ELSEVIER

Journal of Econometrics 106 (2002) 217–241

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Determination of cointegrating rank in fractional systems

Peter M. Robinson[a,*], Yoshihiro Yajima[b]

[a] *Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK*

[b] *Faculty of Economics, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan*

Received 6 July 2000; revised 4 May 2001; accepted 1 June 2001

Abstract

This paper develops methods of investigating the existence and extent of cointegration in fractionally integrated systems. We focus on stationary series, with some discussion of extension to nonstationarity. The setting is semiparametric, so that modelling is effectively confined to a neighbourhood of frequency zero. We first discuss the definition of fractional cointegration. The initial step of cointegration analysis entails partitioning the vector series into subsets with identical differencing parameters, by means of a sequence of hypothesis tests. We then estimate cointegrating rank by analysing each subset individually. Two approaches are considered here, both of which are based on the eigenvalues of an estimate of the normalized spectral density matrix at frequency zero. An empirical application to a trivariate series of oil prices is included.

JEL classification: C22

Keywords: Fractional cointegration; Long memory

1. Introduction

Cointegration analysis has principally been developed theoretically, and applied empirically, in the "$I(1)/I(0)$" context, in which components of a multivariate empirical series displaying evidence of unit root nonstationarity

* Corresponding author. Tel.: +44-20-7955-7516; fax: +44-20-7831-1840.
E-mail address: p.m.robinson@lse.ac.uk (P.M. Robinson).

0304-4076/02/$ - see front matter
PII: S0304-4076(01)00096-3

($I(1)$)) are examined for the existence of one or more stationary, short memory ($I(0)$) linear relations. We call a scalar time series $u_t, t = 0, \pm 1, \pm 2, \ldots, I(0)$ if it is covariance stationary and has spectral density that is finite and positive at zero frequency; for such a u_t series we call the scalar series $v_t = (u_1 + \cdots + u_t)1(t > 0), t = 0, \pm 1, \pm 2, \ldots, I(1)$, where 1(.) is the indicator function. Then a $p \times 1$ vector variate X_t with $I(1)$ elements has been said to be cointegrated if there exists a linear combination $\beta' X_t$, the prime denoting transposition, that is $I(0)$. If there are r, $1 \leqslant r \leqslant p-1$, such relations, with linearly independent coefficients, we can express X_t in terms of unobservable components,

$$X_t = AF_t + V_t, \quad t \geqslant 1, \tag{1}$$

where F_t is a $(p-r) \times 1$ vector of unobservable (not cointegrated) $I(1)$ series, A is a full rank $p \times (p-r)$ matrix, and V_t is a $p \times 1$ vector of unobservable $I(0)$ series; F_t can be interpreted as a vector of common trends (see e.g., Stock and Watson, 1988). The integer r is termed the *cointegrating rank* and the $\beta' X_t$ are termed *cointegrating errors*. This basic setup has been extended in various directions, for example to incorporate deterministic trends, but the basic notion of observables with an $I(1)$ component, and of $I(0)$ cointegrating errors, is standard. Many tools of statistical inference have been developed to investigate the existence of such cointegration, and widely applied empirically (see e.g. Banerjee et al., 1993; Engle and Granger, 1991; Hatanaka, 1996; Johansen, 1996), for example in connection with consumption and income data, term structure of interest rates, and purchasing power parity between exchange rates.

The possibility of a long-run stable relationship existing between nonstationary series X_t exists irrespective of whether or not the series are $I(1)$, however. Indeed, one can also envisage the possibility of cointegration with $\beta' X_t$ stationary but not necessarily $I(0)$, while cointegration can also be defined for stationary X_t. There is considerable interest in identifying structure in multivariate series, and thus a strong case for a flexible approach. This is permitted by the class of $I(d)$ series, with real-valued d. For $d < 1/2$, we say v_t is $I(d)$ if $u_t = (1-L)^d v_t$ is $I(0)$, where L is the lag operator and

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j, \tag{2}$$

where $\Gamma(a) = \int_0^\infty x^{a-1} \mathrm{e}^{-x}\, \mathrm{d}x$ for $a > 0$, while for $a = -n$, $n = 0, 1, \ldots, \Gamma(a)$ has simple poles with residues $(-1)^n/n$, and for other $a < 0, \Gamma(a)$ is obtainable by the recursion $\Gamma(a) = \Gamma(a+1)/a$. Then v_t is covariance stationary. For $d \geqslant 1/2$, we define a nonstationary $I(d)$ series $v_t = (1-L)^{-d}\{u_t 1(t > 0)\}$. Clearly, this $I(d)$ class nests the $I(0)$ and $I(1)$ series. The parameter d, called the *fractional differencing* parameter, can be said to describe the *memory* of v_t. In fact, while the early paper of Engle and Granger (1987) focussed principally on cointegration in the "$I(1)/I(0)$" context, it included a definition

which applies to $I(d)$ series (though its $I(d)$ class is much narrower than ours): X_t can be said to be cointegrated $CI(d,b)$ if X_t has $I(d)$ elements and, for some $b>0$, there exists β such that $\beta' X_t$ is $I(d-b)$. Thus, the original definition takes $d=b=1$. If either d and/or b is non-integral, we have *fractional cointegration*. Representation (1) now applies with F_t a vector with $I(d)$ components and V_t a vector with $I(d-b)$ components.

Some empirical study of fractional cointegration has already been carried out, see e.g. Cheung and Lai (1993), Robinson and Marinucci (1998). However, rigorously justified procedures are currently in short supply, especially in the most practically interesting situation in which d and/or b are unknown. Here, the study of fractional cointegration clearly presupposes a good understanding of statistical inference on $I(d)$ series, in particular on a theory of estimation of d. At present this has been much better developed in case of stationary and invertible series, that is when $-1/2<d<1/2$, than for nonstationary ones, and partly for that reason the present paper focusses principally on the possibility of cointegration in stationary series. This may be of direct interest in financial series, for example exchange rates between three or more currencies, and some empirical series that have been regarded as having unit roots could be better modelled as $I(d)$ with even $d<1/2$. Indeed, Robinson (1994) considered cointegration of stationary $I(d)$ series, showing that β can be consistently estimated here, and Robinson and Marinucci (1998) developed his approach. This estimate of β converges at only a nonparametric rate, requiring only the assumptions on spectral behaviour of X_t entailed in our $I(d)$ definition: defining the spectral density matrix $f(\lambda)$ of X_t to satisfy $\mathrm{E}\{(X_t-\mu)(X_{t+j}-\mu)'\}=\int_{-\pi}^{\pi} f(\lambda)\,\mathrm{e}^{ij\lambda}\,\mathrm{d}\lambda$, where $\mu=\mathrm{E}(X_t)$, we have

$$f(\lambda)\backsim G\lambda^{-2d} \quad \text{as } \lambda\to 0+, \tag{3}$$

where '$\backsim$' is taken elementwise, to mean that the ratio of real parts, and of imaginary parts, of left and right sides tends to 1, and G is a finite nonnegative definite matrix each of whose diagonal elements is non-null, G being positive definite if and only if X_t is not cointegrated (cf. Robinson and Marinucci, 1998). Under similar assumptions, with G positive definite, a theory of "semiparametric" estimation of d has been developed that entails similar convergence rates. More efficient inferences are possible on the basis of parametric models for the autocorrelation in X_t, such as fractionally integrated autoregressive moving average (FARIMA) models, developed by Adenstedt (1974) and subsequent authors, but model misspecification (under-specifying either the autoregressive or moving average orders or over-specifying both) is liable to lead to inconsistent estimation of fractional differencing and other parameters. Thus, the present paper focusses on a semiparametric, low-frequency, approach, as seems natural in the sense that cointegration is essentially a low-frequency phenomenon, and justifiable insofar as many financial series are sufficiently long that nonparametric rates afford acceptable precision.

A basic question of interest is the existence of cointegration, or, more generally, the value of the cointegrating rank r. This problem has been quite well solved in some cases of integer d, b, especially $d = b = 1$ (see e.g. Johansen, 1996; Phillips and Ouliaris, 1988, 1990). Of course, the methods developed there would not necessarily be expected to detect cointegration when in fact alternative values of d and/or b prevail (see e.g. Abadir and Taylor, 1999), and indeed the unit root tests directed against stationary autoregressive (AR) alternatives of Dickey and Fuller (1981) that are commonly employed do not have very good power against fractional alternatives (see e.g., Cheung and Lai, 1993; Diebold and Rudebusch, 1991). While one might envisage some relatively straightforward extension to a fractional context, involving specifying and testing null values of d and b (ideally using tests directed against fractional alternatives), it seems preferable to treat d and b as nuisance parameters throughout. Here, while proposals have been made for analysing cointegration, little rigorous theoretical justification is available. The present paper makes some attempt to redress this situation. We do not discuss estimation of β or b in the event of cointegration, and consider estimation of d only for the purposes of testing for cointegration.

