COINTEGRATION ANALYSIS

Herman J. Bierens[1]
Pennsylvania State University

April 7, 2010

1. Introduction

1.1 What is cointegration?

The basic idea behind cointegration is that if all the components of a vector time series process z_t have a unit root, or in other words, if z_t is a multivariate $I(1)$ process, there may exist linear combinations $\beta^T z_t$ without a unit root. These linear combinations may then be interpreted as long term relations between the components of z_t, or in economic terms as static equilibrium relations.

For bivariate economic $I(1)$ processes, cointegration often manifests itself by more or less parallel shapes of the plots of the two series involved. Figure 1 displays a typical example of such a pair of cointegrated economic time series, namely the log of nominal income (upper curve) and the log of nominal consumption (lower curve) in Sweden[2] from 1861 to 1988.

According to Friedman's (1957) permanent income theory, the long run marginal propensity to consume from permanent income should be close to one. With the logs of consumption and income being unit root with drift processes, the modern interpretation of the permanent income hypothesis therefore is that the difference of the logs of consumption and income is stationary: $\beta = (1,-1)^T$. However, income in Friedman's theory is net income rather than gross income, so that the long run marginal propensity to consume from gross income might be less than one. Anyhow, Friedman's theory predicts that the logs of consumption and income are cointegrated. The time series displayed in Figure 1 will be used in an empirical application in section 6.

---

[1] This paper is an updated and extended version of Bierens (1997b).

[2] I like to thank Philip Hans Franses for providing me with this data set. The original sources of these time series are Krantz and Nilson (1975) and Melander, Vredin and Warne (1992).

