Nonparametric Cointegration Analysis

Herman J. Bierens[1]

Pennsylvania State University, U.S.A.,

and

Tilburg University, the Netherlands

In this paper we propose consistent cointegration tests, and estimators of a basis of the space of cointegrating vectors, that do not need specification of the data-generating process, apart from some mild regularity conditions, or estimation of structural and/or nuisance parameters. This nonparametric approach is in the same spirit as Johansen's LR method in that the test statistics involved are obtained from the solutions of a generalized eigenvalue problem, and the hypotheses to be tested are the same, but in our case the two matrices in the generalized eigenvalue problem involved are constructed independently of the data-generating process. We compare our approach empirically as well as by a limited Monte Carlo simulation with Johansen's approach, using the series for ln(wages) and ln(GNP) from the extended Nelson-Plosser data.

Key words: Cointegration, unit roots, nonparametric, nuisance parameter free, hypotheses testing, estimation

JEL Codes: C12, C14, C32

Final version: February 1996

---

[1] Correspondence address: Department of Economics, Pennsylvania State University, University Park, PA 16802, U.S.A. This paper was written and revised while affiliated with Southern Methodist University, Dallas, and enjoying the hospitality of Tilburg University during the summers of 1993, 1994 and 1995. The helpful comments of Manfred Deistler, Philip Hans Franses, Noud van Giersbergen, Esfandiar Maasoumi, Rolf Tschernig, Ben Vogelvang, and four referees, are gratefully acknowledged. Previous versions of this paper have been presented at the University of Amsterdam, Free University of Amsterdam, University of Houston-Rice University, Texas A&M University, Tinbergen Institute Rotterdam, the University of British Columbia, the University of Virginia, the University of Wisconsin, ESEM 1994, and the ERNSI Econometric Workshop 1995, the Netherlands.

1. Introduction

The concept of cointegration was introduced by Granger (1981) and elaborated further by Engle and Granger (1987), Engle (1987), Engle and Yoo (1987), Stock and Watson (1988), Phillips and Ouliaris (1990), Park (1990), Phillips (1991), Boswijk (1993,1994), Perron and Campbell (1993), Johansen (1988, 1991, 1994), and Harris (1995), among others. The basic idea behind cointegration is that if all the components of a vector time series process z_t have a unit root there may exist linear combinations $\xi^T z_t$ without a unit root. These linear combinations may then be interpreted as long term relations between the components of z_t.

In a recent series of influential papers, Johansen (1988, 1991, 1994) and Johansen and Juselius (1990) propose an ingenious and practical full maximum likelihood estimation and testing approach, based on a Gaussian *E*rror *C*orrection *M*odel (*ECM*). This ECM is based on the Engle-Granger (1987) error correction representation theorem for cointegrated systems, and the asymptotic inference involved is related to the work of Sims, Stock and Watson (1990). By stepwise concentrating all the parameter matrices in the likelihood function out, except the matrix of cointegrating vectors, Johansen shows that the ML estimators of the cointegrating vectors can be derived from the eigenvectors of a generalized eigenvalue problem, and LR tests of the number of cointegrating vectors from the eigenvalues. This approach has become the standard tool in macroeconometrics for analyzing long term economic relations.

All cointegration approaches in the literature require consistent estimation of nuisance and/or structural parameters. In this paper we propose consistent cointegration tests that do not need specification of the data-generating process, apart from some mild regularity conditions, or estimation of (nuisance) parameters. Thus these tests are completely nonparametric. Our tests are conducted analogously to Johansen's tests, inclusive the test for parametric restrictions on the cointegrating vectors, namely on the basis of the ordered solutions of a generalized eigenvalue problem. Moreover, similarly to Johansen's approach we can consistently estimate a basis of the space of cointegrating vectors, using the eigenvectors of the generalized eigenvalue problem involved. However, in our case the two matrices involved are constructed independently of the data-generating process, and we can use the same set of tables of critical values for all the cointegration cases considered in Stock and Watson (1988) and Johansen (1988, 1991, 1994).

The plan of the paper is as follows: First, in section 2, we formulate our maintained hypotheses. In section 3 we propose a class of pairs of random matrices for which the generalized eigenvalues have similar properties as in the Johansen approach, based on weighted means of the level variables z_t and the first differences Δz_t. On the basis of these eigenvalues, we propose in section 4 tests for the number of cointegrating vectors similar to Johansen's (1988, 1991) lambda-max test. In section 5 we discuss the choice of the weight functions. In section 6 we propose tests for linear restrictions on cointegrating vectors, and a procedure for consistently estimating a basis of the space of cointegrating vectors. Up to this point we have maintained the assumption that the data-generating process is an integrated vector time series process with drift, where the vector of drift parameters is orthogonal to the cointegrating vectors. In section 7 we show how to make our approach invariant to unconstrained drift, including seasonal drift. Finally, in section 8 we compare our approach with Johansen's ML approach, empirically using the logs of wages and GNP from the extended Nelson-Plosser (1982) data set, as well as by a limited Monte Carlo simulation.

Proofs of all the lemmas are given in a separate appendix to this paper. Also, additional Monte Carlo results regarding the limiting null distributions of the tests, unit root test results for the extended Nelson-Plosser data, and further details of the cointegration test results for ln(wages) and ln(GNP), can be found in this separate appendix, which is available from the author on request. The empirical applications have been conducted using a computer program package developed by the author.[2]

2. The data-generating process

Consider the q-variate unit root process with drift $z_t = \mu + z_{t-1} + u_t$, where u_t is a zero mean stationary process, and μ is a vector of drift parameters. We assume that z_t is observable for $t = 0,1,2,..,n$. Due to the Wold decomposition theorem, we can write (under some mild

---

[2] This package, called *EasyReg International*, conducts our nonparametric cointegration analysis together with Johansen's tests, various unit root tests, VAR innovation response analysis, OLS, IV, Probit and Logit, and much more. It is freely downloadable from URL http://econ.la.psu.edu/~hbierens/EASYREG.HTM

regularity conditions),

$$u_t = \sum_{j=0}^{\infty} C_j v_{t-j} = C(L)v_t, \tag{1}$$

where v_t is a q-variate stationary white noise process, and $C(L)$ is a $q \times q$ matrix of lag polynomials in the lag operator L. For convenience we assume that $C(L)$ is a rational lag polynomial, and that the v_t's are Gaussian white noise, so that u_t is a Gaussian VARMA process:

Assumption 1. The process u_t can be written as (1), *with v_t i.i.d. $N_q(0,I_q)$ and* $C(\mathrm{L}) = C_1(L)^{-1}C_2(L)$, *where $C_1(L)$ and $C_2(L)$ are finite-order lag polynomials, with all the roots of* $\det(C_1(L))$ *lying outside the complex unit circle.*

This assumption is more restrictive than necessary, but it will keep the argument below transparent, and focused on the main issues. See Phillips and Solo (1992) for weaker conditions in the case of linear processes. Also, we could assume instead of Assumption 1 that u_t is stationary and ergodic, so that we can write $u_t = \varepsilon_t + w_t - w_{t-1}$, where ε_t is a martingale difference process with variance matrix comparable with $C(1)C(1)^T$. Cf. Hall and Heyde (1980, p.136), and equation (2) below. Note that we do not restrict the lag polynomial $C_2(L)$, except for the implicit restrictions imposed by Assumption 2 below.

Since by construction the lag polynomial $C(L) - C(1)$ is zero at $L = 1$, we can write

$$\begin{aligned} u_t &= C(L)v_t = C(1)v_t + (C(L)-C(1))v_t = C(1)v_t + (1-L)D(L)v_t \\ &= C(1)v_t + w_t - w_{t-1}, \end{aligned} \tag{2}$$

where $w_t = D(L)v_t$ and $D(L) = (C(L)-C(1))/(1-L) = \Sigma_{k=0}^{\infty} D_k L^k$. Thus:

$$z_t = z_0 - w_0 + \mu t + w_t + C(1)\sum_{j=1}^{t} v_j. \tag{3}$$

The process z_t is cointegrated with r linear independent cointegrating vectors ξ_j, $j = 1,..,r$, say, if rank(C(1)) = $q - r < q$. Then $\xi_j^T C(1) = 0^T$ for j = 1,..,r, hence it follows from (3) that $\xi_j^T z_t$ is trend stationary, with trend function $\xi_j^T(z_0 - w_0) + \xi_j^T \mu t$.