The paper is organized as follows. Cointegration requires at least two elements of X_t to have the same differencing parameters, but in the following section we discuss a definition of fractional cointegration that allows for some variation in the differencing parameters of elements of X_t when $p \geqslant 3$ (in which case a more general representation than (1) would apply). When all elements have the same differencing parameter, as in (3), $r = p - rank(G)$. When they do not, on reordering and partitioning the elements of X_t into subvectors that have the same differencing parameter, we have a representation of type (3) for the spectral density matrix of each subvector, and the overall cointegrating rank is p minus the sum of the ranks of the "G" matrices for each subvector. Thus, the problem essentially reduces to one of determining the rank of G in (3). First, however, in Section 3, we describe a testing algorithm, of a type previously used in the analysis of variance, for the initial partitioning of X_t; this problem is complicated by the fact that the theory of estimating differencing parameters varies depending on whether or not there is cointegration, which is not known at the partitioning stage. We may then focus on (3) in Section 4, deriving the limit distribution of eigenvalues of an estimate of G under no-cointegration, and thence proposing two methods for determining its rank, one an extension of that of Phillips and Ouliaris (1988) for investigating $CI(1,1)$ cointegration, the other a model selection procedure. In Section 5, we apply the procedures to a trivariate series of prices of crude oil. Section 6 contains final comments, including a brief discussion of implications of our work for investigating fractional cointegration in a nonstationary context. Proofs of theoretical results are left to Appendix A.

2. Fractional integration and cointegration

We first extend our $I(d)$ definition to X_t whose p elements are permitted to have distinct differencing parameters. For $|d_a| < 1/2$, $a = 1, \ldots, p$, and $\lambda > 0$, define the $p \times p$ matrix

$$\Lambda(\lambda) = \mathrm{diag}\{e^{i\pi d_1/2}\lambda^{-d_1}, \ldots, e^{i\pi d_p/2}\lambda^{-d_p}\} \tag{4}$$

and its complex conjugate $\bar{\Lambda}(\lambda)$.

Definition 1. We call X_t an $I(d_1, \ldots, d_p)$ series if and only if

$$f(\lambda) \sim \Lambda(\lambda) G \bar{\Lambda}(\lambda) \quad \text{as } \lambda \to 0+ \tag{5}$$

and all diagonal elements of the $p \times p$ matrix G are nonzero.

Remark 1. (i) If $d_1 = \cdots = d_p = d$, then (5) reduces to (3), and so an $I(d, \ldots, d)$ series is a vector of $I(d)$ series.

(ii) We can relate the definition to fractionally differenced time-series models such as FARIMAs. Define the $p \times p$ matrix functions

$$\mathrm{E}(L) = \mathrm{diag}\{(1-L)^{d_1}, \ldots, (1-L)^{d_p}\}, C(L) = \sum_{j=0}^{\infty} C_j L^j \tag{6}$$

for $p \times p$ matrices C_j such that $\sum_{j=0}^{\infty} \mathrm{tr}\{C_j C_j'\} < \infty$, and consider X_t given by

$$\mathrm{E}(L)(X_t - \mu) = C(L)e_t, \quad t = 0, \pm 1, \ldots, \tag{7}$$

where the $p \times 1$ vectors e_t are such that $\mathrm{E}e_t = 0$, $\mathrm{E}e_s e_t' = 0$ for $s \neq t$, and $\mathrm{E}e_t e_t'$ is positive definite, and thus taken with no loss of generality to be the identity matrix. Then

$$f(\lambda) = \mathrm{E}(e^{i\lambda})^{-1} C(e^{i\lambda}) C(e^{-i\lambda})' \mathrm{E}(e^{-i\lambda})^{-1}/(2\pi). \tag{8}$$

Now for $\lambda \neq 0, \mathrm{mod}(\pi)$, and $|d| < 1/2$, $(1 - e^{i\lambda})^{-d} = \sum_{j=0}^{\infty} a_j(\cos(j\lambda) + i\sin(j\lambda))$, where $a_j = \Gamma(j+d)/(\Gamma(d)\Gamma(j+1))$. Because the a_j decrease monotonically to zero and $a_j \sim j^{d-1}$ it follows from Theorem III-I of Yong (1974) that as $\lambda \to 0+$, $(1 - e^{i\lambda})^{-d} \to \lambda^{-d}(\sin((1-d)\pi/2) + i\cos((1-d)\pi/2)) = \lambda^{-d} e^{id\pi/2}$. Thus, Definition 1 is satisfied with

$$G = C(1)C(1)'/(2\pi) \tag{9}$$

if all rows of $C(1)$ are non-null. One choice of $C(L)$ is $C(L) = \Phi(L)^{-1}\Psi(L)$, where $\Phi(L)$ and $\Psi(L)$ are finite-degree polynomial matrices such that all zeros of $\det\{\Phi(z)\}$ and $\det\{\Psi(z)\}$ lie outside the unit circle. Then X_t has the

vector FARIMA representation

$$\Phi(L)\mathrm{E}(L)(X_t-\mu)=\Psi(L)e_t,\quad t=0,\pm1,\ldots, \tag{10}$$

see e.g. Lobato (1995), Robinson (1995a). Note that Engle and Granger (1987) defined a vector $I(d)$ series by (10) with $d_1=\cdots=d_p=d$. With unequal d_a we can alternatively define a vector FARIMA as

$$\mathrm{E}(L)\Phi(L)(X_t-\mu)=\Psi(L)e_t,\quad t=0,\pm1,\ldots \tag{11}$$

(see Lobato, 1995). To relate this to (5) we order elements of X_t such that

$$d_1=\cdots=d_{i_1}>d_{i_1+1}=\cdots=d_{i_2}>\cdots>d_{i_{s-1}+1}=\cdots=d_{i_s}, \tag{12}$$

where $1\leqslant s\leqslant p$, $1\leqslant i_1<i_2<\cdots<i_s=p$ with $i_0=0$; Eq. (12) will also be employed in our discussion of cointegration. Writing ϕ^{ab} for the (a,b)th element of $\Phi(1)^{-1}$ and ψ_b' for the bth row of $\Psi(1)$, define $\theta_{al}=\sum_{b=i_{l-1}+1}^{i_l}\phi^{ab}\psi_b$ and $\tilde{d}_a=d_{i_{l(a)}}$, where $l(a)=\arg\max_{l:\theta_{al}\neq0}d_{i_l}$. Then if $\theta_{al(a)}\neq0$, $a=1,\ldots,p$, it follows that X_t is an $I(\tilde{d}_1,\ldots,\tilde{d}_p)$ series, such that in (5) $\Lambda(\lambda)$ is replaced by $\mathrm{diag}\{\mathrm{e}^{\mathrm{i}\pi\tilde{d}_1/2}\lambda^{-\tilde{d}_1},\ldots,\mathrm{e}^{\mathrm{i}\pi\tilde{d}_p/2}\lambda^{-\tilde{d}_p}\}$ and $G=\Theta\Theta'/2\pi$, where $\Theta=(\theta_{1l(1)},\ldots,\theta_{pl(p)})'$.

If X_t is generated by (10) or (11) it cannot be cointegrated under the above conditions on $\Phi(z),\Psi(z)$, but we now discuss cointegration of $I(d_1,\ldots,d_p)$ series X_t. Let $\beta=(\beta_1,\beta_2,\ldots,\beta_p)'$ be a p-dimensional vector, and under (12) write

$$\beta=(\beta(1)',\ldots,\beta(s)')', \tag{13}$$

where for $l=1,\ldots,s$, $\beta(l)=(\beta_{i_{l-1}+1},\ldots,\beta_{i_l})'$ is a $p_l=i_l-i_{l-1}$-dimensional vector, conforming to the partition of d. We correspondingly write $X_t=(X_t^{(1)'},\ldots,X_t^{(s)'})'$ where $X_t^{(l)}=(X_{i_{l-1}+1,t},\ldots,X_{i_l,t})'$, for $l=1,\ldots,s$.

Definition 2. If there exists a non-null vector $\beta(l)$ such that $\beta(l)'X_t^{(l)}$ is $I(d_u)$ with $d_u<d_{i_l}$, then we say X_t is cointegrated with cointegrating vector $\beta=(0',\ldots,0',\beta(l)',0',\ldots,0')'$. The number of such linearly independent $\beta(l)$ is r_l, and the cointegrating rank of X_t is $r=\sum_{l=1}^{p}r_l$.

Remark 2. (i) An individual cointegrating vector can be uniquely defined by a normalization. The r cointegrating errors $\beta'X_t$ can have different differencing parameters d_u. Our definition does not cover "polynomial cointegration", in which two or more $X_t^{(l)}$ are linked in a cointegrating relation.

(ii) If $d_1=\cdots=d_p=d$, and d and $d_u=d-b$ are integers, our definition is identical with $CI(d,b)$ given by Engle and Granger (1987), except that their $I(d)$ definition pertains only to FARIMA series (10).

(iii) The definition implies $p_l\geqslant2$, but this is not entailed if we adapt the definition of Johansen (1996) (whose treatment concerned only $I(d)$ processes

for integer d, and in an AR context). Johansen's definition would include as a cointegrating vector any β such that $\beta' X_t$ is $I(d_u)$ for $d_u < d_1$.

(iv) Flôres and Szafarz (1996) define X_t to be cointegrated if there exists β with $\beta(1) \neq 0$ such that $\beta' X_t$ is $I(d_u)$ with $d_u < d_1$. If Definition 2 provides no cointegrating vector of form $\beta = (\beta(1)', 0', \ldots, 0')'$, then X_t is not cointegrated by Flôres and Szafarz's definition.

(v) Robinson and Marinucci (1998) define X_t to be cointegrated if there exists $\beta \neq 0$ such that $\beta' X_t$ is $I(d_u)$ with $d_u < d_p$.