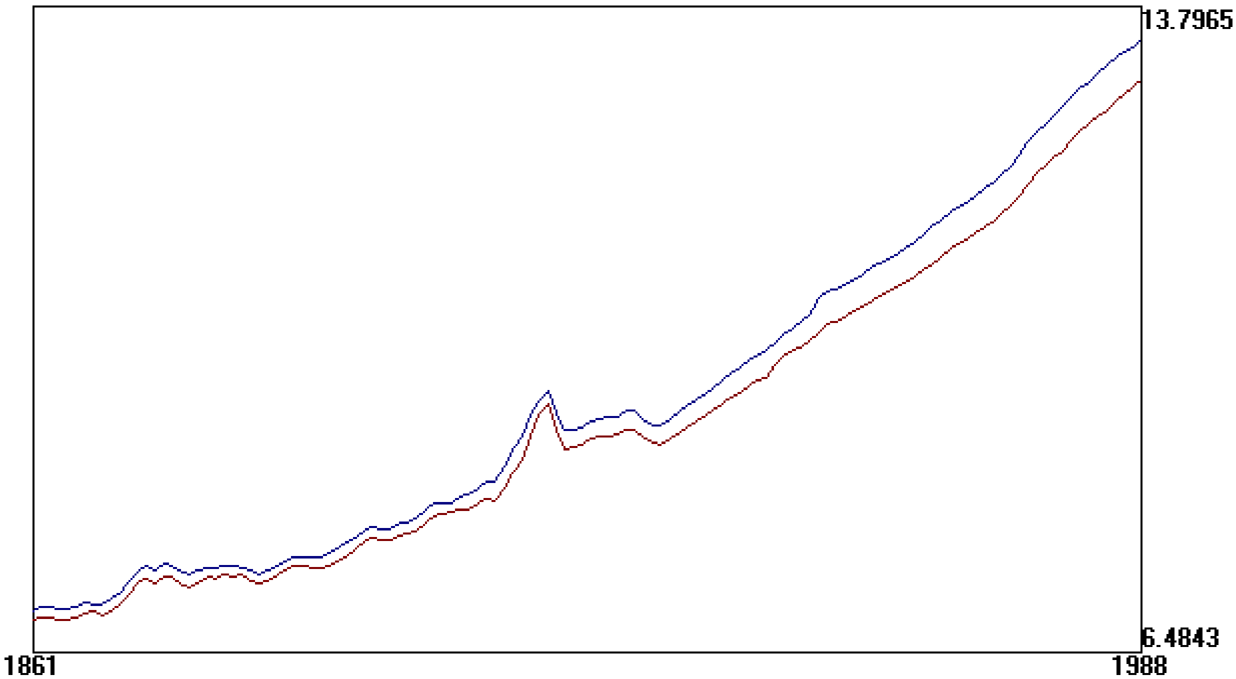

Figure 1: *Logs of Income and Consumption in Sweden*

1.2 The literature on cointegration

The concept of cointegration was first introduced by Granger (1981) and elaborated further by Engle and Granger (1987), Engle and Yoo (1987, 1991), Phillips and Ouliaris (1990), Stock and Watson (1988), Phillips (1991) and Johansen (1988, 1991, 1994), among others.

Working in the context of a bivariate system with at most one cointegrating vector, Engle and Granger (1987) propose to estimate the cointegrating vector $\beta = (1,\beta_2)^T$ by regressing the first component $z_{1,t}$ of z_t on the second component $z_{2,t}$, using OLS (which is called the cointegrating regression), and then testing whether the OLS residuals of this regression have a unit root, using the Augmented Dickey-Fuller (ADF) test. See Fuller (1976), Dickey and Fuller (1979, 1981) and Said and Dickey (1984) for the latter. However, since the ADF test is conducted on estimated residuals, the tables of the critical values of this test in Fuller (1976) do not apply anymore. The correct critical values involved can be found in Engle and Yoo (1987). Phillips and Ouliaris' (1990) tests are also based on these residuals, but instead of using the ADF test for testing the presence of a unit root they use further elaborations of the Phillips (1987) and Phillips-Perron

(1988) unit root tests. Both types of tests have absence of cointegration as the null hypothesis. Park (1990) proposes a test for unit root and cointegration using the variable addition approach, by regressing the OLS residuals of the cointegrating regression on powers of time and testing whether the coefficients involved are jointly zero. The same idea has been used by Bierens and Guo (1993) to test (trend) stationarity against the unit root hypothesis. However, also Park's approach requires consistent estimation of the long-run variance of the errors of the true cointegrating regression by a Newey-West (1987) type estimator, which sacrifices a substantial amount of asymptotic power of the test. Cf. Bierens and Guo (1993) for the latter. Also the tests of Hansen (1992) and Park (1992) are based on a single cointegrating regression, and both tests employ variants of the instrumental variables estimation method of Phillips and Hansen (1990). Finally, Boswijk (1994, 1995) links the single-equation and system approaches by using structural single-equations as a basis for cointegration analysis.

The above approaches test the null or alternative hypothesis of absence of cointegration, but if the tests indicate the presence of cointegration in systems with three variables or more we still don't know how many linear independent cointegrating vectors there are. In such cases one may use the approach of Stock and Watson (1988), which is a multivariate extension of the Engle-Granger and Phillips-Ouliaris tests. The basic idea is to linearly transform the q-variate cointegrated process z_t with say r linear independent cointegrating vectors such that the first r components of the transformed z_t are stationary and the last $q-r$ components, stacked in a vector w_t, say, are integrated. The transformation matrix involved can be consistently estimated using principal components of z_t. Then test whether w_t is a $q-r$ variate unit root process, using a multivariate version of the ADF test or the Phillips (1987) test. The critical values of this test differ according to whether the initial value z_0 is non-zero or not and whether the unit root process z_t has drift or not.

In a series of influential papers, Johansen (1988, 1991) and Johansen and Juselius (1990) propose an ingenious and practical full maximum likelihood estimation and testing approach, based on the following *V*ector *E*rror *C*orrection *M*odel (hereafter indicted by ECM) for the q-variate unit root process z_t:

$$\Delta z_t = \Pi_0 d_t + \sum_{j=1}^{p-1}\Pi_j \Delta z_{t-j} + \alpha\beta^T z_{t-p} + e_t. \tag{1}$$

Here $\Delta z_t = z_t - z_{t-1}$, d_t is a vector of deterministic variables, such as a constant and seasonal dummy variables, the Π_j, $j > 0$, are $q \times q$ and β and α are $q \times r$ parameter matrices, where β and α are of full column rank, with r the number of linear independent cointegrating vectors (the columns of β), the e_t are i.i.d. $N_q(0,\Sigma)$ errors, and $\det(I - \Sigma_{j=1}^{p-1} \Pi_j L^j)$ has all its roots outside the complex unit circle. Note that if $r = q$, so that then the matrix $\alpha\beta^T$ is of full rank, and if $d_t = 1$, then model (1) generates a stationary AR(p) process z_t.