Note that Assumption 1 guarantees that $C(L)v_t$ and $D(L)v_t$ are well-defined stationary processes and that $\sum C_k$, $\sum C_k C_k^T$, $\sum D_k$ and $\sum D_k D_k^T$ converge. Cf. Engle (1987). For later reference it will be convenient to write the latter matrix as:

$$\sum_{k=0}^{\infty} D_k D_k^T = D_* D_*^T. \tag{4}$$

Assumption 1 will be our maintained hypothesis, together with the following assumption:

Assumption 2. Let R_r be the matrix of the eigenvectors of $C(1)C(1)^T$ corresponding to the r zero eigenvalues. Then the matrix $R_r^T D(1)D(1)^T R_r$ is nonsingular.

Moreover, for the time being we shall assume that the cointegration relations $R_r^T z_t$ are stationary about a possible intercept but not about a trend. Thus:

Assumption 3. $R_r^T \mu = 0$.

This assumption will be dropped in due course.

3. Convergence in distribution of a class of random matrices and their generalized eigenvalues

Our tests will be based on the following pair of random matrices:

$$\hat{A}_m = \sum_{k=1}^{m} a_{n,k} a_{n,k}^T, \quad \hat{B}_m = \sum_{k=1}^{m} b_{n,k} b_{n,k}^T,$$

depending on a natural number $m \geq q$, where

$$a_{n,k} = \frac{M_n^z(F_k)/\sqrt{n}}{\sqrt{\int\int F_k(x)F_k(y)\min(x,y)dxdy}}, \quad b_{n,k} = \frac{\sqrt{n}\,M_n^{\Delta z}(F_k)}{\sqrt{\int F_k(x)^2 dx}},$$

with

$$M_n^z(F_k) = \frac{1}{n}\sum_{t=1}^{n} F_k(t/n)z_t, \quad M_n^{\Delta z}(F_k) = \frac{1}{n}\sum_{t=1}^{n} F_k(t/n)\Delta z_t,$$

where $\{F_k\}$ is a class of differentiable real functions on the unit interval [0,1]. As will be shown below, the functions F_k can be chosen such that

$$\begin{aligned} \hat{A}_m &\overset{D}{\to} \left(C(1)C(1)^T\right)^{1/2}\left(\sum_{k=1}^{m} X_k X_k^T\right)\left(C(1)C(1)^T\right)^{1/2}, \\ \hat{B}_m &\overset{D}{\to} \left(C(1)C(1)^T\right)^{1/2}\left(\sum_{k=1}^{m} Y_k Y_k^T\right)\left(C(1)C(1)^T\right)^{1/2}, \end{aligned} \tag{5}$$

where the X_k's and Y_k's are independent q-variate standard normal random vectors, and $\overset{D}{\to}$ indicates convergence in distribution. In order to apply the result of Andersen, Brons and Jensen (1983), saying:

> *if for a pair of square random matrices* P_n, Q_n, (P_n, Q_n) *converges in distribution to* (P,Q), *where Q is a.s. nonsingular, then the ordered solutions of the generalized eigenvalue problem* $\det(P_n - \lambda Q_n) = 0$ *converge in distribution to the ordered solutions of the generalized eigenvalue problem* $\det(P - \lambda Q) = 0$,

we need to transform one of our matrices such that its limiting matrix becomes a.s. nonsingular. As will be shown below, choosing $P_n = \hat{A}_m$ and $Q_n = \hat{B}_m + n^{-2}\hat{A}_m^{-1}$ yields a suitable pair (P_n, Q_n), such that if $\text{rank}(C(1)C(1)^T) = q - r$ then the $q - r$ largest solutions of $\det(P - \lambda Q) = 0$ are a.s. positive and free of nuisance parameters, whereas the r smallest solutions are zero. However, notice that the above specification of the matrix Q_n is not the only possibility. For example the asymptotic results below will not change if we specify $Q_n = \hat{B}_m + n^{-2}c\hat{A}_m^{-1}$, for arbitrary scalar $c > 0$.

Now choose the functions F_k such that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} F_k(t/n) = o(1), \tag{6}$$

$$\frac{1}{n\sqrt{n}}\sum_{t=1}^{n} tF_k(t/n) = o(1), \tag{7}$$

and for $i \neq j$,

$$\int\int F_i(x)F_j(y)\min(x,y)\,dx\,dy = 0, \tag{8}$$

$$\int F_j(x)\int_0^x F_i(y)\,dy\,dx = 0, \tag{9}$$

$$\int F_i(x)F_j(x)\,dx = 0. \tag{10}$$

Note that the integrals involved are taken over the unit interval [0,1] if not otherwise indicated. It is a standard exercise in Wiener measure calculus to show (see, e.g., Billingsley 1968, Phillips 1987, Bierens 1994, Ch.9) that for each k,

$$\begin{pmatrix} M_n^z(F_k)/\sqrt{n} \\ M_n^{\Delta z}(F_k)\sqrt{n} \end{pmatrix} \overset{D}{\rightarrow} \begin{pmatrix} C(1)\int F_k(x)W(x)dx \\ C(1)\left(F_k(1)W(1) - \int f_k(x)W(x)dx\right) \end{pmatrix} \sim N_{2q}(0,\ (C(1)C(1)^T)\otimes\Sigma_k), \tag{11}$$

where W is a q-variate standard Wiener process, f_k is the derivative of F_k, and

$$\Sigma_k = \begin{pmatrix} \int\int F_k(x)F_k(y)\min(x,y)dx\,dy & 0 \\ 0 & \int F_k(x)^2 dx \end{pmatrix}. \tag{12}$$

The absence of the drift parameter vector μ in the right-hand side of (11) is due to conditions (6) and (7). Since the matrix Σ_k in (12) is diagonal, due to condition (6), and the two components on the right-hand side of (11) are linear functionals of a Wiener process and thus normally distributed, they are independent. They are also independent over k, due to the conditions (8), (9) and (10). Thus we have:

Lemma 1. Under Assumption 1 and conditions (6) *through* (10),

$$\begin{pmatrix} M_n^z(F_k)/\sqrt{n} \\ M_n^{\Delta z}(F_k)\sqrt{n} \end{pmatrix} \overset{D}{\to} \begin{pmatrix} C(1)X_k\sqrt{\int\int F_k(x)F_k(y)\min(x,y)\,dx\,dy} \\ C(1)Y_k\sqrt{\int F_k(x)^2dx} \end{pmatrix},$$

jointly for k = 1,...,*m, with m a fixed natural number, where the* X_k*'s and* Y_k*'s are independent q-variate standard normally distributed random vectors depending on* F_k *in the following way:*

$$X_k = \frac{\int F_k(x)W(x)dx}{\sqrt{\int\int F_k(x)F_k(y)\min(x,y)dxdy}}, \quad Y_k = \frac{F_k(1)W(1) - \int f_k(x)W(x)dx}{\sqrt{\int F_k(x)^2dx}}. \tag{13}$$

This result holds regardless of the possible existence of cointegration. Thus Lemma 1 proves (5), with $C(1)X_k$ replaced by $[C(1)C(1)^T]^{1/2}X_k$, and similarly for Y_k.

Next, assume that there are r linear independent cointegrating vectors. As is well-known, we can write

$$C(1)C(1)^T = R\Lambda R^T = (R_{q-r},\ R_r)\begin{pmatrix} \Lambda_{q-r} & O \\ O & O \end{pmatrix}\begin{pmatrix} R_{q-r}^T \\ R_r^T \end{pmatrix},$$

where Λ_{q-r} is the diagonal matrix of the $q-r$ positive eigenvalues, R_{q-r} is the corresponding matrix of orthonormal eigenvectors, and R_r is the matrix of orthonormal eigenvectors corresponding to the r zero eigenvalues. Then:

Lemma 2. Under Assumption 3 and the conditions of Lemma 1,

$$\begin{pmatrix} R_r^T M_n^z(F_k)\sqrt{n} \\ R_r^T M_n^{\Delta z}(F_k)\, n \end{pmatrix} \stackrel{D}{\to} \begin{pmatrix} R_r^T D(1) Y_k \sqrt{\int F_k(x)^2 dx} \\ F_k(1) R_r^T D_* Z \end{pmatrix},$$

jointly in $k = 1,...,m$, *where the* Y_k*'s and Z are independent q-variate standard normally distributed, with* Y_k *defined by* (13). *Moreover, Z does not depend on* F_k.

Such weight functions F_k do exist. In particular,

Lemma 3. If $F_k(x) = \cos(2k\pi x)$, *then the conditions* (6) *through* (10) *hold. Moreover, we then have* $F_k(1) = 1$, $\int\int F_k(x)F_k(y)\min(x,y)\,dx\,dy = \frac{1}{8}(k\pi)^{-2}$, $\int F_k(x)^2 dx = \frac{1}{2}$.

There are many ways to choose these functions F_k, but as will be shown in section 5, the above choice is optimal in some sense.