To illustrate the differing implications of the various definitions, consider the example

$$X_t = (u_t + \varepsilon_{1t},\ au_t + \varepsilon_{2t},\ v_t + \varepsilon_{3t},\ bv_t + \varepsilon_{4t},\ \varepsilon_{5t})', \tag{14}$$

where $a \neq 0,\ b \neq 0,\ u_t$ is $I(d),\ v_t$ is $I(e)$ with $0 < e < d$, and $\varepsilon_{it},\ i = 1, \ldots, 5$, are $I(0)$. We list the values of r and β provided by each of the four definitions:

Definition 2: $r = 2;\ \beta = (a, -1, 0, 0, 0)',\ (0, 0, b, -1, 0)'$.
Johansen: $r = 4;\ \beta = (a, -1, 0, 0, 0)',\ (0, 0, b, -1, 0)',\ (0, 0, 0, 1, 0)',\ (0, 0, 0, 0, 1)'$.
Flôres and Szafarz: $r = 1;\ \beta = (a, -1, 0, 0, 0)'$.
Robinson and Marinucci: $r = 0$.

On the other hand if the system consists of only the first two elements of (14) then we get $r = 1;\ \beta = (a, -1)'$ from all four definitions.

We introduce

Assumption A. X_t *is* $I(d_1, \ldots, d_p)$.

Theorem 1. Let Assumption A *hold and impose* (12). *Then if* $G_{(l)}$ *is the* $p_l \times p_l$ *matrix whose* (i, j)th *element is the* $(p_1 + \cdots + p_{l-1} + i, p_1 + \cdots + p_{l-1} + j)$th *element of* G (*meaning, when* $l = 1$, *the* (i, j)th *element*), *then*

$$r_l = p_l - \operatorname{rank}(G_{(l)}), \quad r = p - \sum_{l=1}^{s} \operatorname{rank}(G_{(l)}). \tag{15}$$

Remark 3. Under Definitions 1 and 2 and (12), r can be estimated by a two-stage procedure, to be discussed subsequently, namely first partitioning X_t into subvectors $X_t^{(l)},\ l = 1, \ldots, s$, for $1 \leqslant s \leqslant p$, such that (12) holds, and then estimating $r_l,\ l = 1, \ldots, s$. On the other hand, by applying a modified version of such a procedure repeatedly, we can also detect polynomial cointegration and cointegration covered by definitions of Flôres and Szafarz (1996), Robinson and Marinucci (1998), but not covered by Definition 2. If

Definition 2 has identified cointegrating vectors of form $(\beta_a(1)', 0', \ldots, 0')'$, $1 \leqslant a \leqslant q$, where $1 \leqslant q < i_1$, we may define the $(p - i_1 + q) \times 1$ vector Y_t to have ath element

$$Y_{at} = \begin{cases} \beta_a(1)' X_t^{(1)}, & 1 \leqslant a \leqslant q, \\ X_{a+i_1-q,\,t}, & q+1 \leqslant a \leqslant p - i_1 + q. \end{cases}$$

(Of course, in practice, it is only possible to statistically estimate q and the $\beta_a(1)$ from finitely many data.) We might then investigate Y_t for cointegration in the same way as we did X_t, then redefine Y_t, and continue in this fashion until all cointegrating vectors under these other definitions have been determined. However, we do not pursue this approach, but rather the one referred to at the start of the paragraph.

3. Testing the homogeneity of fractional difference parameters

We propose a specific-to-general type of procedure for partitioning X_t into subvectors $X_t^{(l)}$, $l = 1, \ldots, s$, with common differencing parameters, as in (12). A similar procedure is employed in other problems, for example the analysis of variance (see Marcus et al., 1976). Hsu (1996) provides a comprehensive survey of possible solutions. For $s = 1, \ldots, p$, let $(v_1, v_2, \ldots, v_s)$ be a partition of $(1, 2, \ldots, p)$ and define the hypothesis

$$\mathrm{H}_{v_1 v_2, \ldots, v_s}\colon \{d_a = d_b \text{ for } a, b \in v_l \text{ and } d_a \neq d_b \text{ for } a \in v_l, b \in v_{l'}, l \neq l', l, l' = 1, \ldots, s\}.$$

We first take $s = 1$ and test

$$\mathrm{H}_{v_1}\colon d_1 = d_2 = \cdots = d_p. \tag{16}$$

If H_{v_1} is not rejected, the procedure terminates. If H_{v_1} is rejected, we test each hypothesis $\mathrm{H}_{v_1 v_2}$ and terminate if at least one of these is not rejected. Otherwise, we proceed to test each hypothesis $\mathrm{H}_{v_1 v_2 v_3}$, and so on. If we eventually reach $s = p - 1$, and then reject each $\mathrm{H}_{v_1 v_2, \ldots, v_{p-1}}$, we can conclude there is no cointegration, and so in no case do we test $\mathrm{H}_{v_1 v_2, \ldots, v_p}$. Thus, we test at most $\sum_{s=1}^{p} S(p,s) - 1$ hypotheses, where $S(p,s) = 1/s! \sum_{i=0}^{s} (-1)^i \binom{s}{i} (s-i)^p$ is the Stirling number of the second kind (see Liu, 1968, pp. 38–39). On the other hand if we do not reject more than one of $\mathrm{H}_{v_1 v_2, \ldots, v_s}$, for some $s = 2, \ldots, p-1$, then we can either investigate each of the corresponding subvectors of X_t, or else select the one with the most insignificant test statistic.

We now consider the testing of $\mathrm{H}_{v_1 v_2, \ldots, v_s}$. Noting that $\mathrm{H}_{v_1 v_2, \ldots, v_s}$ is a joint hypothesis, composed of the hypotheses $\mathrm{H}_{\rho_l}, l = 1, \ldots, s$, where

$$\mathrm{H}_{\rho_l}\colon \{d_a \text{ are equal for } a \in v_l\},$$

we test each such H_{ρ_l} individually and reject $H_{v_1 v_2,\ldots,v_l}$ if and only if at least one is rejected . Of course, since we do not need to test H_{ρ_l} with $n_l = \#\{v_l\} = 1$, actually $H_{v_1 v_2,\ldots,v_s}$ is composed only of H_{ρ_l} such that $n_l > 1$.

A test of H_{ρ_l} can be based on $\hat{d}_a - \hat{d}_b$, for $a < b, a, b \in v_l$, where $\hat{d}_a, \hat{d}_b$ are estimates of d_a, d_b. Several consistent semiparametric estimates of differencing parameters have been proposed (see e.g. Geweke and Porter-Hudak, 1983; Künsch, 1987; Robinson, 1994, 1995a, b). We shall use the Gaussian semiparametric or local Whittle estimate of Künsch (1987), Robinson (1995b), because it has similarly nice asymptotic properties to the log periodogram estimate, but is more efficient.

Given observations $X_t, t = 1, 2, \ldots, n$, introduce the discrete Fourier transform

$$w_a(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_{at} e^{it\lambda}, \quad a = 1, \ldots, p,$$

where mean-correction is unnecessary because $w_a(\lambda)$ will be computed only at frequencies $\lambda_j = 2\pi j/n$ for $j = 1, \ldots, m < n/2$. The (cross-)periodogram of X_{at} and X_{bt} is

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

For brevity, we write $I_{ab,\,j}$ for $I_{ab}(\lambda_j)$. Define

$$\hat{d}_a = \arg \min_{\Delta \in [\Delta_1, \Delta_2]} R_a(\Delta), \tag{17}$$

where Δ_1 and Δ_2 satisfy $-1/2 < \Delta_1 < \Delta_2 < 1/2$, and

$$R_a(\Delta) = \log \left(\frac{1}{m} \sum_{j=1}^{m} j^{2\Delta} I_{aa,\,j} \right) - 2\Delta \frac{1}{m} \sum_{j=1}^{m} \log j.$$

We introduce the following further assumptions.

Assumption B. Defining $A(L) = \sum_{j=0}^{\infty} A_j L^j$, *where the* A_j *are* $p \times p$ *matrices*,

$$X_t = \mu + A(L) e_t, \quad t = 0, \pm 1, \ldots,$$

where, almost surely, $E(e_t | \mathscr{F}_{t-1}) = 0, E(e e_t' | \mathscr{F}_{t-1}) = I_p$, the $p \times p$ *identity matrix, and the matrices* $E(e_t \otimes e_t e_t' | \mathscr{F}_{t-1})$, $E(e_t e_t \otimes e_t e_t' | \mathscr{F}_{t-1})$ *are finite, nonstochastic and constant in* t, $\mathscr{F}_t$ *being the* σ*-field of events generated by* $e_s, s \leqslant t$.

Assumption C. Writing $A_{ab}(L)$ *for the* (a, b)th *element of* $A(L)$,

$$\frac{d}{d\lambda} A_{ab}(e^{i\lambda}) = O(\{|A_{aa}(e^{i\lambda})||A_{bb}(e^{i\lambda})|\}^{1/2}/\lambda) \quad \text{as } \lambda \to 0+,$$

for $a, b = 1, \ldots, p$.