The VECM (1) 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 step-wise concentrating all the parameter matrices in the likelihood function out, except the matrix β, Johansen shows that the maximum likelihood estimator of β can be derived as the solution of a generalized eigenvalue problem. Likelihood ratio tests of hypotheses about the number of cointegrating vectors can then be based on these eigenvalues. Moreover, Johansen (1988) also proposes likelihood ratio tests for linear restrictions on the cointegrating vectors.

Initially, Johansen (1988) considered the case where d_t is absent. Later on, Johansen (1991) extended his approach to the case where d_t contains an intercept and seasonal dummy variables, and in Johansen (1994) also a time trend in d_t (but no seasonal dummy variables) is allowed. These three cases lead to different null distributions of the likelihood ratio tests of the number of cointegrating vectors. Moreover, also possible restrictions on the vector of intercepts or the vector of trend coefficients may lead to different null distributions. Thus, application of Johansen's tests actually requires some a priori knowledge about the true parameters of the VECM (1).

Phillips' (1991) efficient error correction modeling approach differs from that of Johansen (1988) in that Phillips specifies the VECM directly on the basis of the cointegrating relations $z_{1,t} = Bz_{2,t} + u_t$, with u_t a stationary zero mean Gaussian process, leading to an VECM of the form

$$\Delta z_t = \begin{pmatrix} I_r & -B \\ O & O \end{pmatrix} z_{t-t} + v_t, \tag{2}$$

where r is the number of cointegrating relations and v_t is a stationary Gaussian process with long run variance matrix $\Omega = \lim_{n\to\infty} Var[(1/\sqrt{n})\Sigma_{t=1}^{n} v_t]$. Phillips shows that under the i.i.d. assumption on v_t the maximum likelihood estimator of B is efficient, and that this efficiency carries over to the case with dependent errors v_t if B is estimated by maximum likelihood on the basis of model (2) with i.i.d. N(0,Ω) errors v_t, provided Ω is replaced by a consistent estimator. In contrast with Johansen's maximum likelihood method, however, Phillips' efficient maximum likelihood approach has not yet been widely applied in empirical research, probably due to the fact that the limiting distribution of the maximum likelihood estimator of the matrix B depends on the long run variance matrix Ω.

The Stock and Watson (1988), Phillips (1991) and Johansen (1988, 1991, 1994) approaches require consistent estimation of nuisance and/or structural parameters. In Bierens (1997a) I have proposed 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. My 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 one can consistently estimate a basis of the space of cointegrating vectors, using the eigenvectors of the generalized eigenvalue problem involved. However, the two matrices involved are constructed independently of the data generating process on the basis of weighted means of z_t and Δz_t, respectively, where the weights involved are Chebishev time polynomials [cf. Hamming (1973)] of even order.

1.3 Contents

In these lecture notes I will review some new developments in cointegration analysis, in particular Johansen's (1988, 1991, 1994) maximum likelihood approach on the basis of the VECM (1), and my nonparametric cointegration approach. First, in section 2, I will explain in more detail what cointegration is about, and in section 3 I will discuss (in an informal way) the Granger representation theorem that gives rise to the VECM specification (1). In sections 4 and 5 I will review Johansen's and my nonparametric approach, respectively. The main reason for focusing on Johansen's approach is that it is presently the most popular one in empirical macroeconomic cointegration research, due to its own merits as well as the fact that Johansen's approach is now available in most time series oriented econometric software packages. Finally, in section 6 I will apply both the Johansen approach and my nonparametric approach to the Swedish data on the logs of consumption and income displayed in Figure 1.