Denoting

$$\gamma_k = \frac{\sqrt{\int F_k(x)^2 dx}}{\sqrt{\int\int F_k(x)F_k(y)\min(x,y)dxdy}}, \quad \delta_k = \frac{F_k(1)}{\sqrt{\int F_k(x)^2 dx}}, \tag{14}$$

it follows now easily from Lemmas 1-2:

Lemma 4. Let rank $C(1) = q - r$. Then

$$\begin{pmatrix} I_{q-r} & O \\ O & nI_r \end{pmatrix} R^T \hat{A}_m R \begin{pmatrix} I_{q-r} & O \\ O & nI_r \end{pmatrix} = \begin{pmatrix} R_{q-r}^T \hat{A}_m R_{q-r} & nR_{q-r}^T \hat{A}_m R_r \\ nR_r^T \hat{A}_m R_{q-r} & n^2 R_r^T \hat{A}_m R_r \end{pmatrix}$$

$$\overset{D}{\to} \begin{pmatrix} R_{q-r}^T C(1) \sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r} & R_{q-r}^T C(1) \sum_{k=1}^{m} \gamma_k X_k Y_k^T D(1)^T R_r \\ R_r^T D(1) \sum_{k=1}^{m} \gamma_k Y_k X_k^T C(1)^T R_{q-r} & R_r^T D(1) \sum_{k=1}^{m} \gamma_k^2 Y_k Y_k^T D(1)^T R_r \end{pmatrix} \tag{15}$$

and

$$\begin{pmatrix} I_{q-r} & O \\ O & \sqrt{n} I_r \end{pmatrix} R^T \hat{B}_m R \begin{pmatrix} I_{q-r} & O \\ O & \sqrt{n} I_r \end{pmatrix} = \begin{pmatrix} R_{q-r}^T \hat{B}_m R_{q-r} & \sqrt{n} R_{q-r}^T \hat{B}_m R_r \\ \sqrt{n} R_r^T \hat{B}_m R_{q-r} & nR_r^T \hat{B}_m R_r \end{pmatrix}$$

$$\overset{D}{\to} \begin{pmatrix} R_{q-r}^T C(1) \sum_{k=1}^{m} Y_k Y_k^T C(1)^T R_{q-r} & R_{q-r}^T C(1) \sum_{k=1}^{m} \delta_k Y_k Z^T D_*^T R_r \\ R_r^T D_* \sum_{k=1}^{m} \delta_k Z Y_k^T C(1)^T R_{q-r} & R_r^T D_* \sum_{k=1}^{m} \delta_k^2 Z Z^T D_*^T R_r \end{pmatrix}, \tag{16}$$

where the random vectors X_i, Y_j and Z are the same as in Lemmas 1-2. Moreover,

$$\frac{R^T \hat{A}_m^{-1} R}{n^2} \overset{D}{\to} \begin{pmatrix} O & O \\ O & V_{r,m}^{-1} \end{pmatrix}, \tag{17}$$

where

$$V_{r,m} = R_r^T D(1) \sum_{k=1}^{m} \gamma_k^2 Y_k Y_k^T D(1)^T R_r - \left(R_r^T D(1) \sum_{k=1}^{m} \gamma_k Y_k X_k^T C(1)^T R_{q-r} \right)$$

$$\times \left(R_{q-r}^T C(1) \sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r} \right)^{-1} \left(R_{q-r}^T C(1) \sum_{k=1}^{m} \gamma_k X_k Y_k^T D(1)^T R_r \right).$$

Note that Assumption 2 guarantees that the matrix $V_{r,m}$ is a.s. nonsingular.

Denoting

$$X_k^* = \left(R_{q-r}^T C(1)C(1)^T R_{q-r}\right)^{-\frac{1}{2}} R_{q-r}^T C(1)X_k, \quad Y_k^* = \left(R_{q-r}^T C(1)C(1)^T R_{q-r}\right)^{-\frac{1}{2}} R_{q-r}^T C(1)Y_k, \tag{18}$$

and using the result of Andersen, Brons and Jensen (1983), it follows straightforwardly from Lemma 4:

Theorem 1. Let $\hat{\lambda}_{1,m} \geq .. \geq \hat{\lambda}_{q,m}$ *be the ordered solutions of the generalized eigenvalue problem*

$$\det\left[\hat{A}_m - \lambda(\hat{B}_m + n^{-2}\hat{A}_m^{-1})\right] = 0, \tag{19}$$

and let $\lambda_{1,m} \geq ... \geq \lambda_{q-r,m}$ *be the ordered solution of the generalized eigenvalue problem*

$$\det\left(\sum_{k=1}^{m} X_k^* X_k^{*\,T} - \lambda \sum_{k=1}^{m} Y_k^* Y_k^{*\,T}\right) = 0. \tag{20}$$

where the X_i^**'s and* Y_j^**'s are i.i.d.* $N_{q-r}(0,I_{q-r})$*. If* z_t *is cointegrated with r linear independent cointegrating vectors then under Assumptions 1-3,* $(\hat{\lambda}_{1,m},..,\hat{\lambda}_{q,m})$ *converges in distribution to* $(\lambda_{1,m},...,\lambda_{q-r,m},0,..,0)$.

In order to show how fast $(\hat{\lambda}_{q-r+1,m},..,\hat{\lambda}_{q,m})$ converges to (0,....,0), observe from Lemma 4 that

$$n^{-2}\left(R^T\hat{A}_m R\right)^{-\frac{1}{2}}\left(R^T(\hat{B}_m + n^{-2}\hat{A}_m^{-1})R\right)\left(R^T\hat{A}_m R\right)^{-\frac{1}{2}}$$

$$= \left(R^T n^{-2}\hat{A}_m^{-1}R\right)^{\frac{1}{2}}\left(R^T(\hat{B}_m + n^{-2}\hat{A}_m^{-1})R\right)\left(R^T n^{-2}\hat{A}_m^{-1}R\right)^{\frac{1}{2}} \overset{D}{\to} \begin{pmatrix} O & O \\ O & V_{r,m}^{-2} \end{pmatrix}.$$

Moreover, it is easy to see that the solutions $\hat{\mu}_{j,m}$ of the generalized eigenvalue problem

$$\det\left[\frac{1}{n^2}\left(R^T\hat{A}_n R\right)^{-\frac{1}{2}}\left(R^T(\hat{B}_m + n^{-2}\hat{A}_m^{-1})R\right)\left(R^T\hat{A}_n R\right)^{-\frac{1}{2}} - \mu I_q\right] = 0$$

are just the reciprocals of $n^2\hat{\lambda}_{j,m}$. Thus again referring to Andersen, Brons and Jensen (1983) it follows that $n^2(\hat{\lambda}_{q-r+1,m}, .. , \hat{\lambda}_{q,m})$ converges in distribution to the ordered eigenvalues of the matrix $V_{r,m}^2$. Finally, observe that, with X_k^* defined by (18) and

$$Y_k^{**} = \left(R_r^T D(1)D(1)^T R_r\right)^{-\frac{1}{2}} R_r^T D(1) Y_k \quad (\sim N_r[0, I_r]), \tag{21}$$

we can write the matrix $V_{r,m}$ as

$$V_{r,m} = \left(R_r^T D(1)D(1)^T R_r\right)^{\frac{1}{2}} V_{r,m}^* \left(R_r^T D(1)D(1)^T R_r\right)^{\frac{1}{2}} \tag{22}$$

where

$$V_{r,m}^* = \left(\sum_{k=1}^m \gamma_k^2 Y_k^{**} Y_k^{**\,T}\right) - \left(\sum_{k=1}^m \gamma_k Y_k^{**} X_k^{*\,T}\right)\left(\sum_{k=1}^m X_k^* X_k^{*\,T}\right)^{-1}\left(\sum_{k=1}^m \gamma_k X_k^* Y_k^{**\,T}\right). \tag{23}$$

Thus we have:

Theorem 2. Under the conditions of Theorem 1, $n^2(\hat{\lambda}_{q-r+1,m}, .. , \hat{\lambda}_{q,m})$ *converges in distribution to* $(\lambda_{1,m}^{*\,2},...,\lambda_{r,m}^{*\,2})$, *where* $\lambda_{1,m}^* \geq ... \geq \lambda_{r,m}^*$ *are the ordered solutions of the generalized eigenvalue problem*

$$\det\left[V_{r,m}^* - \lambda\left(R_r^T D(1)D(1)^T R_r\right)^{-1}\right] = 0,$$

where the matrix $V_{r,m}^*$ *is defined in* (23) *with the* X_i^**'s and* Y_j^{**}*'s independent* $q-r$*-variate and* r*-variate, respectively, standard normally distributed random vectors.*

4. Testing the number of cointegrating vectors

4.1. The lambda-min test, and a comparison with Johansen's tests

The results in Theorems 1-2 suggest to use the test statistic $\hat{\lambda}_{q-r,m}$ for testing the null hypothesis H_r that there are r cointegrating vectors against the alternative H_{r+1}. We shall call this test the lambda-min test, which (as will be shown below) is in the same spirit as Johansen's lambda-max test.