Assumption D. With $C(L)=\mathrm{E}(L)A(L)$ *as in* (6), (7), *we have for* $\xi\in(0,2]$,

$$C(\mathrm{e}^{\mathrm{i}\lambda})=C(1)(1+\mathrm{O}(\lambda^{\xi}))\quad \text{as } \lambda\to 0+.$$

Assumption E. For the same ξ *as in Assumption D*,

$$\frac{1}{m}+\frac{m^{1+2\xi}(\log m)^2}{n^{2\xi}}\to 0\quad \text{as } n\to\infty.$$

Assumption F.

$$d_a\in(\Delta_1,\Delta_2), a=1,\ldots,p.$$

Assumptions B, C and E are the same as Assumptions A2, A3 and A4 of Lobato (1999), respectively, and are analogous to Assumptions A3′, A2′ and A4′ of Robinson (1995b) for scalar X_t. Assumption D implies that the error in approximating the left-hand side of (3) by the right is $\mathrm{O}(\lambda^{\xi})$, which is a similar assumption to A1′ of Robinson (1995b) and A1 of Lobato (1999). The case $\xi=2$ applies to FARIMA models, for example, and is standard in other circumstances of smoothed nonparametric estimation, such as probability and spectral density estimation. The interior-point Assumption F is standard in central limit theory for implicitly defined extremum estimates.

For the sequel we redefine d as the $p\times 1$ vector $d=(d_1,\ldots,d_p)'$. Define also $\hat{d}=(\hat{d}_1,\ldots,\hat{d}_p)'$ and $D=\mathrm{diag}\{G_{11},\ldots,G_{pp}\}$ where G_{ab} is the (a,b)th element of G, and denote Hadamard product by $\circ$ and transposition combined with complex conjugation by *.

Proposition 1. Under Assumptions A–F,

$$m^{1/2}(\hat{d}-d)\to_{\mathrm{d}}\mathrm{N}\left(0,\tfrac{1}{4}D^{-1}(G\circ G)D^{-1}\right)\quad \text{as } n\to\infty.$$

A consistent estimate of G *is*

$$\hat{G}=\frac{1}{m}\sum_{j=1}^{m}\mathrm{Re}\{\hat{\Lambda}(\lambda_j)^{-1}I_j\hat{\Lambda}(\lambda_j)^{-1*}\},\tag{18}$$

where $I_j=I(\lambda_j)$, $I(\lambda)=w(\lambda)w(\lambda)^*$, $w(\lambda)=(w_1(\lambda),\ldots,w_p(\lambda))'$, *and* $\hat{\Lambda}(\lambda)$ *is* $\Lambda(\lambda)$ *with* d *replaced by* $\hat{d}$.

We apply this result to a single test of pairwise equality

$$\mathrm{H}_{ab}: d_a=d_b.\tag{19}$$

Denote by G_{ab}, $\hat{G}_{ab}$ the (a,b)th elements of $G,\hat{G}$, respectively. Noting that $G_{aa}G_{bb}-G_{ab}^2$ is the determinant of the matrix formed by omitting from G all rows and columns but the ath and bth, if X_{at} and X_{bt} are cointegrated

$G_{ab}^2 = G_{aa}G_{bb}$, while if they are not, $G_{ab}^2 < G_{aa}G_{bb}$. It thus follows from Proposition 1 and the delta method that under H_{ab} the limit distribution of

$$\tilde{T}_{ab} = m^{1/2}(\hat{d}_a - \hat{d}_b)/\{\tfrac{1}{2}(1 - \hat{G}_{ab}^2/(\hat{G}_{aa}\hat{G}_{bb}))\}^{1/2}$$

is standard normal if X_{at} and X_{bt} are not cointegrated, but is not well defined if they are. Since the presence or absence of cointegration is not known at the time of testing H_{ab}, we require a test statistic that is informative under both circumstances. Define

$$\hat{T}_{ab} = \frac{m^{1/2}(\hat{d}_a - \hat{d}_b)}{\{\tfrac{1}{2}(1 - \hat{G}_{ab}^2/(\hat{G}_{aa}\hat{G}_{bb}))\}^{1/2} + h(n)},$$

where $h(n) > 0$. We introduce

Assumption G.

$$h(n) + \frac{(\log m)m^{1/2+\xi}/n^{\xi} + (\log m)^2 m^{-1/6}}{h(n)} \to 0 \quad \text{as } n \to \infty.$$

Theorem 2. Let Assumptions A–G *hold. Then under* H_{ab} (19), *as* $n \to \infty$:

(i) *If* X_{at} *and* X_{bt} *are not cointegrated,* $\hat{T}_{ab} \to_{\mathrm{d}} \mathrm{N}(0,1)$;
(ii) *If* X_{at} *and* X_{bt} *are cointegrated,* $\hat{T}_{ab} \to_{\mathrm{p}} 0$.

Of course, $\hat{T}_{ab}$ is potentially highly sensitive to choice of $h(n)$, but a significantly large value of $|\hat{T}_{ab}|$, relative to the N(0, 1) distribution, can be taken as evidence against H_{ab}, irrespective of whether or not there is cointegration, whereas a decision of non-rejection under $h(n) = 0$ (i.e. based on $\tilde{T}_{ab}$) would be made with greater confidence for any positive $h(n)$.

To test H_{ρ_l} we can consider

$$\hat{T}_{\rho_l} = \max_{a,\, b \in v_l} |\hat{T}_{ab}|.$$

The level of the test of each H_{ρ_l} must provide a level-α test of $H_{v_1 v_2,\ldots,v_s}$. Unlike in classical analysis of variance, the test statistics $\hat{T}_{\rho_l}$ are not mutually independent. For $s' = \#\{l\colon n_l > 1, l = 1,\ldots,s\}$, a level-$(1 - (1-\alpha)^{1/s'})$ test based on $\hat{T}_{\rho_l}$ for each H_{ρ_l} does thus not necessarily assure a level-α test for $H_{v_1 v_2,\ldots,v_s}$. However, by the Bonferroni inequality, if a level-α/s' test for H_{ρ_l} with $n_l > 1$ is given by $\hat{T}_{\rho_l}$, which in turn is given by a level-$2\alpha/(s' n_l(n_l - 1))$ test for H_{ab} based on the $n_l(n_l-1)/2$ statistics $|\hat{T}_{ab}|$, $a < b$, $a, b \in v_l$, it assures a level-α test for $H_{v_1 v_2,\ldots,v_s}$. Specifically, if we reject $H_{v_1 v_2,\ldots,v_s}$ when at least one of $|\hat{T}_{ab}|$, $a < b$, $a, b \in v_l$, $l = 1,\ldots,s$, exceeds $z_{\alpha/(s'(n_l(n_l-1))}$ where z_c is the $1-c$ quantile of N(0, 1), we achieve, asymptotically, a level-α test for $H_{v_1 v_2,\ldots,v_s}$.

An alternative approach, which avoids a user-chosen sequence such as $h(n)$, involves applying in the same way

$$\bar{T}_{ab} = (2m)^{1/2}(\hat{d}_a - \hat{d}_b), \quad \bar{T}_{\rho_l} = \max_{a,\, b \in v_l} |\bar{T}_{ab}|,$$

which is more conservative than $\hat{T}_{\rho_l}$ because the asymptotic variance of $\bar{T}_{ab}$ is generally less than one.

According to the above pairwise approach, $n_l(n_l - 1)/2$ separate tests are involved in the testing of H_{ρ_l} alone. These ideas can be extended, when $n_l \geqslant 3$, to permit a single test of H_{ρ_l}. We can write H_{ρ_l} as $Sd = 0$, where, with $v_l = (a_1, \ldots, a_{n_l})$, S is the $(n_l - 1) \times p$ matrix whose jth row has 1 as its a_jth element and -1 as its a_{j+1}th element and the remainder 0, $j = 1, \ldots, n_l - 1$. Denote $\hat{D} = \mathrm{diag}\{\hat{G}_{11}, \ldots, \hat{G}_{pp}\}$. Then, for example, in view of Proposition 1, under H_{ρ_l} the statistic

$$(S\hat{d})' \left(S \tfrac{1}{4} \hat{D}^{-1} (\hat{G} \circ \hat{G}) \hat{D}^{-1} S' + h(n)^2 I_{n_l - 1} \right)^{-1} (S\hat{d})$$

has a limiting $\chi^2_{n_l - 1}$ distribution when there is no cointegration, and converges in probability to zero otherwise.

4. Determination of fractional cointegrating rank

We now suppose that X_t has already been partitioned into subvectors $X_t^{(l)}$, $l = 1, \ldots, s$, satisfying (12), perhaps by applying the procedures of the previous section. We shall only attempt to estimate the cointegrating ranks of such $X_t^{(l)}$ individually, in which case for notational convenience we can take $s = 1$, so that (16) holds, and consider the cointegrating rank of X_t itself. We denote by d_* the common value of $d_1, \ldots, d_p$. In view of (3) and Theorem 1 we commence by obtaining estimates of G and its eigenvalues, and determine their limit distribution. Consider (cf. (18))

$$\hat{G}(d_*) = \frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d_*} \mathrm{Re}(I_j). \tag{20}$$

Note that $\pi \hat{G}(0)$ was used by Phillips and Ouliaris (1988) in testing the cointegrating rank of an $I(1)$ vector. Let G_a be the ath column of G.

