mean and covariance matrix

$$\begin{bmatrix} 1 & 8 & 1 \\ 8 & 65 & 9 \\ 1 & 9 & 3 \end{bmatrix}, \tag{8.3}$$

while

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8.4}$$

and

$$\gamma_1 = 0.1, \quad \gamma_2 = 0.4, \quad \gamma_3 = 0.2. \tag{8.5}$$

Thus, z_t has elements that do not all have the same integration order, as discussed in the penultimate paragraph of Section 6. In the notation there, $z_t = x_t$, $s = 2$, $P_1 = 2$, $P_2 = 1$, $P = 3$, $R_1 = 1$, $R_2 = 0$, $R = 1$; the first two elements of z_t are $I(0.4)$, the last one is $I(0.2)$, and the cointegrating error is $I(0.1)$. This setting illustrates the need for our robustified test statistic X^{**}, which was employed in precisely the algorithm described for the case $P = 3$ in Section 6 . Two different rules for choosing the χ^2 nominal sizes were employed. In one (see Table 3) we used 0.05 and 0.01 for each of the (up to 4) tests needed in the algorithm to estimate R from a given data set. In the other (see Table 4) we approximated these sizes (for testing $R = 0$ over the whole algorithm) by means of Bonferroni's inequality, as described in Section 6; thus, nominal sizes $0.05/4 = 0.0125$ and $0.01/4 = 0.0025$) were used for each χ^2 test. Relative frequencies of $\hat{R}$ for these two rules are displayed in Tables 3 and 4, with the same choices of n and m as before.

(Tables 3 and 4 about here)

Throughout, we see a clear tendency to underestimate R; it is seldom over-estimated. This is especially notable for $n = 128$, but the results improve markedly with increasing n. For $n = 512$ there is definite improvement with increasing m, while for $n = 1024$ the opposite effect is observed. Of course the choice of nominal significance level is always arbitrary in any case. While the use or not of Bonferroni does not make a huge difference to the results, it is obvious that were we to choose a somewhat larger nominal size, say 0.1, especially in Table 4, the results would improve. Bearing in mind the stress we have placed on computational simplicity in developing tests, the results overall do not seem disappointing.

We also examined the performance of the tapered statistic X_h^{**} in a nonstationary setting. To examine size, we generated bivariate vectors

$$v_t = \Delta^{-0.2} e_t, \quad t = 1, 2, ..., n, \tag{8.6}$$

where the e_t were as in (8.1). Then we formed the partial sums

$$z_t = \sum_{i=1}^{t} v_i \tag{8.7}$$

(i.e. (7.11) with $s_1 = s_2 = 1$). Thus the elements of z_t are non-cointegrated $I(1.2)$ series. We formed the tapered vectors $y_t = h_t z_t$, where $h_t = h\left((t - \frac{1}{2})/n\right)$ in which $h(u)$ is the cosine bell

$$h(u) = \frac{1}{2}\left(1 - \cos(2\pi u)\right), \quad 0 \leq u \leq 1. \tag{8.8}$$

In this case $q = 3$, and we constructed the estimates $\tilde{\delta}_{(1)h}, \tilde{\delta}_{(2)h}$, and thence the statistic X_h^{**}, on this basis, as described in the previous section. We took $n = 129$ with $m = 12, 21, 42$; $n = 513$ with $m = 21, 42, 81, 150$; and $n = 1023$ with $m = 81, 150, 300$. Once again nominal sizes $\alpha = 0.01$ and 0.05 were employed in the test justified in Theorem 4.

(Table 5 about here)

Empirical sizes, and MSE of $\tilde{\delta}_{(1)h}, \tilde{\delta}_{(2)h}$, are presented in Table 5. Sizes are almost uniformly too small, seriously so for small m, though when $n = 512$ and 1024 with $\alpha = 0.05$ they seem satisfactory for the largest m. MSEs of the δ estimates are worse than those in Table 1, perhaps predictably in view of the tapering, and this may be partly to blame for the more disappointing of the results.