Johansen's (1988) original approach is based on the following ECM of the q-variate unit root process z_t:

$$\Delta z_t = \sum_{j=1}^{p-1} \Pi_j \Delta z_{t-j} + \gamma\beta^T z_{t-p} + e_t, \tag{24}$$

where the Π_j, $j > 0$, are $q\times q$ and β and γ are $q\times r$ parameter matrices with r the number of cointegrating vectors (the columns of β), and the e_t's are i.i.d. $N_q(0,\Sigma)$ errors. By stepwise concentrating all the parameter matrices in the likelihood function out, except the matrix β, Johansen shows that the ML estimator of β can be derived from the eigenvectors of the generalized eigenvalue problem $\det(S_{po}S_{oo}^{-1}S_{op} - \lambda S_{pp}) = 0$, where $S_{ij} = (1/n)\Sigma_{t=1}^{n} R_{i,t}R_{j,t}^{T}$, $i,j = o,p$, with $R_{o,t}$ the residual vector of the regression of Δz_t on $\Delta z_{t-1},\ldots,\Delta z_{t-p+1}$, and $R_{p,t}$ the residual vector of the regression of z_{t-p} on $\Delta z_{t-1},\ldots,\Delta z_{t-p+1}$. Moreover, the ordered eigenvalues $\hat{\lambda}_1 \geq \ldots. \geq \hat{\lambda}_q$ involved can be used for testing hypotheses about the number of cointegrating vectors. In particular, Johansen proposes two LR tests for the number of cointegrating vectors, the trace test and the lambda-max test. The test statistic of the latter test, for testing H_r against H_{r+1}, is $n\hat{\lambda}_{r+1}$. The trace test tests H_r against H_q, which is equivalent to the alternative that z_t is stationary. Johansen proves that $(\hat{\lambda}_1,\ldots,\hat{\lambda}_q)$ converges in distribution to $(c_1,\ldots,c_r,0,\ldots,0)$, where the c_j's are positive constants, and $n(\hat{\lambda}_{r+1},\ldots.,\hat{\lambda}_q)$ converges in distribution to $(\varepsilon_1,\ldots,\varepsilon_{q-r})$, where the ε_j's are positive random variables. Comparing Johansen's generalized eigenvalue results with Theorems 1-2 we see that we can mimic Johansen's tests by transforming our generalized eigenvalues $\hat{\lambda}_{j,m}$ by $\hat{\mu}_{j,m} = 1/(n\sqrt{\hat{\lambda}_{q+1-j,m}})$ and replacing Johansen's eigenvalues in his lambda-max and trace tests by these $\hat{\mu}_{j,m}$'s. In this paper we shall focus on the lambda-min test only, because for this test it is possible to optimize the power of the test to m, and the order in which it is applied is more natural than for a trace test.

4.2. The choice of "m"

The limiting distribution of the lambda-min test under the null as well as under the alternative depends on the test parameter m, and so do the α×100% critical values $K_{\alpha,q-r,m}$, say, as well as the power function. These critical values, which are presented in the separate appendix to this paper for $q-r = 1,..,5$ and $m = q-r,...,20$, with the weight functions F_k chosen as in Lemma 3, are calculated on the basis of 10,000 replications of the generalized eigenvalue problem (20). These critical values increase with m. Now the power of the test against the alternative H_{r+1} is

$$P(\hat{\lambda}_{q-r,m} \le K_{\alpha,q-r,m}) \approx P\left(\lambda^*_{1,m} \le n\sqrt{K_{\alpha,q-r,m}}\right),$$

where, by Chebishev's inequality, the latter probability is bounded from below as follows:

Lemma 5.

$$P\left(\lambda^*_{1,m} \le n\sqrt{K_{\alpha,q-r,m}}\right) \ge 1 - \frac{\left(1 - \frac{q-r-1}{m}\right)\sum_{k=1}^{m}\gamma_k^2}{\sqrt{K_{\alpha,q-r,m}}} \times \frac{\left[R_{r+1}^T D(1)D(1)^T R_{r+1}\right]}{n}. \tag{25}$$

This result suggests to choose m such that the right hand side of (25) is maximal, subject to the condition $m \ge q$. The values of m involved are presented in Table 1, for the case where the weight function F_k are chosen as in Lemma 3 (for which $\gamma_k = 2\pi k$), and the corresponding critical values are presented in Table 2.

Table 1: Optimal values of m if $F_k(x) = \cos(2k\pi x)$

20% significance level						*10% significance level*						*5% significance level*					
r: \ q:	1	2	3	4	5	r: \ q:	1	2	3	4	5	r: \ q:	1	2	3	4	5
0	1	2	3	4	5	0	1	2	4	5	6	0	1	3	4	5	6
1		2	3	4	5	1		2	3	4	5	1		2	3	4	5
2			3	4	5	2			3	4	5	2			3	4	5
3				4	5	3				4	5	3				4	5
4					5	4					5	4					5

Table 2: Critical values of the lambda−min test for

$F_k(x) = \cos(2k\pi x)$ *and m as in Table 1*

20% *significance level*

r:\q:	1	2	3	4	5
0	0.10927	0.01680	0.00647	0.00318	0.00202
1		0.24145	0.07695	0.03702	0.02337
2			0.34138	0.13448	0.07389
3				0.40009	0.18198
4					0.44898

10% *significance level*

r:\q:	1	2	3	4	5
0	0.02490	0.00451	0.01696	0.01107	0.00722
1		0.11106	0.03429	0.01696	0.01107
2			0.18732	0.07598	0.04309
3				0.24428	0.11266
4					0.29513

5% *significance level*

r:\q:	1	2	3	4	5
0	0.00598	0.01691	0.00842	0.00543	0.00357
1		0.05416	0.01691	0.00842	0.00543
2			0.11052	0.04622	0.02562
3				0.15818	0.07456
4					0.19710

4.4. Estimating the number of cointegrating vectors

Rather than testing for the number of cointegrating vectors, we can also estimate it consistently, as follows. Denote

$$\hat{g}_m(r) = \left(\prod_{k=1}^{q}\hat{\lambda}_{k,m}\right)^{-1} \quad if\ r = 0,$$

$$= \left(\prod_{k=1}^{q-r}\hat{\lambda}_{k,m}\right)^{-1}\left(n^{2r}\prod_{k=q-r+1}^{q}\hat{\lambda}_{k,m}\right) \quad if\ r = 1,....,q-1, \tag{26}$$

$$= n^{2q}\prod_{k=1}^{q}\hat{\lambda}_{k,m} \quad if\ r = q.$$

where m is chosen from Table 1 for one of the three significance levels and the test result for r, provided $r < q$, and $m = q$, say, if the test result is $r = q$. Then $\hat{g}_m(r)$ converges in probability to infinity if the true number of cointegrating vectors is unequal to r, and $\hat{g}_m(r) = O_p(1)$ if the true number of cointegrating vectors is indeed r. Thus, taking $\hat{r}_m = \text{argmin}_{0\le r\le 1}\{\hat{g}_m(r)\}$ we have $\lim_{n\to\infty}P(\hat{r}_m = r) = 1$. This approach may be useful as a double-check on the test results for r.

5. The choice of the weight functions F_k

The best choice of the weight functions F_k is such that the power of the lambda-min test is maximal, but again this is not feasible because the power depends on nuisance parameters. However, Lemma 5 suggests that the second best choice is to choose the F_k's as to minimize the squared γ_k's, subject to the conditions (6) through (10). In doing so, it will be convenient to replace first the conditions (6) and (7) by the weaker conditions

$$\int F_k(x)dx = 0 \tag{27}$$

and

$$\int xF_k(x)dx = 0, \tag{28}$$

respectively, and to verify afterwards that the optimal weight functions F_k satisfy the stronger conditions (6) and (7).