Proposition 2. Let Assumptions A–G *hold. Then as* $n \to \infty$

$$m^{1/2} vec(\hat{G}(d_*) - G) \to_{\mathrm{d}} \mathrm{N}\left(0, \tfrac{1}{2}(G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p))\right). \tag{21}$$

Since d_* is unknown this result is not of direct use. We might think of estimating d_* by versions of the multivariate log periodogram method of

Robinson (1995a) or multivariate local Whittle method (cf. (17)) of Lobato (1995, 1999) that gain efficiency by exploiting the fact that the elements of X_t have the same differencing parameters; denoting such an estimate $\hat{d}_*$, we might then estimate G by $\hat{G}(\hat{d}_*)$. There are two problems with this approach. First, if $\hat{d}_*$ uses the same bandwidth m as in (20) then, as found by Robinson (1995a), $\hat{G}(\hat{d}_*)$ does not have the limit distribution of (21), but rather, though it is normal, it is undesirably asymptotically perfectly correlated with $\hat{d}_*$ and indeed is only $m^{1/2}/\log n$-consistent. Second, these results of Robinson (1995a), Lobato (1995, 1999) assume G has full rank and are thus not valid if X_t is cointegrated. We deal with both difficulties by instead pooling estimates of d_* based on the individual elements of X_t (which make no presumption about cointegration) and computing the latter with a bandwidth m_1 that increases sufficiently faster than m that the effect of estimating d_* has no effect on (21). Denote by $\tilde{d}_a$ the estimate d_a given by (17) but with m replaced by m_1 and define

$$\bar{d}_* = \frac{1}{p}\sum_{a=1}^{p} \tilde{d}_a.$$

Assumption H. For any $\xi > 0$

$$\frac{m^{1/2-\zeta}n^{\zeta}}{m_1^{1/2}} + \frac{m_1^{1+2\xi}(\log m_1)^2}{n^{2\xi}} \to 0, \quad \text{as } n \to \infty.$$

Proposition 3. Let Assumptions A–F *and* H *hold. Then as* $n \to \infty$,

$$m^{1/2}vec(\hat{G}(\bar{d}_*) - G) \to_{\mathrm{d}} \mathrm{N}\left(0, \frac{1}{2}(G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p))\right).$$

Assumption I. $rank(G) = p - r$, *for* $0 \leqslant r < p$, *and the nonzero eigenvalues of* G *are distinct.*

Let $\delta_a(\hat{\delta}_a)$ be the ath eigenvalue of G $(\hat{G}(\bar{d}_*)), a = 1, \ldots, p$, ordered such that $\delta_1 > \delta_2 > \cdots > \delta_{p-r} > 0$, with $\delta_{p-r+1} = \cdots = \delta_p = 0$, for $r \geqslant 1$, and $\hat{\delta}_1 \geqslant \hat{\delta}_2 > \cdots \geqslant \hat{\delta}_p$.

Proposition 4. Let Assumptions A–F, H *and* I *hold. Then the* $m^{1/2}(\hat{\delta}_a - \delta_a)$ *are asymptotically independent for* $a = 1, \ldots, p$, *converge in distribution to* $\mathrm{N}(0, \delta_a^2)$ *variates for* $a = 1, \ldots, p - r$, *and are* $\mathrm{o_p}(1)$ *for* $a = p - r + 1, \ldots, p$.

Proposition 4 can be interpreted as a variant of Theorem 9.4.4 of Brillinger (1975) as m and n tend to infinity simultaneously, and suggests that

Theorem 1 of Anderson (1963) and Theorem 13.5.1 of Anderson (1984) are still true in the singular case.

Now define, for $j = 1, \ldots, p-1$,

$$\pi_j = \frac{\sigma^{(1)}_{p-j+1,\,p}}{\sigma^{(1)}_{1,\,p}}, \quad \hat{\pi}_j = \frac{\hat{\sigma}^{(1)}_{p-j+1,\,p}}{\hat{\sigma}^{(1)}_{1,\,p}},$$

where

$$\sigma^{(i)}_{k,\,l} = \sum_{a=k}^{l} \delta_a^i, \quad \hat{\sigma}^{(i)}_{k,\,l} = \sum_{a=k}^{l} \hat{\delta}_a^i.$$

Also define

$$s_j = \frac{\left(\hat{\sigma}^{(1)2}_{p-j+1,\,p}\hat{\sigma}^{(2)}_{1,\,p-j} + \hat{\sigma}^{(1)2}_{1,\,p-j}\hat{\sigma}^{(2)2}_{p-j+1,\,p}\right)^{1/2}}{\hat{\sigma}^{(1)2}_{1,\,p}}.$$

Theorem 3. Let Assumptions A–F, H *and* I *hold, and let* $r = 0$. *Then for* $j = 1, \ldots, p-1$

$$m^{1/2}(\hat{\pi}_j - \pi_j)/s_j \to_{\mathrm{d}} \mathrm{N}(0,1) \quad \textit{as } n \to \infty.$$

The application of Theorem 3 in determining r by hypothesis testing is hampered by the assumption that $r = 0$. We propose two rather ad hoc solutions, both of which might be applied for increasing values of r:

(i) This directly follows the proposal of Phillips and Ouliaris (1988) for the case of $CI(1,1)$ cointegration. To test that the cointegrating rank is r we consider the $100(1-\alpha)\%$ upper confidence interval for π_r based on Theorem 3, namely

$$\hat{\pi}_r + s_r z_\alpha / m^{1/2}. \tag{22}$$

We find evidence in favour of the hypothesis that the cointegrating rank is r if this is smaller than some prescribed threshold, such as $0.1/p$ (as suggested by Phillips and Ouliaris, 1988).

(ii) We can apply Theorem 3 to test that π_r is some sufficiently small positive value γ, e.g. $\gamma = 0.01/p$. Then if we reject this hypothesis in favour of the alternative $\pi_r > \gamma$ we find evidence against a cointegrating rank of r. The specification of a small null value of π_r might seem more attractive than the specification of a threshold for (22), but one expects that the central limit theory of Theorem 3 for $\hat{\delta}_a$ may provide a poor approximation as $\delta_a \to 0$.

Alternatively, we may consider a model selection procedure which consistently estimates r (cf. Fujikoshi, 1985; Fujikoshi and Veitch, 1979; Gunderson and Muirhead, 1997). Define, for $v(n) > 0$,

$$L(u) = v(n)(p-u) - \hat{\sigma}^{(1)}_{1,\,p-u} \tag{23}$$

and estimate r by

$$\hat{r} = \arg \min_{u=1,\ldots,p-1} L(u).$$

Assumption J.

$$v(n) + \frac{1}{m^{1/2} v(n)} \to 0 \quad \text{as } n \to \infty.$$

Theorem 4. Let Assumptions A–F *and* H–J *hold. Then*

$$\lim_{n\to\infty} P(\hat{r} = r) = 1.$$

If elements of X_t are measured in different units, scale-invariant statistics might seem preferable, as suggested in the $CI(1,1)$ case by Phillips and Ouliaris (1988). For example, we might wish to base the analysis not on $\hat{G}(\bar{d}_*)$ but rather on the correlation matrix $\hat{P}(\bar{d}_*) = \hat{D}(\bar{d}_*)^{-1/2} \hat{G}(\bar{d}_*) \hat{D}(\bar{d}_*)^{-1/2}$, where $\hat{D}(\bar{d}_*)$ is the diagonal matrix whose ath diagonal element is the same as that of $\hat{G}(\bar{d}_*)$. Unfortunately, the limiting covariance structure of the eigenvalues of $\hat{P}(\bar{d}_*)$ is much more complicated than those of $\hat{G}(\bar{d}_*)$, and so testing procedures of corresponding simplicity to those derived from Theorem 3 are not available. However, because its probability limit has the same rank as $\hat{G}(\bar{d}_*)$, $\hat{P}(\bar{d}_*)$ can be used in a model choice procedure analogous to that justified in Theorem 4.

5. Empirical example

We apply the procedure developed in the preceding sections to a trivariate series of 146 observations on spot closing prices of crude oil, namely West Texas Intermediate (WTI), Brent, and Dubai (which are said to be key markers in the US, European and Asian markets, respectively), recorded on the last trading day of each month from January 1986 through February 1998. (The price of WTI for October 1991 was not observed due to a pipeline accident and is replaced by the mean of the September and November 1991 observations.) We analyse log prices, taking $p = 3$ and $X_{1t} = \log WTI_t$, $X_{2t} = \log Brent_t$ and $X_{3t} = \log Dubai_t$ in the notation of the paper.

Since unit root analysis in an AR setting is standard in econometrics, we commenced in this fashion, using the model

$$X_{at} = c + \rho_a X_{a,t-1} + \sum_{i=1}^{N_a - 1} \phi_{ai} (X_{a,t-i} - X_{a,t-1-i}) + \varepsilon_{at}, \quad a = 1,2,3.$$

We applied the augmented Dickey–Fuller test (see e.g. Dickey and Fuller, 1981; Said and Dickey, 1984) because the series (see Fig. 1) appear to have

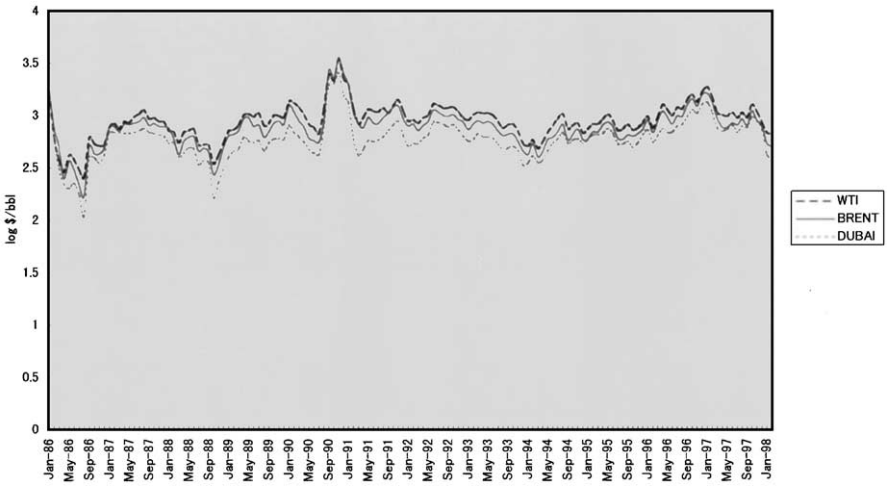

Fig. 1. Log prices of crude oil.