2. Introduction to cointegration

Consider the q-variate unit root process $z_t = z_{t-1} + u_t$, where u_t is a zero mean stationary process, and let z_t be observable for $t = 0,1,2,..,n$. Due to the Wold decomposition theorem, we can write (under some mild regularity conditions), $u_t = C(L)v_t$, where v_t is a q-variate stationary white noise process with unit variance, i.e.,

$$E[v_t] = 0,\ E[v_t v_t^T] = I_q,\ E[v_t v_{t-j}^T] = O \text{ for } j \neq 0, \tag{3}$$

and $C(L)$ is a $q \times q$ matrix of lag series: $C(L) = \sum_{k=0}^{\infty} C_k L^k$, where L is the lag operator. Since by construction the lag polynomial $C(L) - C(1)$ is zero at $L = 1$, we can write

$$C(L)v_t = C(1)v_t + (C(L) - C(1))v_t = C(1)v_t + (1-L)D(L)v_t, \tag{4}$$

where

$$D(L) = \sum_{k=0}^{\infty} D_k L^k = (C(L) - C(1))/(1-L). \tag{5}$$

Denoting $w_t = D(L)v_t$ we now have $u_t = C(1)v_t + w_t - w_{t-1}$, hence

$$z_t = z_0 - w_0 + w_t + C(1)\sum_{j=1}^{t} v_j. \tag{6}$$

If v_t is a Gaussian process then by the white noise assumption (3) the v_t's are i.i.d. $N_q(0,I_q)$. Since Johansen's approach is based on Gaussian maximum likelihood theory, this normality condition will be assumed. Moreover, we need regularity conditions that ensure that u_t and w_t are stationary processes. Therefore, it will be assumed that:

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

Note that this condition on $C(L)$ implies that ΣC_k, $\Sigma C_k C_k^T$, ΣD_k, and $\Sigma D_k D_k^T$ converge, so that together with the normality condition, it follows that u_t and w_t are stationary Gaussian processes.

Cf. Engle and Yoo (1991). Moreover, Assumption 1 excludes the usual condition that also $\det(C_2(L))$ has roots all outside the unit circle. This is necessary because for cointegration we need to allow the matrix $C(1)$ to be singular.

As far as the nonparametric cointegration approach is concerned, Assumption 1 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, in the nonparametric cointegration case 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).

Now if $\mathrm{rank}(C(1)) = q - r < q$ then the process z_t is cointegrated: there exist r linear independent cointegrating vectors β_j, $j = 1,..,r$, say, such that $\beta_j^T C(1) = 0^T$, hence it follows from (6) that $\beta_j^T z_t = \beta_j^T(z_0 - w_0) + \beta_j^T w_t,\ j = 1,....,r$. Thus the $\beta_j^T z_t$'s are now asymptotically stationary processes, in the sense that the stochastic intercept $\beta_j^T(z_0 - w_0)$ becomes independent of w_t if t approaches infinity, so that we then may condition on $z_0 - w_0$ and treat it as a constant.

The factorization (4) can be applied to the matrix $D(L)$ as well, so that similarly to (4), $C(L)v_t = C(1)v_t + D(1)(1-L)v_t + (1-L)^2 G(L)v_t$, where $G(L) = (D(L) - D(1))/(1-L)$. However, if there would exist a cointegrating vector β such that $\beta^T D(1) = 0^T$ then, with $\varepsilon_t = G(L)v_t$ a stationary process, we would have $\beta^T u_t = \Delta^2 \beta^T \varepsilon_t$, hence $\sum_{j=1}^t \beta^T z_t = \beta^T(z_0 - \varepsilon_0 + \varepsilon_{-1})t + \beta^T \varepsilon_t$ is trend stationary. As we will see in the next section, this would violate one of the conditions for the existence of an autoregressive error correction representation of a cointegrated system. Also, we need to exclude this case for the nonparametric cointegration approach. Therefore I assume:

Assumption 2: *Let R_r be the matrix of 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.*

3. The error correction form of a cointegrated system

Following the approach of Engle and Yoo (1991) I will show now that under some regularity conditions a cointegrated process can be modelled as an VECM of the type (1). This result is due to Granger. Cf. Engle and Granger (1987). For convenience the discussion will be confined to the bivariate case ($q = 2$) with one cointegrating vector.