Without loss of generality we may represent the functions F_k by their Fourier series expansion

$$F_k(x) = \alpha_{0,k} + \sum_{j=1}^{\infty} \alpha_{j,k}\cos(2j\pi x) + \sum_{j=1}^{\infty} \beta_{j,k}\sin(2j\pi x)$$

Then by some tedious but straightforward calculations it can be shown that:

Lemma 6. The conditions (27), (28), (8), (9), *and* (10) *now read as:*

$$\int F_k(x)dx = \alpha_{0,k} = 0$$

$$\int xF_k(x)dx = -\frac{1}{2\pi}\sum_{j=1}^{\infty}\frac{\beta_{j,k}}{j} = 0$$

$$\int\int F_k(x)F_m(y)\min(x,y)dxdy = \frac{1}{8\pi^2}\left(\sum_{j=1}^{\infty}\frac{\alpha_{j,k}\alpha_{j,m}}{j^2} + \sum_{j=1}^{\infty}\frac{\beta_{j,k}\beta_{j,m}}{j^2}\right) = 0 \;\; if \;\; k \neq m,$$

$$\int F_k(x)\int_0^x F_m(y)dydx = \frac{1}{4\pi}\left(\sum_{j=1}^{\infty}\frac{\alpha_{j,k}\beta_{j,m}}{j} - \sum_{j=1}^{\infty}\frac{\alpha_{j,m}\beta_{j,k}}{j}\right) = 0$$

$$\int F_k(x)F_m(x)dx = \frac{1}{2}\left(\sum_{j=1}^{\infty}\alpha_{j,k}\alpha_{j,m} + \sum_{j=1}^{\infty}\beta_{j,k}\beta_{j,m}\right) = 0 \;\; if \;\; k \neq m.$$

Combining (14) and the results of Lemma 6, we have

$$\gamma_k^2 = 4\pi^2\frac{\sum_{j=1}^{\infty}\alpha_{j,k}^2 + \sum_{j=1}^{\infty}\beta_{j,k}^2}{\sum_{j=1}^{\infty}\frac{\alpha_{j,k}^2}{j^2} + \sum_{j=1}^{\infty}\frac{\beta_{j,k}^2}{j^2}},$$

which we are going to minimize subject to the conditions in Lemma 6, as follows: First, choose for some large natural number N and all $j,k > N$, $\alpha_{j,k} = \beta_{j,k} = 0$. Denote $\theta_k = (\alpha_{1,k},\beta_{1,k},\alpha_{2,k},\beta_{2,k},...,\alpha_{N,k},\beta_{N,k})^T$, $J = \text{diag}(1,1,2,2,...,N,N)$. Then

$$\gamma_k^2 = 4\pi^2 \frac{\theta_k^T \theta_k}{\theta_k^T J^{-2} \theta_k},$$

hence the unconstrained minimum of γ_k^2 corresponds to the minimum solution of the generalized eigenvalue problem $\det(I - \lambda J^{-2}) = 0$. Taking $k = 1$, this minimum eigenvalue is 1 (twice), with corresponding normalized eigenvectors $(1,0,0,...,0)^T$ and $(0,1,0,...,0)^T$. Thus, the unconstrained optimal solution satisfies $\alpha_{j,1} = \beta_{j,1} = 0$ for $j > 1$, and then the conditions in Lemma 6 imply that also $\beta_{1,1} = 0$, whereas without loss of generality we may take $\alpha_{1,1} = 1$. Next, for $k = 2$ the conditions in Lemma 6, except condition (28), imply that the optimal solution corresponds to the minimum eigenvalue for which the corresponding eigenvector is orthogonal to $(1,0,0,...,)^T$ and $(0,1,0,...,0)^T$. Clearly, this minimum eigenvalue is $k^2 = 4$, and the corresponding normalized eigenvectors are $(0,0,1,0,...,0)^T$ and $(0,0,0,1,...,0)^T$. Thus, $\alpha_{j,2} = \beta_{j,2} = 0$ for $j \neq 2$, and condition (28) implies that also $\beta_{2,2} = 0$, whereas again we may choose $\alpha_{2,2} = 1$. Continuing this argument shows that the optimal solution for F_k is the one in Lemma 3, provided that the stronger conditions (6) and (7) also hold. The latter has already been established in Lemma 3. Since N was chosen arbitrary, it follows by induction that:

Theorem 3. The choice $F_k(x) = \cos(2k\pi x)$ *for the weight functions is optimal in the sense that then for any fixed positive integer m the lower bound* (25) *of the power of the lambda-min test is maximal.*

6. Testing linear restrictions

6.1. Design of the generalized eigenvalue problem, and asymptotic distribution theory

Following Johansen (1988,1991), we now focus on the problem of how to test whether there exists a cointegrating vector ξ satisfying a linear relation of the form

$$H_0: \xi = H\varphi, \text{ where } \operatorname{rank}(H) = s \leq r, \quad \varphi \in \mathbb{R}^s. \tag{29}$$

Thus, the matrix H is of full column rank s. At first sight we may think of mimicking Johansen's test for these linear restrictions, on the basis of the matrices $\hat{A}_m$ and $\hat{B}_m + n^{-2}\hat{A}_m^{-1}$. However, that leads to a case-dependent asymptotic null distribution. Therefore we propose the following alternative approach, on the basis of the matrix $\hat{A}_m$ only.

First, note that the null hypothesis (29) implies

$$H = R_r\Gamma, \tag{30}$$

where Γ is a $r \times s$ matrix of rank s. Then it follows straightforwardly from (15), (17) and (30) that

$$n^2 H^T\hat{A}_m H = n^2\Gamma^T R_r^T\hat{A}_m R_r\Gamma \xrightarrow{D} \Gamma^T R_r^T D(1)\sum_{k=1}^{m}\gamma_k^2 Y_k Y_k^T D(1)^T R_r\Gamma$$

and

$$H^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)^{-1}H = \Gamma^T R_r^T R\left(R^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)R\right)^{-1}R^T R_r\Gamma \xrightarrow{D} \Gamma^T V_{r,m}\Gamma .$$

Since similarly to (21) we can write

$$Y_k^{**} = \left(\Gamma^T R_r^T D(1)D(1)^T R_r\Gamma\right)^{-\frac{1}{2}}\Gamma^T R_r^T D(1)Y_k \quad (\sim N_s[0,I_s]),$$

we have that

Theorem 4. If there are r cointegrating vectors then under the null hypothesis (29) *the ordered solutions of the eigenvalues problem*

$$\det\left[H^T\hat{A}_m H - \lambda H^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)^{-1}H\right] = 0 \tag{31}$$

times n^2, *converge jointly in distribution to the ordered solutions of the generalized eigenvalue problem* $\det[W_{s,m}^* - \lambda V_{s,q-r,m}^*] = 0$, *where* $W_{s,m}^* = \Sigma_{k=1}^m\gamma_k^2 Y_k^{**}Y_k^{**T}$,

$$V_{s,q-r,m}^* = \left(\sum_{k=1}^{m}\gamma_k^2 Y_k^{**}Y_k^{**T}\right) - \left(\sum_{k=1}^{m}\gamma_k Y_k^{**}X_k^{*T}\right)\left(\sum_{k=1}^{m}X_k^{*}X_k^{*T}\right)^{-1}\left(\sum_{k=1}^{m}\gamma_k X_k^{*}Y_k^{**T}\right). \tag{32}$$

[*cf.* (23)], *and the* Y_k^{**} *and* X_k^{*} *involved are independent s-variate and* $q-r$*-variate, respectively, standard normal random vectors.*

Note that the matrix $V^*_{s,q-r,m}$ in (32) differs from the matrix $V^*_{s,m}$ defined by (23) with r replaced by s in that in the latter case the vectors X^*_k are $(q-s) \times 1$ rather than $(q-r) \times 1$.

If the null hypothesis (29) is false, then the matrix H can be written as

$$H = R_{q-r}\Gamma_1 + R_r\Gamma_2, \quad \textit{where} \ \operatorname{rank}(\Gamma_1) = s_1 \geq 1, \ \operatorname{rank}\begin{pmatrix}\Gamma_1 \\ \Gamma_2\end{pmatrix} = s. \tag{33}$$

Then again it follows straightforwardly from (15), (17) and (30) that

$$H^T\hat{A}_m H \overset{D}{\to} \Gamma_1^T R_{q-r}^T C(1)\sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r}\Gamma_1$$

and

$$H^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)^{-1}H = H^T R\left(R^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)R\right)^{-1}R^T H$$

$$= \left(\Gamma_1^T, \Gamma_2^T\right)\left(R^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)R\right)^{-1}\begin{pmatrix}\Gamma_1 \\ \Gamma_2\end{pmatrix}$$

$$\overset{D}{\to} \Gamma_1^T\left(R_{q-r}^T C(1)\sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r}\right)^{-1}\Gamma_1 + \Gamma_2^T V_{r,m}\Gamma_2,$$

where the latter limit matrix is of full rank s. Therefore,

Theorem 5. If the null hypothesis (29) *is false, then the* s_1 *ordered largest solutions of the generalized eigenvalue problem* (31) *converge in distribution to the ordered solutions of the generalized eigenvalue problem*

$$\det\left[\Gamma_1^T R_{q-r}^T C(1)\sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r}\Gamma_1 - \lambda\left(\Gamma_1^T\left(R_{q-r}^T C(1)\sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r}\right)^{-1}\Gamma_1 + \Gamma_2^T V_{r,m}\Gamma_2\right)\right] = 0, \tag{34}$$

whereas the remaining $s - s_1$ *solutions of* (31) *converge in probability to zero, where* Γ_1, Γ_2 *and* s_1 *are defined in* (33).