Table 1
ADF test

Series	t-ratio	5% critical value	1% critical value
WTI	−3.489	−2.883	−3.480
Brent	−3.019	−2.883	−3.479
Dubai	−4.691	−2.882	−3.477

nonzero mean but reveal no distinctive seasonal pattern or trend except for the upturn around the period of the Gulf War from late 1990 to early 1991. Choosing the AR order N_a to attain the minimum AIC between 0 and 15, we obtained $N_1 = 11$, $N_2 = 9$ and $N_3 = 5$. Table 1 shows t-ratios for $\rho_a - 1$ with 5% and 1% critical values, based on Table 1 of MacKinnon (1991), allowing for a constant but no trend. We rejected the unit root null hypothesis at 1% for WTI and Dubai and at 5% for Brent.

The AR orders chosen by AIC are on the large side, particularly for WTI and Brent, and so stationary long memory is an alternative possibility. Table 2 presents the estimates $\hat{d}_a$, $a = 1, 2, 3$ of (17), using $m = 20$ and 15, along with standard errors, obtained from the asymptotic variance formula $1/4m$ for univariate estimates (cf. Proposition 1). With the exception of WTI with $m = 20$, all estimates are less than 0.5. We focus on the estimates with $m = 15$, which are consistent with stationarity.

In Table 3, we report $\hat{G}$ (18) with $m = 13$, thereby deducing the test statistics $\tilde{T}_{12} = 2.626$, $\tilde{T}_{13} = 2.768$ and $\tilde{T}_{23} = 1.273$. Thus, $H_{123}: d_1 = d_2 = d_3$ is not

Table 2
Estimates $\hat{d}_a$ of d_a, $a = 1, 2, 3$, with standard errors in parentheses

Series	$m = 20$	$m = 15$
WTI	0.5336 (0.112)	0.4714 (0.129)
Brent	0.4538 (0.112)	0.3748 (0.129)
Dubai	0.4367 (0.112)	0.3076 (0.129)

Table 3
Estimates $\widehat{G}$ of G, $m = 13$

Series	WTI	Brent	Dubai
WTI	0.0037456	0.0046078	0.0055116
Brent	0.0046078	0.0064323	0.0073090
Dubai	0.0055116	0.0073090	0.0090621

Table 4
Estimates $\hat{G}(\bar{d}_*)$ of G based on $\bar{d}_* = (\hat{d}_1 + \hat{d}_2 + \hat{d}_3)/3$, and eigenvalues of $\hat{G}(\bar{d}_*)$, $\hat{P}(\bar{d}_*)$; $m = 13$, $m_1 = 15$

Series	WTI	Brent	Dubai
WTI	0.0049261	0.0054196	0.0057460
Brent	0.0054196	0.0062486	0.0065317
Dubai	0.0057460	0.0065317	0.0072951
Eigenvalues	1	2	3
$G(\times 10^{-2})$	1.80704	0.02750	0.01244
P	2.93521	0.04298	0.02182

rejected at 1%, and would be more strongly supported by $\hat{T}_{12}, \hat{T}_{13}$ and $\hat{T}_{23}$ with $h(m) > 0$, while $H_{23}: d_2 = d_3$ is not rejected at level 5%.

In consequence, we investigate the presence of cointegration on the basis of both $s = 1$ and 2, where in the latter case Brent and Dubai are supposed to have a common differencing parameter. Table 4 considers the case $s = 1$, showing $\hat{G}(\bar{d}_*)$ and its eigenvalues and those of $\hat{P}(\bar{d}_*)$, while Table 5 considers $s = 2$, showing the 2×2 matrix $\hat{G}_{(1)}(\bar{d}_*)$ and its eigenvalues and those of $\hat{P}_{(1)}(\bar{d}_*)$, $\hat{G}_{(1)}(\bar{d}_*)$ estimating $G_{(1)}$ and $\hat{P}_{(1)}(\bar{d}_*)$ thence defined analogously to $\hat{P}(\bar{d}_*)$; in each case $m = 13$, $m_1 = 15$. The largest eigenvalue greatly dominates throughout, so that for any reasonable value of δ the objective function $L(u)$ will support the conclusion $r = 2$ when $s = 1$, and $r = 1$ when $s = 2$.

Table 5
Estimates $\hat{G}_{(1)}(\bar{d}_*)$ of $G_{(1)}$ (Brent and Dubai only) based on $\bar{d}_* = (\hat{d}_1 + \hat{d}_2)/2$, and eigenvalues of $\hat{G}_{(1)}(\bar{d}_*)$, $\hat{P}_{(1)}(\bar{d}_*)$; $m = 13$, $m_1 = 15$

Series	Brent	Dubai
Brent	0.0071133	0.0073984
Dubai	0.0073984	0.0082350
Eigenvalues	1	2
$G(\times 10^{-2})$	1.50938	0.02545
P	1.96665	0.03335

Also, with $r = 2$ and $\alpha = 0.05$, (22) is $0.0338 \backsimeq 1/30 = 0.10/p$. This is the threshold suggested by Phillips and Ouliaris (1988) so that the conclusion concerning $r = 2$ is rather uncertain. Since H_{23} seemed more strongly supported than H_{123} we might thus prefer $r = 1$, for which Table 5 provides support.

6. Final comments

1. The procedures suggested all involve choice of a bandwidth m (and also m_1 involved in $\bar{d}_*$). User-chosen bandwidth numbers are inevitable in smoothed nonparametric estimation, and as usual the results will be sensitive to the choice made. Some proposals for choosing m in (17) are made by Henry and Robinson (1996).

2. Perhaps more seriously, the procedures also depend on other user-chosen numbers, namely $h(n), v(n)$ and the threshold for (22) (or the null hypothesis on π_r). Our introduction of these is an indication of the difficulty of the problems tackled, but clearly it would be desirable to develop more objective methods.

3. There is interest in extending the methods to cover nonstationary X_t. The extension of Definition 2 here is straightforward. If, on the other hand, we have sufficient prior knowledge to first-difference the raw series to the stationary/invertible region $(-1/2, 1/2)$, then our results may be applicable, though this may depend also on the differenced cointegrating errors having differencing parameters that lie in this region. If we then go on to estimate β it is important that the undifferenced series be used here in order to achieve a fast rate of convergence (see Robinson and Marinucci,1998). So far as applicability of our present procedures to raw nonstationary X_t is concerned, Velasco (1999) showed that in case of scalar series, at least suitably tapered

versions of estimates (17) can still be $m^{1/2}$-consistent and asymptotically normal. Thus, it seems that suitable modifications of our procedures may directly apply to nonstationary X_t, though it remains to provide rigorous justification.

Acknowledgements

The authors are especially grateful for help with data analysis provided by Professor K. Nawata and Mr. T. Okimoto of the University of Tokyo and by Dr. Y. Matsuda of Niigata University. They thank two referees for a careful reading of the paper and helpful comments. They also thank Dr. L. Giraitis of the London School of Economics for helpful comments and Dr. S. Kuriki of the Institute of Statistical Mathematics for advice on the closed testing method. Robinson's research was supported by a Leverhulme Trust Personal Research Professorship and ESRC Grants R000235892 and R000238212. Yajima's research was supported by the Ministry of Education of Japan, reference number 09CE2002.

Appendix A. Proofs

Proof of Theorem 1. From (3) the spectral density matrix $f_{\beta_l}(\lambda)$ of $\beta'(l)X_t^{(l)}$ satisfies

$$f_{\beta_l}(\lambda) \backsim \lambda^{-2d_l}\beta'(l)G_{(l)}\beta(l) \quad \text{as } \lambda \to 0+.$$

Hence, $\beta(l)$ is a cointegrating vector for $X_t^{(l)}$ if and only if $\beta'(l)G_{(l)}\beta(l) = 0$. Then (15) follows immediately. □

Proof of Proposition 1. We only briefly summarize the proof since it follows like that of Theorem 2 of Robinson (1995b). We write $R_a^{(i)}(\Delta) = \partial^i R_a(\Delta)/\partial\Delta^i$. With probability approaching 1, as $n \to \infty$, $\hat{d}_a$ satisfies

$$0 = R_a^{(1)}(\hat{d}_a) = R_a^{(1)}(d_a) + R_a^{(2)}(\bar{d}_a^*)(\hat{d}_a - d_a),$$

where $|\bar{d}_a^* - d_a| \leqslant |\hat{d}_a - d_a|$. Now

$$m^{1/2}(R_1^{(1)}(d_1), \ldots, R_p^{(1)}(d_p))' \to_{\mathrm{d}} \mathrm{N}(0, 4D^{-1/2}(G \circ G)D^{-1/2})$$

and $R_a^{(2)}(\bar{d}_a^*) \to_{\mathrm{p}} 4$, whence the result follows immediately. □

Proof of Theorem 2. The proof of (i) follows quickly from Proposition 1 and $h(n) \to 0$. To prove (ii) it suffices, from Assumption G, to show that $\hat{d}_a - \hat{d}_b = \mathrm{O_p}((\log m)(m/n)^{\xi} + (\log m)^2 m^{-2/3})$. Note first that under H_{ab}, with