Finally, the power of the X_h^{**} test was examined in bivariate nonstationary cointegrated series. We generated

$$v_t = diag\left\{\Delta^{-0.4}, \Delta^{-0.2}\right\} e_t, \quad t = 1, 2, ..., n, \tag{8.9}$$

with e_t as before, then

$$u_{1t} = v_{1t}, \tag{8.10}$$

$$u_{2t} = \sum_{i=1}^{t} v_{2i}, \tag{8.11}$$

,and finally $z_t = B^{-1}u_t$ with B as in (8.2). Thus the elements of z_t are cointegrated $I(1.2)$ series with $I(0.4)$ cointegrating errors. Then $\tilde{\delta}_{(1)h}, \tilde{\delta}_{(2)h}$ and X_h^{**} were computed as before.

(Table 6 about here)

The results are presented in Table 6. When $n = 129$ the powers are extremely poor, and except for $m = 150$ they are disappointing also when $n = 513$. This must at least in part be due to the under-sizing. When $n = 1023$ power is not too bad with $m = 150$, and high with $m = 300$. In the latter case there is over-sizing, but this is only substantial when $\alpha = 0.01$. It is important to recall that (to reduce correlation across frequencies induced by tapering) we skip two Fourier frequencies between each included one ($q = 3$), so in Tables 5 and 6 the actual number of frequencies used is only about $m/3$.

Looking back at all our Monte Carlo results, an overall conclusion is that, except perhaps in case of the X^* statistic, $n = 128$ is too short a series for these semipara-

metric procedures. On the other hand, results for the largest n are in several instances promising.

A notable feature of the Monte Carlo results is sensitivity to bandwidth, m. This is common in smoothed estimation, and a popular reaction is to employ a data-dependent bandwidth that has some optimality properties. We have deliberately avoided pursuing this approach on several grounds.

The first is due to ambiguity about how to base such a choice. For the series z_{it}, for given i, we could employ an approximate minimum-MSE rule (see e.g. Henry and Robinson, 1996) to estimate its integration order (though strictly such rules do not seem to have been explicitly studied in the nonstationary case). But this would not be optimal for some other element of z_t, while on the other hand if we used different m's for each series this could affect the limit distribution of our test statistics. Nor is the viability of employing a bandwidth choice procedure based on the multivariate local Whittle function (see e.g. Lobato, 1999) assuming common integration order clear, because its properties vary depending on whether or not there is cointegration (cf Robinson and Yajima, 2002).

Another kind of problem with "optimal" bandwidth choices, is that the simplest ones assume twice differentiability near frequency zero of the ratio between the spectral density and its power law approximation. As noted earlier in this section, this smoothness property is violated when, as is plausible for cointegrated systems, series contain components of different integration orders. Thus, the usual rules may lead to over-smoothing. More elaborate procedures that take account of the lesser smoothness in a systematic way can be developed, but at additional computational cost.

Furthermore, the relevance of using a minimum-MSE bandwidth for integration order estimation in our statistics is dubious. What seems more appropriate is an m that somehow makes the error in the null χ^2 distribution small. It would be possible

to develop a theory that could lead to a data-dependent m of this type but it would require considerable work, involving Edgeworth expansion (cf. Giraitis and Robinson, 2003).

Finally, as is common, our asymptotics are all based on data-free m, and asymptotic behaviour with data-dependent m cannot be taken for granted. Rather than encourage the practitioner to rely on the outcome provided by a data-dependent bandwidth selection rule, we prefer to recommend in these circumstances a more informal approach, in which the test statistic is computed across a grid of m values. Sensitivity can then be assessed and qualified judgements made.

9. FINAL COMMENTS

We have presented computationally simple tests for cointegration, and embedded one of them in an algorithm for estimating cointegrating rank. The tests require no knowledge of integration order, and apply to fractional series as well as non-fractional ones, and cover both stationary and nonstationary data. The tests are all semiparametric in character, at the cost of requiring a user-chosen bandwidth. Versions based on a parametric specification of the autocorrelation of the input $I(0)$ vector could be developed; given correct specification they would have faster rates of convergence, and thus probably better finite-sample performance. However, such tests could not be justified in the same general way, and at least for long series our more robust approach seems preferable, especially in view of the relatively simple computation of semiparametric integration order estimates compared to parametric ones. Unlike in some rival procedures no other user-chosen tuning numbers are required. All null limit distributions are standard. A careful preliminary study of the data may provide information that can be used in more powerful tests, but our approach has the benefit of computational simplicity and wide generality.