6.2. The lambda-max and trace tests for linear restrictions

Theorems 4 and 5 suggest to use the maximum solution, or the sum $T_m(H)$, say, of all solutions, of eigenvalue problem (31) as a basis for a test of the null hypothesis (29). We only discuss the trace test in detail, as the asymptotic properties of the lambda-max tests can be derived along similar lines as for the trace test.

It follows straightforwardly from Theorems 4 and 5 that under the null hypothesis (29),

$$n^2\hat{T}_m(H) \overset{D}{\to} \text{trace}\left(W^*_{s,m}V^{*\,-1}_{s,q-r,m}\right), \tag{35}$$

whereas if this null hypothesis is false, $T_m(H)$ converges in distribution to the sum, $T_{1,m}(H)$, say, of the s_1 solutions of (34), hence $\text{plim}_{n\to\infty}n^2T_m(H) = \infty$. Thus, denoting the critical value of the trace test at the α×100% significance level by $M_{\alpha,s,q-r,m}$, we reject the null if $n^2T_m(H) \geq M_{\alpha,s,q-r,m}$.

6.3. The choice of "m"

A Monte Carlo simulation of the limiting distribution (35), based on 10,000 replications of the random vectors Y^{**}_k and X^*_k for $k = 1,...,m$, with $m = \max(s,q-r),...,20$, reveals that $M_{\alpha,s,q-r,m}$ is decreasing in m for $m \geq q-r+s$, and infinite for $m < q-r+s$, due to (near−) singularity of the matrix (32). See the separate appendix to this paper. Using the approximation

$$P(n^2\hat{T}_m(H) \geq M_{\alpha,s,q-r,m}) \approx P\left(T_{1,m}(H) \geq n^{-2}M_{\alpha,s,q-r,m}\right)$$

and

Lemma 7.

$$P\left(T_{1,m}(H) \geq n^{-2}M_{\alpha,s,q-r,m}\right) \geq P\left[\lambda_{\min}\left(\sum_{k=1}^{m}X^*_kX^{*\,T}_k\right) \geq \frac{n^{-1}\sqrt{M_{\alpha,s,q-r,m}}}{\lambda_{\min}\left(R^T_{q-r}C(1)C(1)^TR_{q-r}\right)}\right],$$

where the right-hand side probability is an increasing function of m, and $\lambda_{\min}(.)$ *stands for the minimum eigenvalue of the matrix involved,* it follows that in order to boost the power of the test we should choose m "large". On the other hand, m should not be too large, as otherwise m acts as being dependent on the sample size n, which may distort the size of the test. Since the critical values $M_{\alpha,s,q-r,m}$ hardly change anymore for $m > 2(q-r+s)$, and since we have to choose $m \geq q$ as otherwise the matrix $\hat{A}_m$ becomes singular, we recommend the rule-of-thumb $m = 2q$. The corresponding critical values are presented in Table 3. As is easy to see, the same rule-of-thumb applies to the lambda-max test. The critical values of the lambda-max test, for $m = 2q$, and the weight functions F_k chosen as in Lemma 3, are given in Table 4.

Table 3: *Critical values of the trace test* $\left(m = 2q,\ F_k(x) = \cos(2k\pi x)\right)$:

		$s=1$			$s=2$			$s=3$			$s=4$		
q	r	20%	10%	5%	20%	10%	5%	20%	10%	5%	20%	10%	5%
2	1	1.91	2.89	4.70									
3	1	2.24	3.14	4.44									
	2	1.45	1.82	2.35	3.23	4.11	5.36						
4	1	2.32	3.14	4.14									
	2	1.71	2.17	2.76	3.77	4.77	5.96						
	3	1.29	1.53	1.83	2.71	3.16	3.70	4.37	5.20	6.26			
5	1	2.37	3.12	4.03									
	2	1.87	2.32	2.86	4.12	5.06	6.16						
	3	1.51	1.80	2.14	3.14	3.68	4.32	5.05	5.97	6.96			
	4	1.22	1.39	1.58	2.50	2.79	3.16	3.87	4.32	4.89	5.38	6.08	6.96

Table 4: Critical values of the lambda−max test $\left(m = 2q,\ F_k(x) = \cos(2k\pi x)\right)$:

		$s=1$			$s=2$			$s=3$			$s=4$		
q	r	20%	10%	5%	20%	10%	5%	20%	10%	5%	20%	10%	5%
2	1	1.91	2.89	4.70									
3	1	2.24	3.14	4.44									
	2	1.45	1.82	2.35	2.23	3.11	4.36						
4	1	2.31	3.11	4.16									
	2	1.71	2.15	2.72	2.68	3.58	4.87						
	3	1.29	1.52	1.79	1.74	2.18	2.78	2.33	3.14	4.27			
5	1	2.41	3.13	4.08									
	2	1.85	2.31	2.85	2.85	3.71	4.78						
	3	1.50	1.79	2.13	2.08	2.60	3.22	2.83	3.73	4.84			
	4	1.22	1.38	1.58	1.50	1.78	2.13	1.86	2.31	2.85	2.41	3.12	4.02

6.4. Estimation of the cointegrating vectors

The results in section 6.1 can also be used to derive consistent estimators of the cointegrating vectors, as follows. Choose again $m = 2q$, and let H be the matrix of the r eigenvectors corresponding to the r smallest eigenvalues of the generalized eigenvalue problem

$$\det\left[\hat{A}_m - \lambda\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)^{-1}\right] = 0, \tag{36}$$

where H is standardized such that $\hat{H}^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)^{-1}\hat{H} = I_r$. Then similarly to (33) we can write

$$\hat{H} = R_{q-r}\hat{\Gamma}_1 + R_r\hat{\Gamma}_2, \quad \textit{where}\ \operatorname{rank}(\hat{\Gamma}_1) = s \geq 0,\ \operatorname{rank}\begin{pmatrix}\hat{\Gamma}_1\\ \hat{\Gamma}_2\end{pmatrix} = r, \tag{37}$$

with Γ_1 and Γ_2 stochastically bounded matrices. It follows now similarly to Theorem 4 that $n^2H^T\hat{A}_mH = O_p(1)$. Moreover, using (37) we can write

$$\hat{H}^T\hat{A}_m\hat{H} = \hat{\Gamma}_1^TR_{q-r}^T\hat{A}_mR_{q-r}\hat{\Gamma}_1 + \hat{\Gamma}_1^TR_{q-r}\hat{A}_mR_r\hat{\Gamma}_2 + \hat{\Gamma}_2^TR_r^T\hat{A}_mR_r\hat{\Gamma}_2 + \hat{\Gamma}_2^TR_r\hat{A}_mR_{q-r}\hat{\Gamma}_1.$$

Therefore, it follows easily from part (15) of Lemma 4 that $\Gamma_1 = O_p(1/n)$. Since by (37), $\Gamma_1 = R_{q-r}^TH$, we now have $R_{q-r}^TH = O_p(1/n)$. Thus:

Theorem 6. If there are r linear independent cointegrating vectors then the matrix H of standardized eigenvectors corresponding to the r smallest eigenvalues of the generalized eigenvalue problem (36) (*with m chosen from Table 1*) *satisfies* $R_{q-r}^{T}H = O_p(1/n)$, *where* R_{q-r} *is the matrix of eigenvectors of* $C(1)C(1)^T$ *corresponding to the positive eigenvalues.*