$d_a = d_b = d_*$, the mean value theorem gives, with probability approaching 1 as $n \to \infty$,

$$\hat{d}_a - \hat{d}_b = \left(\frac{R_b^{(1)}(d_*)}{R_b^{(2)}(d_*)} - \frac{R_a^{(1)}(d_*)}{R_a^{(2)}(d_*)} \right) - \frac{1}{2} \left(\frac{R_a^{(3)}(\bar{d}_a)}{R_a^{(2)}(d_*)} (\hat{d}_a - d_*)^2 - \frac{R_b^{(3)}(\bar{d}_b)}{R_b^{(2)}(d_*)} (\hat{d}_b - d_*)^2 \right),$$

where $|\bar{d}_a - d_*| \leqslant |\hat{d}_a - d_*|$, $|\bar{d}_b - d_*| \leqslant |\hat{d}_b - d_*|$. The second term on the right is $O_p(m^{-1})$, by the same argument as that of Theorem 2 of Robinson (1995b) and, for all $a, R_a^{(2)}(d_*) \to_p 4$, $R_a^{(3)}(\bar{d}_*) \to_p -16$ and $\hat{d}_a - d_* = O_p(m^{-1/2})$. The first term is

$$\frac{R_b^{(1)}(d_*) - R_a^{(1)}(d_*)}{R_a^{(2)}(d_*)} + R_b^{(1)}(d_*) \left(\frac{1}{R_b^{(2)}(d_*)} - \frac{1}{R_a^{(2)}(d_*)} \right).$$

As found by Robinson (1995b), $R_a^{(i)}(d_*)$ is scale-free, so bearing in mind also that $G_{ab}^2 = G_{aa} G_{bb}$, we may take

$$G_{aa} = |G_{ab}| = G_{bb} = 1. \tag{24}$$

Also as found by Robinson (1995b), $R_a^{(i)}(d_*)$ is a differentiable function of

$$Y_a^{(l)} = \frac{1}{m} \sum_{j=1}^{m} (\log j)^l I_{aa,j} \lambda_j^{2d}, \quad l = 0, 1, \ldots, i.$$

As shown by Robinson (1995b) and Lobato (1999, Appendix C), for $l \geqslant 0$, $Y_a^{(l)}$ differs from

$$Z_a^{(l)} = \frac{1}{m} \sum_{j=1}^{m} (\log j)^l C_a' J_j C_a \tag{25}$$

by $O((\log m)^l (m/n)^{\xi} + (\log m)^{l+1} m^{-2/3})$, $l = 0, 1, \ldots, i$, where C_a' is the ath row of $C(1)$ and J_j is defined like I_j but with e_t replacing X_t. Thus, consideration of $R_a^{(i)}(d_*) - R_b^{(i)}(d_*)$ reduces to consideration of $Z_a^{(l)} - Z_b^{(l)}$, $l = 0, \ldots, i$. Now note from (9) that for all $a, b, C_a' C_b = 2\pi G_{ab}$. Then from (26) it follows that $||C_a||^2 = ||C_b||^2 = |C_a' C_b|$, with $||\cdot||$ denoting Euclidean norm, to imply, by the Schwarz inequality, that $C_a = \pm C_b$, so that $Z_a^{(l)} - Z_b^{(l)}$ is identically zero. Thus, the first term in (25) is $O_p((\log m)(m/n)^{\xi} + (\log m)^2 m^{-2/3})$. Since $R_b^{(1)}(d_*) = O_p(m^{-1/2})$, the second term in (25) is $O_p((\log m)^2 (m/n)^{\xi} m^{-1/2} + (\log m)^2 m^{-7/6})$, and the proof is completed. □

Proof of Proposition 2. We give only partial details as much of the proof is similar to that of Theorem 2 of Robinson (1995b) and the Theorem of Lobato (1999). We have

$$m^{1/2}\{\hat{G}(d_*) - G\} = \frac{1}{m^{1/2}}\sum_{j=1}^{m} \lambda_j^{2d_*}(\mathrm{Re}(I_j) - G) = \eta + \zeta,$$

where

$$\eta = m^{-1/2}\sum_{j=1}^{m} \{\lambda_j^{2d_*}(\mathrm{Re}(I_j - A_j J_j A_j^*)) + (\lambda_j^{2d} f(\lambda_j) - G)\},$$

$$\zeta = m^{-1/2}\sum_{j=1}^{m} \mathrm{Re}(B_j(J_j - I_p/2\pi)B_j^*) = \frac{m^{-1/2}}{2}\sum_{j}{}' B_j(J_j - I_p/2\pi)B_j^*,$$

where $\sum_j' = \sum_{j=1}^{m} + \sum_{j=n-m}^{n-1}$ and we write $A_j = A(\mathrm{e}^{i\lambda_j})$, $B_j = \lambda_j^d A_j, 1 \leqslant j \leqslant m$, $B_j = \lambda_{n-j}^d A_j, n - m \leqslant j \leqslant n - 1$, and use the fact that $\lambda_j^{2d_*} f(\lambda_j) = B_j B_j^*/2\pi$, $1 \leqslant j \leqslant m$.

Appendices C and D of Lobato (1999) assure under Assumptions A, C and D that $\eta = \mathrm{o_p}(1)$. Now $\zeta = \zeta_1 + \zeta_2$, where

$$\zeta_1 = \frac{m^{-1/2}}{2}\sum_{j}{}' B_j\left(\frac{1}{2\pi n}\sum_{t=1}^{n}(e_t e_t' - I_p)\right)B_j^*,$$

$$\zeta_2 = \frac{m^{-1/2}}{2}\sum_{j}{}' B_j\left(\frac{1}{2\pi n}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n} e_t e_s' \mathrm{e}^{-\mathrm{i}(t-s)\lambda_j}\right)B_j^*.$$

Because $\sum_{t=1}^{n}(e_t e_t' - I_p) = \mathrm{O_p}(n^{1/2})$, we have $\zeta_1 = \mathrm{O_p}((m/n)^{1/2}) = \mathrm{o_p}(1)$. By the Cramer–Wold device it remains to show that for any non-null $p^2 \times 1$ vector τ,

$$\tau' vec(\zeta_2) \to_{\mathrm{d}} \mathrm{N}\left(0, \tfrac{1}{2}\tau'\{G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p\}\tau\right).$$

Now, writing $\mathrm{H}_j = \bar{B}_j \otimes B_j$, $vec(\zeta_2)$ is $m^{-1/2}/4\pi n$ times

$$\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\sum_{j}{}' \mathrm{H}_j \mathrm{e}^{\mathrm{i}(s-t)\lambda_j}(e_s \otimes e_t)$$

$$= \sum_{t=1}^{n}\sum_{s=1,s<t}^{n}(C_{st}(e_s \otimes e_t) + C_{ts}(e_t \otimes e_s))$$

$$= \sum_{t=1}^{n}\left\{\sum_{s=1,s<t}^{n}(C_{st} + C_{ts}P)(e_s \otimes I_p)\right\}e_t,$$

where $C_{st} = \sum_j' \mathrm{H}_j \mathrm{e}^{\mathrm{i}(s-t)\lambda_j}$ and P is the $p^2 \times p^2$ permutation matrix such that $x \otimes y = P(y \otimes x)$ for all $p \times 1$ vectors x, y. Thus, $\tau' vec(\zeta_2) = \sum_{t=1}^n Z_t$, where, with $U_t = (m^{-1/2}/4\pi n)\tau'(\sum_{s=1, s<t}^n (C_{st} + C_{ts}P)(e_s \otimes I_p))$, $Z_t = U_t e_t$ is a martingale difference with respect to $\mathscr{F}_t$. Put $V_n^2 = \sum_{t=1}^n \mathrm{E}\{Z_t^2 | \mathscr{F}_{t-1}\}$, where $\mathscr{F}_0$ is the trivial σ-field $\{\phi, \Omega\}$ and $s_n^2 = \sum_{t=1}^n \mathrm{E}(Z_t)^2$. By the same argument as in the proof of Theorem 2 of Robinson (1995b), and in Appendix A of Lobato (1999), as $n \to \infty V_n^2 s_n^{-2} \to_{\mathrm{p}} 1$ and, for any $\varepsilon > 0, s_n^{-2} \sum_{t=1}^n \mathrm{E}\{Z_t^2 1(|Z_t| \geqslant \varepsilon s_n)\}$ $\to 0$. Hence by Theorem 2 of Brown (1971), $\sum_{t=1}^n Z_t / s_n \to_{\mathrm{d}} \mathrm{N}(0,1)$. It thus remains to show that $s_n \to \frac{1}{2}\tau'\{G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p)\}\tau$,which is easily seen to reduce to showing that

$$\frac{1}{mn^2} \sum_{t=1}^n \sum_{s=1,\, s<t}^n (C_{st} + C_{ts}P)(C_{st}' + P'C_{ts}')$$

$$\to 8\pi^2\{G \otimes G + (G \otimes G_1, \ldots, G \otimes G_p)\}.$$

The left-hand side is

$$\frac{1}{2mn^2} \sum_{t=1}^n \sum_{s=1, s\neq t}^n \sum_j' \mathrm{H}_j(\mathrm{e}^{\mathrm{i}(s-t)\lambda_j} + \mathrm{e}^{\mathrm{i}(t-s)\lambda_j}P) \sum_k' (\mathrm{e}^{\mathrm{i}(t-s)\lambda_k} + \mathrm{e}^{\mathrm{i}(s-t)\lambda_k}P)\mathrm{H}_k^*$$

$$= \frac{1}{m} \sum_j' \mathrm{H}_j(\mathrm{H}_j^* + P\mathrm{H}_{n-j}^*) - \frac{1}{2mn} \sum_j' \mathrm{H}_j(I_{p^2} + P) \sum_k' (I_{p^2} + P)\mathrm{H}_k^*.$$