Acknowledgements

I am very grateful to Supachoke Thawornkaiwong for carrying out the Monte Carlo simulations reported in Section 8, and three referees for numerous comments which have improved the paper. This research was supported by ESRC Grants R000239936 and RES-062-23-0036.

REFERENCES

Giraitis, L., Robinson, P.M., 2003. Edgeworth expansions for semiparametric Whittle estimation of long memory. Annals of Statistics 31, 1325-1375.

Hausman, J., 1978. Misspecification tests in econometrics. Econometrica 46, 1251-1271.

Henry, M., Robinson, P.M., 1996. Bandwidth choice in Gaussian semiparametric estimation of long range dependence. Athens Conference onApplied Probability and Time Series Analysis. Volume II: Time Series Analysis. In Memory of E.J.Hannan, (P.M. Robinson and M. Rosenblatt, eds.) Lecture Notes in Statistics Volume 115, Springer-Verlag, 220-232.

Hualde, J., 2004. Testing for the equality of orders of integration. Preprint, Universidad de Navarra.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231-254.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551-1580.

Johansen, S., 1996. Likelihood Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

Künsch, H.R., 1987. Statistical aspects of self-similar processes. Proceedings of the First World Congress of the Bernoulli Society. VNU Science Press, Utrecht, 67-74.

Lobato, I.N,. 1999. A semiparametric two-step estimator in a multivariate long memory model. Journal of Econometrics 90, 129-153.

Marinucci, D., Robinson, P.M., 2001. Semiparametric fractional cointegration analysis. Journal of Econometrics 105, 225-247.

Phillips, P.C.B., Ouliaris, S, 1988. Testing for cointegration using principal components methods. Journal of Economic Dynamics and Control 12, 205-230.

Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residual-based cointegration tests. Econometrica 58, 165-193.

Robinson, P.M., 1994a. Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515-539.

Robinson, P.M., 1994b. Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89, 1420-1437.

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630-1661.

Robinson, P.M., Yajima, Y., 2002. Determination of cointegrating rank in fractional systems. Journal of Econometrics 106, 217-241.

Shimotsu, K,. 2007. Gaussian semiparametric estimation of multivariate fractionally integrated processes. Journal of Econometrics 137, 277-310.

Shimotsu, K., Phillips, P.C.B., 2005. Exact local Whittle estimation of fractional integration. Annals of Statistics 33, 1890-1933.

Velasco, C., 1999. Gaussian semiparametric estimation of non-stationary time series. Journal of Time Series Analysis 20, 87-127.

Table 1: Rejection frequencies of the correct H_0 of no cointegration at level α, and MSE of $\tilde{\delta}$ (stationary observables).

		$\alpha = 0.01$			$\alpha = 0.05$			
n	m	X	X^*	X^{**}	X	X^*	X^{**}	$MSE\left(\tilde{\delta}\right)$
128	10	0.001	0.000	0.000	0.006	0.006	0.006	0.027
	20	0.012	0.011	0.008	0.032	0.032	0.022	0.016
	40	0.020	0.020	0.014	0.038	0.037	0.032	0.008
512	20	0.008	0.005	0.005	0.027	0.028	0.018	0.014
	40	0.015	0.015	0.011	0.035	0.032	0.028	0.007
	80	0.014	0.014	0.011	0.042	0.038	0.029	0.003
	150	0.011	0.010	0.009	0.058	0.054	0.050	0.002
1024	80	0.013	0.011	0.009	0.038	0.036	0.034	0.003
	150	0.010	0.009	0.008	0.032	0.035	0.032	0.002
	300	0.006	0.006	0.006	0.041	0.044	0.042	0.001

Table 2: Rejection frequencies of the incorrect H_0 of no cointegration at level α, and MSE of $\tilde{\delta}$ (stationary observables).

		$\alpha = 0.01$			$\alpha = 0.05$			
n	m	X	X^*	X^{**}	X	X^*	X^{**}	$MSE\left(\tilde{\delta}\right)$
128	10	0.000	0.000	0.000	0.000	0.003	0.003	0.028
	20	0.001	0.053	0.017	0.002	0.215	0.069	0.017
	40	0.014	0.440	0.034	0.077	0.688	0.063	0.009
512	20	0.000	0.045	0.030	0.001	0.183	0.091	0.014
	40	0.000	0.405	0.093	0.000	0.650	0.201	0.007
	80	0.001	0.901	0.164	0.002	0.973	0.244	0.004
	150	0.442	0.998	0.305	0.812	1.000	0.385	0.002
1024	80	0.000	0.885	0.316	0.001	0.964	0.475	0.004
	150	0.000	0.996	0.275	0.004	0.999	0.406	0.002
	300	0.968	1.000	0.395	0.998	1.000	0.473	0.001

Table 3: Frequencies of $\hat{R}$ at level α (without Bonferroni's inequality).