7. Cointegrating systems with unconstrained drift

7.1. Non-seasonal drift

Until so far all our derivations were based on the assumption that the drift parameter vector μ is orthogonal to the cointegrating vectors. Cf. Assumption 3. The problem is that without Assumption 3 the result of Lemma 2 no longer holds, due to the fact that $\Sigma_{t=1}^{n} t\cos(2k\pi t/n) = n/2$, although $\Sigma_{t=1}^{n}\cos(2k\pi t/n) = 0$, so that, with $F_k(x) = \cos(2k\pi x)$, the result of Lemma 2 now becomes:

$$\begin{pmatrix} R_r^T M_n^z(F_k)\sqrt{n} - \frac{1}{2}R_r^T\mu\sqrt{n} \\ R_r^T M_n^{\Delta z}(F_k)n \end{pmatrix} \overset{D}{\to} \begin{pmatrix} R_r^T D(1)Y_k\sqrt{\int F_k(x)^2 dx} \\ F_k(1)R_r^T D_* Z \end{pmatrix}.$$

and consequently, part (15) of Lemma 4 becomes:

$$R^T\hat{A}_m R \overset{D}{\to} \begin{pmatrix} R_{q-r}^T C(1)\sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r} & R_{q-r}^T C(1)\sum_{k=1}^{m} X_k \mu^T R_r \\ R_r^T\mu\sum_{k=1}^{m} X_k^T C(1)R_{q-r} & R_r^T\mu\mu^T R_r \end{pmatrix}.$$

Clearly, this will render all our test results invalid. However, a minor change of the functions F_k will cure the problem:

Lemma 8. Let $F_{n,k}(x) = \cos[2k\pi(nx - ½)/n]$ *and* $F_k(x) = \cos(2k\pi x)$. *Then*

$$\sum_{t=1}^{n} F_{n,k}(t/n) = \sum_{t=1}^{n} tF_{n,k}(t/n) = 0, \tag{38}$$

and

$$\lim_{n\to\infty}\left(n \sup_{0\le x\le 1} |F_{n,k}(x) - F_k(x)|\right) = k\pi. \tag{39}$$

The proof of this lemma is straightforward. Note that the functions $\cos(2k\pi(t-.5)/n)$ are known as Chebishev time polynomials, of even order. See, e.g., Hamming(1973).

It follows now easily from Lemma 8 that:

Theorem 7. With the weight functions F_k replaced by $F_{n,k}$, the results of Theorems 1 through 6 carry over to cointegrated systems with drift, without the need for Assumption 3.

Note that, due to (39), the optimality of the modified weight functions $F_{n,k}$ is preserved. Moreover, note that without Assumption 3 we allow the cointegration relations to be trend stationary. This case is considered only very recently by Johansen (1994) and, in a slightly different way, by Perron and Campbell (1993). Toda (1994) compares the two approaches involved by Monte Carlo simulation.

7.2. Seasonal drift

Next, consider the case where z_t is a seasonal vector time series process with s seasons. In that case the drift may differ per season:

$$z_t = z_{t-1} + \sum_{\tau=0}^{s-1} c_\tau d_{\tau,t} + u_t,$$

where the $d_{\tau,t}$'s are seasonal dummy variables, i.e., $d_{\tau,t} = 1$ if $t = js + \tau$ for some integer j and $d_{\tau,t} = 0$ if not, and the c_τ's are q-vectors of coefficients. However, the modified weight function $F_{n,k}$ do not sufficiently filter out the seasonal drift:

Lemma 9. For k = 1,2,...,

$$\lim_{n\to\infty}\sum_{t=1}^{n} d_{\tau,t}\cos[2k\pi(t - 0.5)/n] = -\frac{\tau}{2s}, \textit{ and}$$

$$\lim_{n\to\infty}(1/n)\sum_{t=1}^{n}\left(\sum_{j=1}^{t} d_{\tau,j}\right)\cos[2k\pi(t - 0.5)/n] = \frac{4\tau + s - 1}{s^2}.$$

Thus, the problem is now similar to the previous problem of unconstrained drift, but the cure is much simpler:

Theorem 8. If we conduct our tests on the basis of moving averages of s adjacent z_t*'s, where s is the number of seasons, using the weight functions* $F_{n,k}$*, then the results of Theorems 1 through 7 carry over to cointegrated systems with seasonal drift.*

By taking these moving averages, the lag polynomial $C(L)$ becomes $C_s(L) = [(1/s)\Sigma_{j=0}^{s-1}L^j]C(L)$. Since $C_s(1) = C(1)$, all our results go through.

8. Empirical and Monte Carlo comparison with Johansen's approach

8.1. The data

In this section we compare our nonparametric approach with Johansen's tests, using the series on the logs of GNP and wages from the Nelson-Plosser (1982) data set, extended by Schotman and Van Dijk (1991) to 1988. These two series were selected on the basis of the test results of the Phillips-Perron (1988), Bierens-Guo (1993) and Bierens (1993) unit root and trend stationarity tests, as well as by inspection of the plots of the time series (see Schotman and Van Dijk, 1991). The two-dimensional vector time series involved has length n = 80 (from 1909 to 1988). As a double check we applied our nonparametric cointegration test to each of the two series: the unit root hypothesis could not be rejected at the 10% significance level.

8.2. Nonparametric cointegration analysis

The result of our nonparametric cointegration analysis is that the null hypothesis of no cointegration $r = 0$) is rejected at the 5% significance level, whereas the null hypothesis $r = 1$ is not rejected at the 10% significance level. Therefore we conclude that ln(wages) and ln(GNP) are cointegrated: $r = 1$. This result is confirmed by the estimation approach in section 4.4: the function $\hat{g}_m(r)$ defined by (26), with $m = 2$, takes the values $\hat{g}_m(0) = 1382.966$, $\hat{g}_m(1) = 3.087$, $\hat{g}_m(2) = 28164.158$, hence the estimated number of cointegrated vectors is 1. The estimate of the standardized cointegrating vector is $(1, -.70)^T$, i.e., ln(wages) − .7ln(GNP) is (trend) stationary.

In order to see how "significant" the estimated cointegrating vector is, we have conducted a series of trace tests (which in this case coincide with the lambda-max tests), for 2×1 matrices $H = (1,a)^T$ with $a \in \{-.4,-.5,-.6.-.65,-.7,-.75,-.8,-.9,-1\}$. The null hypothesis is not rejected at the 10% significance level for a ranging from $-.6$ to $-.8$, and at the 5% level for a ranging from $-.5$ to $-.9$.

8.3. Johansen's approach

Next, we have applied Johansen's ML approach. Our aim is to verify whether the nonparametric approach is capable of producing the same results as Johansen's approach. The reason for taking this approach as the benchmark for the comparison with our nonparametric cointegration analysis is threefold. First, the hypotheses to be tested are about the same. Second, Johansen's method seems to be the most popular one in applied macroeconomic cointegration research, due to its own merits as well as the fact that Johansen has made his approach available in the form of a RATS program. Third, to the best of our knowledge the only other methods available in the (published) literature that can test for the number of cointegrating vectors are the Stock-Watson (1988) and Phillips (1991) methods. The Stock-Watson method, however is closely related to the Johansen method [see Johansen (1991, p.1566)], and Phillips' efficient ECM method has a case-dependent null distribution.

In first instance we have specified the ECM (24) with an intercept, and we have conducted Johansen's lambda-max and trace tests for the number of cointegrating vectors, r, for the cases where: (i) the intercept vector π_0, say, is not proportional to γ, (ii) π_0 is proportional to γ, but this restriction is not imposed, and (iii) the restriction that π_0 is proportional to γ is imposed. This restriction implies that the cointegration relation has an intercept rather than the ECM itself. We conducted Johansen's tests for $p = 2, 4, 6$. The results (at the 5% and 10% significance level) indicate that there is one cointegrating vector, $r = 1$, provided the order p of the VAR model is chosen equal to 6. For the lower values of p the test results were inconclusive, in the sense that the results of the tests were either contradictory or different for the 5% and 10% significance levels. Moreover, the restriction that π_0 is proportional to γ is then rejected at the 5% significance level.

The corresponding estimated standardized cointegrating vector is now $(1,-0.75)^T$, which is significantly different (at the 5% level) from our estimate $(1,-.7)^T$. In order to analyze the difference between the nonparametric and the parametric estimates of the cointegrating vector, we have run three cointegration regressions, without and with intercept, and with intercept and time trend. The nonparametric estimate of a corresponds to the OLS coefficient of ln[GNP] in the regression with intercept and time trend, whereas Johansen's estimate of a corresponds to the regression with intercept only. Therefore we now include an intercept *plus* linear time trend in the ECM (24), say $\pi_{00} + \pi_{01}t$. However, it seems reasonable to impose cointegration restrictions on π_{01}, i.e., we assume that π_{01} is proportional to γ, as otherwise there would be a quadratic trend in z_t, which seems unlikely. In view of the previous result we first specified $p = 6$, but for that case the test results for r were inconclusive. Therefore we next specified $p = 8$, which yields conclusive test results: $r = 1$. In both cases the LR test of the restriction that π_{01} is proportional to γ, given $r = 1$, is not rejected at the 10% level.