The second term is easily seen to be $\mathrm{O}_{\mathrm{p}}(m/n)$. The first term can be written

$$\frac{1}{m} \sum_j' \{(\bar{B}_j B_j' \otimes B_j B_j^*) + (\bar{B}_j \otimes B_j)(B_j' \otimes B_{1j}^*, \ldots, B_j' \otimes B_{pj}^*)\},$$

where B_{ij}^* is the ith column of B_j^*. Since $B(\mathrm{e}^{\mathrm{i}\lambda})B(\mathrm{e}^{\mathrm{i}\lambda})^*/2\pi \to G$ as $\lambda \to 0$, and G is symmetric, the proof is readily completed. □

Proof of Proposition 3. From Proposition 2 it suffices to show that $\hat{G}(\bar{d}_*) - \hat{G}(d_*) = \mathrm{o}_{\mathrm{p}}(m^{-1/2})$. From Proposition 1 we have $\bar{d}_* - d_* = \mathrm{O}_{\mathrm{p}}(m_1^{-1/2})$. We thus have, for any $\varepsilon > 0$, $\eta > 0$

$$P\{m^{1/2}||\hat{G}(\bar{d}_*) - \hat{G}(d_*)|| > \varepsilon\}$$

$$\leqslant P\{m^{1/2}||\hat{G}(\bar{d}_*) - \hat{G}(d_*)|| > \varepsilon,\ |\bar{d}_* - d_*| \leqslant \eta\} + \mathrm{o}(1)$$

as $n \to \infty$. By the mean value theorem, for small enough η and $|\bar{d}_* - d_*| \leqslant \eta$, $||\hat{G}(\bar{d}_*) - \hat{G}(d_*)||$ is bounded by

$$\frac{2\log n}{m}\ |\bar{d}_* - d_*| \sum_{j=1}^m \lambda_j^{2(d_* - \eta)} \mathrm{tr}\{I_j\} = \mathrm{O}_{\mathrm{p}}\left(\frac{\log n}{m} |\bar{d}_* - d_*| \sum_{j=1}^m \lambda_j^{-2\eta}\right)$$

due to $EI_{aa}(\lambda_j) = \mathrm{O}(\lambda_j^{-2d})$, $a = 1, \ldots, p$ (see Robinson, 1995a, Theorem 2). The right-handside is $\mathrm{O_p}(m_1^{-1/2}(n/m)^{2\eta} \log n) = \mathrm{O_p}(m_1^{-1/2}(n/m)^{\zeta})$ for $\zeta > 2\eta$. From Assumption H, this is $\mathrm{o_p}(m^{-1/2})$, to complete the proof. □

Proof of Proposition 4. Let Q be a $p \times p$ orthogonal matrix such that $QGQ' = \Delta = \mathrm{diag}\{\delta_1, \ldots, \delta_p\}$. Then the eigenvalues of $\hat{G}(\bar{d}_*)$ are identical with those of $Q\hat{G}(\bar{d}_*)Q'$. Put $U(n) = m^{1/2}(Q\hat{G}(\bar{d}_*)Q' - \Delta)$. From Proposition 3 it readily follows that

$$vec\{U(n)\} \to_{\mathrm{d}} \mathrm{N}(0, \tfrac{1}{2}\{\Delta \otimes \Delta + (\Delta \otimes \Delta_1, \ldots, \Delta \otimes \Delta_p)\}),$$

where Δ_i is the ith column of Δ. The proof can then be completed by the argument of Theorem 1 of Anderson (1963) and Theorem 13.5.1 of Anderson (1984), noting that the limit distribution in Proposition 3 is the same as that of $m^{1/2}vec\{W_p(2m, G)/(2m) - G\}$ as $m \to \infty$, W_p denoting a Wishart variate, and, on p. 141 of Anderson (1963), replacing r by $p - r + 1$, taking $q_i = 1$, $q_{p-r+1} = r$, and then putting $\lambda_i = \delta_i$, $i = 1, \ldots, p - r, \lambda_{p-r+1} = 0$. □

Proof of Theorem 3. A straightforward application of Proposition 4 and the delta method. □

Proof of Theorem 4. We have

$$P(\hat{r} > r) \leqslant \sum_{u=r+1}^{p-1} P(L(u) < L(r)) \leqslant pP(\hat{\delta}_{p-r} < pv(n)) \to 0$$

as $n \to \infty$ since $\hat{\delta}_{p-r} \to_p \delta_{p-r} > 0$ and $v(n) \to 0$ by Proposition 4 and Assumption J. On the other hand,

$$P(\hat{r} < r) \leqslant \sum_{u=1}^{r-1} P(L(u) < L(r)) \leqslant rP(m^{1/2}\hat{\delta}_{p-r+1} > m^{1/2}v(n)) \to 0$$

as $n \to \infty$ since $\hat{\delta}_{p-r+1} \to_p 0$ and $m^{1/2}v(n) \to \infty$ by Proposition 4 and Assumption J. The proof is complete. □

References

Abadir, K.M., Taylor, A.M.R., 1999. On the definition of (co-)integration. Journal of Time Series Analysis 20, 129–137.

Adenstedt, R.K., 1974. On large-sample estimation for the mean of a stationary random sequence. Annals of Statistics 2, 1095–1107.

Anderson, T.W., 1963. Asymptotic theory for principal component analysis. Annals of Mathematical Statistics 34, 122–148.

Anderson, T.W., 1984. An Introduction To Multivariate Statistical Analysis, 2nd Edition. Wiley, New York.

Banerjee, A., Dolado, J., Galbraith, J.W., Hendry, D.F., 1993. Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data. Oxford University Press, Oxford.

Brillinger, D.R., 1975. Time Series, Data Analysis and Theory. Holt, Rinehart and Winston, New York.

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

Cheung, Y.-W., Lai, K.S., 1993. A fractional cointegration analysis of purchasing power parity. Journal of Business and Economic Statistics 11, 103–112.

Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Diebold, F.X., Rudebusch, G.D., 1991. On the power of Dickey–Fuller tests against fractional alternatives. Economics Letters 35, 155–160.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Engle, R.F., Granger, C.W.J., 1991. Long-Run Economic Relationships. Oxford University Press, Oxford.

Flôres Jr., R.G., Szafarz, A., 1996. An enlarged definition of cointegration. Economics Letters 50, 193–195.

Fujikoshi, Y., 1985. Two methods for estimation of dimensionality in canonical correlation analysis and the multivariate linear model. In: Matusita, K. (Ed.), Statistical Theory and Data Analysis. Elsevier, Amsterdam, pp. 233–240.

Fujikoshi, Y., Veitch, L.G., 1979. Estimation of dimensionality in canonical correlation analysis. Biometrika 66, 345–351.

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

Gunderson, B.K., Muirhead, R.J., 1997. On estimating the dimensionality in canonical correlation analysis. Journal of Multivariate Analysis 62, 121–136.

Hatanaka, M., 1996. Time-Series-Based Econometrics: Unit Roots and Cointegration. Oxford University Press, Oxford.

Henry, M., Robinson, P.M., 1996. Bandwidth choice in Gaussian semiparametric estimation of long-range dependence. In: Robinson, P.M., Rosenblatt, M. (Eds.), Athens Conference on Applied Probability and Time Series Analysis. Springer-Verlag, New York, pp. 220–232.

Hsu, J.C., 1996. Multiple Comparisons: Theory and Methods. Chapman & Hall, London.

Johansen, S., 1996. Likelihood-Based Inference In Cointegrated Vector Autoregressive Models, 2nd Printing. Oxford University Press, Oxford.

Künsch, H., 1987. Statistical aspects of self-similar processes. In: Prohorov, Yu, Sazonov, V.V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society 1. VNU Science Press, Utrecht, pp. 67–74.

Liu, C.L., 1968. Introduction to Combinatorial Mathematics. McGraw-Hill, New York.

Lobato, I., 1995. Multivariate analysis of long memory series in the frequency domain. Ph.D. Thesis, London School of Economics.

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

Marcus, R., Peritz, E., Gabriel, K.R., 1976. On closed testing procedures with special references to ordered analysis of variance. Biometrika 63, 655–660.

MacKinnon, J., 1991. Critical values for cointegration tests. In: Engle, R.F., Granger, C.W.J. (Eds.), Long-Run Economic Relationships. Oxford University Press, Oxford, pp. 267–276.

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., 1994. Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515–539.
Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.
Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630–1661.
Robinson, P.M., Marinucci, D., 1998. Semiparametric frequency domain analysis of fractional cointegration. STICERD Econometrics Discussion Paper EM/98/348, London School of Economics.
Said, S.E., Dickey, D.A., 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 71, 599–607.
Stock, J.H., Watson, M., 1988. Testing common trends. Journal of the American Statistical Association 83, 1097–1107.
Velasco, C., 1999. Gaussian semiparametric estimation of nonstationary time series. Journal of Time Series Analysis 20, 87–127.
Yong, C.H., 1974. Asymptotic Behaviour of Trigonometric Series. Chinese University of Hong Kong, Hong Kong.