		$\alpha = 0.01$			$\alpha = 0.05$		
n	m	$\hat{R} = 2$	$\hat{R} = 1$	$\hat{R} = 0$	$\hat{R} = 2$	$\hat{R} = 1$	$\hat{R} = 0$
128	10	0.003	0.002	0.995	0.006	0.000	0.994
	20	0.003	0.015	0.982	0.008	0.037	0.955
	40	0.001	0.075	0.924	0.007	0.114	0.879
512	20	0.005	0.017	0.978	0.007	0.044	0.949
	40	0.004	0.129	0.867	0.012	0.209	0.779
	80	0.004	0.246	0.750	0.011	0.316	0.673
	150	0.000	0.332	0.668	0.012	0.384	0.604
1024	80	0.002	0.410	0.588	0.006	0.531	0.463
	150	0.003	0.416	0.581	0.005	0.521	0.474
	300	0.003	0.429	0.568	0.008	0.460	0.532

Table 4: Frequencies of $\hat{R}$ at level α (with Bonferroni's inequality).

		$\alpha = 0.01$			$\alpha = 0.05$		
n	m	$\hat{R} = 2$	$\hat{R} = 1$	$\hat{R} = 0$	$\hat{R} = 2$	$\hat{R} = 1$	$\hat{R} = 0$
128	10	0.004	0.001	0.995	0.006	0.002	0.992
	20	0.004	0.017	0.972	0.008	0.040	0.952
	40	0.002	0.077	0.921	0.007	0.118	0.875
512	20	0.005	0.018	0.977	0.009	0.054	0.937
	40	0.005	0.138	0.857	0.012	0.222	0.766
	80	0.005	0.255	0.740	0.012	0.334	0.654
	150	0.002	0.337	0.661	0.013	0.389	0.598
1024	80	0.002	0.424	0.574	0.006	0.559	0.435
	150	0.003	0.433	0.564	0.006	0.538	0.456
	300	0.003	0.433	0.564	0.008	0.476	0.516

Table 5: Rejection frequencies of the correct H_0 of no cointegration at level α, and MSE of $\tilde{\delta}_{(1)h}$, $\tilde{\delta}_{(2)h}$ (nonstationary observables).

n	m	$\alpha = 0.01$	$\alpha = 0.05$	$MSE\left(\tilde{\delta}_{(1)h}\right)$	$MSE\left(\tilde{\delta}_{(2)h}\right)$
129	12	0.000	0.000	0.281	0.274
	21	0.000	0.001	0.128	0.133
	42	0.006	0.014	0.049	0.048
513	21	0.000	0.001	0.139	0.145
	42	0.004	0.006	0.040	0.042
	81	0.007	0.018	0.016	0.015
	150	0.015	0.042	0.009	0.009
1023	81	0.004	0.016	0.014	0.014
	150	0.007	0.030	0.007	0.007
	300	0.021	0.054	0.005	0.005

Table 6: Rejection frequencies of the incorrect H_0 of no cointegration at level α, and MSE of $\tilde{\delta}_{(1)h}$, $\tilde{\delta}_{(2)h}$ (nonstationary observables).

n	m	$\alpha = 0.01$	$\alpha = 0.05$	$MSE\left(\tilde{\delta}_{(1)h}\right)$	$MSE\left(\tilde{\delta}_{(2)h}\right)$
129	12	0.000	0.000	0.281	0.274
	21	0.000	0.000	0.148	0.133
	42	0.003	0.011	0.066	0.048
513	21	0.000	0.000	0.149	0.145
	42	0.002	0.004	0.044	0.042
	81	0.045	0.167	0.017	0.015
	150	0.672	0.894	0.015	0.009
1023	81	0.019	0.094	0.014	0.014
	150	0.322	0.572	0.008	0.007
	300	0.987	0.998	0.010	0.004