The estimation of the cointegrating vector, and the tests of linear restrictions on the cointegrating vector has been based on the ECM with $p = 8$ without imposing the restriction that π_{01} is proportional to γ, because otherwise we have to test these linear restrictions jointly with linear restriction on π_{01}. Cf. Johansen (1994). The estimate involved of the standardized cointegrating vector is now $(1,-.7)^T$, which is in tune with our nonparametric estimate (the difference is only from the third decimal digit onwards).

The above empirical comparison of our nonparametric cointegration analysis with Johansen's approach demonstrates that our approach is capable of giving the same answers regarding the number of cointegrating vectors and the cointegrating vectors themselves as Johansen's ML method, with less effort.

The details of the test results involved are presented in the separate appendix to this paper.

8.4. Monte Carlo comparison

In order to check whether the above results are typical for this data set or not, we have conducted our nonparametric tests and Johansen's tests on 500 replications of ln(wages) and ln(GNP) with sample size $n = 80$, on the basis of the estimated ECM (24) with $p = 8$, and an intercept plus linear trend $\pi_{00} + \pi_{01}t$, where π_{01} is proportional to γ. The first eight observations were taken from the actual data set, and the errors e_t were drawn independently from the bi-variate normal distribution with zero mean vector and variance matrix equal to the estimated variance matrix. All tests are conducted at the 10% significance level. Johansen's tests are conducted for $p = 6$, 8 and 10, in order to check the sensitivity of these tests for the VAR order p, with an intercept and a linear trend included in the ECM, and cointegration restrictions on the trend parameters imposed.

The Monte Carlo results, presented in Table 5, indicate that for this data-generating process our nonparametric test for testing the number of cointegrating vectors performs somewhat better than Johansen's lambda-max test, even for the correct VAR order $p = 8$. The nonparametric test gives in about 79% of all cases the correct answer $r = 1$, whereas the corresponding percentages for Johansen's lambda-max test are 58% if $p = 6$, 70% if $p = 8$ and 56% if $p = 10$. This illustrates the importance of finding the correct order p of the ECM. Moreover, Johansen's test of linear restrictions on the cointegrating vectors suffers more from size distortions than the nonparametric test, although if we would correct for size distortion Johansen's test seems much more powerful.

Table 5: *Acceptance frequencies* (%)

(500 *simulations*, 10% *significance level*)

	Nonpara-metric	*Johansen* $p = 6$	$p = 8$	$p = 10$
$r = 0$	9.6	21.2	14.6	32.2
$r = 1$	79.2	58.0	70.4	56.2
$r = 2$	11.2	20.8	15.0	11.6
Test of $H^T =^{(*)}$				
(1,−0.40)	20.960	10.345	5.398	4.626
(1,−0.50)	40.152	17.241	11.648	8.897
(1,−0.60)	63.384	40.000	25.568	21.352
(1,−0.65)	75.758	67.586	51.705	37.722
(1,−0.70)	85.859	80.690	68.466	55.872
(1,−0.75)	90.404	42.414	43.182	39.858
(1,−0.80)	90.909	15.517	15.625	18.149
(1,−0.90)	82.071	1.379	1.705	4.626
(1,−1.00)	66.414	0.690	1.420	3.203

(∗) *only for the cases with test result* $r = 1$

We recall that the choice of the parameter *m* in the case of our nonparametric test of parametric restrictions on the cointegrating vectors is somewhat heuristic: we proposed the rule of thumb $m = 2q$. Cf. section 6.3. Therefore, we also have conducted Monte Carlo simulations of these tests for the cases $m = 2q + 1 = 5$ through $5q = 10$, in order to check their sensitivity for the choice of *m*. The results were close to the ones in Table 5. The choice of the parameter *m* in the case of the lambda-min test of the number of cointegrating vectors is much less ambiguous [cf. section 4.2], i.e., the choice of *m* according to Table 1 is just an integral part of the test procedure, so that there is no need for further Monte Carlo simulations to check the sensitivity of the lambda-min test for the choice of *m*.

Of course, this Monte Carlo analysis does not provide sufficient evidence that the nonparametric approach always works better than Johansen's approach. Some preliminary Monte

Carlo simulations by Van Giersbergen (1994) and the author for a class of bivariate cointegrated systems indicate that the small sample power of the nonparametric lambda-min test may be quite poor compared with Johansen's lambda-max test if the fit of the cointegrating regression is low. In that case a full parametric approach may do a much better job than the nonparametric approach.

8.5. Concluding remarks

The above comparison of our nonparametric cointegration analysis with Johansen's ML approach shows that our approach may be a useful addition to the menu of cointegration tests. However, it should be stressed that our approach cannot completely replace Johansen's approach, because the latter provides additional information, in particular regarding possible cointegration restrictions on the drift parameters, and the presence of linear trends in the cointegration relations. Moreover, if one wishes to forecast a cointegrated process or wants to conduct innovation response analysis (cf. Lutkepohl and Saikkonen, 1995), then Johansen's approach seems the only way to go. Thus, rather than being substitutes, the two approaches are complements. In particular, we recommend to conduct our nonparametric cointegration analysis as a preliminary step in specifying an ECM. Deviation of Johansen's test and estimation results from the corresponding nonparametric results may indicate misspecification of the ECM.

References

Anderson,S.A., H.K.Brons and S.T.Jensen, 1983, Distribution of eigenvalues in multivariate statistical analysis, Annals of Statistics 11, 392-415.

Bierens,H.J., 1993, Higher-order sample autocorrelations and the unit root hypothesis, Journal of Econometrics 57, 137-160.

Bierens,H.J., 1994, Topics in advanced econometrics: estimation, testing, and specification of cross-section and time series models (Cambridge, U.K.: Cambridge University Press).

Bierens,H.J. and S.Guo, 1993, Testing stationarity and trend stationarity against the unit root hypothesis, Econometric Reviews 12, 1-32.

Billingsley, P., 1968, Convergence of probability measures (New York: John Wiley).

Boswijk,H.P., 1993, Efficient inference on cointegrating parameters in structural error correction models, to appear in the Journal of Econometrics (Annals).

Boswijk,H.P., 1994, Testing for an unstable root in conditional and structural error correction models", Journal of Econometrics 63, 37-60.

Engle,R.F., 1987, On the theory of cointegrated economic time series, Invited paper presented at the Econometric Society European Meeting 1987, Copenhagen.

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

Engle,R.F. and S.B.Yoo, 1987, Forecasting and testing in cointegrated systems, Journal of Econometrics 35, 143-159.

Engle,R.F. and S.B.Yoo, 1989, Cointegrated economic time series: a survey with new results, mimeo, Department of Economics, University of California, San Diego.

Granger,C.W.J., 1981, Some properties of time series and their use in econometric model specification, Journal of Econometrics 16, 121-130.

Hall,P. and C.C.Heyde, 1980, Martingale limit theory and its applications (San Diego: Academic Press).

Hamming,R.W., 1973, Numerical methods for scientists and engineers (New York: Dover Publications).

Harris, D., 1995, Aspects of estimation of cointegrated time series, Ph.D. Dissertation, Department of Econometrics, Monash University, Australia.

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

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

Johansen,S., 1994, The role of the constant and linear terms in cointegration analysis of nonstationary variables, Econometric Reviews 13, 205-230.

Johansen,S. and K.Juselius, 1990, Maximum likelihood estimation and inference on cointegration: with applications to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169-210.

Lutkepohl,H. and P.Saikkonen, 1995, Impulse response analysis in cointegrated processes, mimeo, Department of Economics, Humboldt University Berlin.

Nelson,C.R. and C.I.Plosser, 1982, Trends and random walks in macroeconomic time series, Journal of Monetary Economics 10, 139-162.

Park,J.Y., 1990, Testing for unit root and cointegration by variable addition, Advances in Econometrics 8, 107-133.

Perron,P. and J.Y.Campbell, 1993, A note on Johansen's cointegration procedure when trends are present, Empirical Economics 18, 777-789.

Phillips,P.C.B., 1987, Time series regression with unit roots, Econometrica 55, 277-302.

Phillips,P.C.B., 1991, Optimal inference in cointegrated systems, Econometrica 59, 283-306.

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

Phillips,P.C.B. and P.Perron, 1988, Testing for a unit roots in time series regression, Biometrika 75, 335-346.

Phillips,P.C.B. and V.Solo, 1992, Asymptotics for linear processes, Annals of Statistics 20, 971-1001.

Schotman,P.C. and H.K.Van Dijk, 1991, On Bayesian routes to unit roots, Journal of Applied Econometrics 6, 387-401.

Stock,J.H. and M.W.Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097-1107.

Sims,C.A., J.H.Stock and M.W.Watson, 1990, Inference in linear time series models with some unit roots, Econometrica 58, 113-144.

Toda, H.Y., 1994, Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present, Review of Economics and Statistics 76, 66-79.