Let β be the cointegrating vector. Without loss of generality we may normalize $\beta = (1, \beta_2)^T$. Consider the matrices

$$\Phi = \begin{pmatrix} 1 & \beta_2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \beta^T \\ \varphi_2^T \end{pmatrix},\ V^{-1}(L) = \begin{pmatrix} \beta^T D(L) \\ \varphi_2^T C(L) \end{pmatrix},$$
$$M(L) = \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix},\ M^*(L) = \begin{pmatrix} 1 & 0 \\ 0 & 1-L \end{pmatrix}. \tag{7}$$

Then

$$\Phi\Delta z_t = \begin{pmatrix} (1-L)\beta^T D(L) \\ \varphi_2^T C(L) \end{pmatrix} v_t = M(L)V^{-1}(L)v_t. \tag{8}$$

and $M^*(L)M(L) = (1-L)I_2$. Next, assume that $V^{-1}(L)$ is invertible with inverse $V(L)$. This assumption is related to Assumption 2: if Assumption 2 does not hold then $V^{-1}(1)$ is singular so that $V^{-1}(L)$ is not invertible. Furthermore, denote $A(L) = V(L)M^*(L)\Phi$. Then $(1-L)A(L)z_t = (1-L)v_t$, which yields the AR form of the model:

$$A(L)z_t = \mu_0 + v_t, \tag{9}$$

where $\mu_0 = A(L)z_0 - v_0$. Now observe that

$$A(1) = V(1)M^*(1)\Phi = V(1)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & \beta_2 \\ 0 & 1 \end{pmatrix} = V(1)\begin{pmatrix} 1 & \beta_2 \\ 0 & 0 \end{pmatrix} = \gamma_1\beta^T, \tag{10}$$

where γ_1 is the first column of $V(1)$. Moreover, similarly to (4) we can write

$$A(L) = A(1)L + (1-L)B(L). \tag{11}$$

Combining (9), (10) and (11) we get the VECM $B(L)\Delta z_t = \mu_0 - \gamma_1\beta^T z_{t-1} + v_t$. Finally, assume that $B(0)$ is invertible and that $B(L)$ is a $(p-1)$-order lag polynomial, so that we may write $\Pi(L) = B(0)^{-1}B(L) = I - \Pi_1 L - \ldots \Pi_{p-1}L^{p-1}$. Denoting $\pi_0 = B(0)^{-1}\mu_0$, $\alpha = -B(0)^{-1}\gamma_1$, and $e_t = B(0)^{-1}v_t$, we get the VECM

$$\Delta z_t = \pi_0 + \Sigma_{j=1}^{p-1}\Pi_j\Delta z_{t-j} + \alpha\beta^T z_{t-1} + e_t. \tag{12}$$

.

Note that the lag of the level variable z_{t-1} does not matter. Without loss of generality we may replace $\Pi(L)$ by the lag polynomial $\Pi_*(L) = \Pi(L) - \Sigma_{j=1}^{p-1}\alpha\beta^T L^j$, which yields an VECM of the form (1) with $d_t = 1$.

4. Johansen's maximum likelihood approach

4.1 Introduction

Johansen's cointegration testing approach is based on maximum likelihood estimation and likelihood ratio testing of the VECM (1), by step-wise concentrating the parameters out (i.e., maximizing the likelihood function over a subset of parameters, treating the other parameters as known), given the number r of cointegrating vectors, where the matrix β is the last to be concentrated out. Denoting the concentrated likelihood, given r and β, by $\hat{L}(r,\beta)$, and the maximum likelihood estimator of β given r by $\hat{\beta}_r$, where β and its maximum likelihood estimate are interpreted as zero vectors if $r = 0$, Johansen proposes two tests for the number of cointegrated vectors, namely the likelihood ratio test $-2\ln\left(\hat{L}(r,\hat{\beta}_r)/\hat{L}(r+1,\hat{\beta}_{r+1})\right)$ of the null hypothesis that there are r cointegrated vectors (for $r = 0,1,..,q-1$) against the alternative that there are r+1 cointegrating vectors, and the likelihood ratio test $-2\ln\left(\hat{L}(r,\hat{\beta}_r)/\hat{L}(q,\hat{\beta}_q)\right)$ of the same null hypothesis against the alternative that there are q cointegrated vectors. The latter alternative corresponds to the case where β is square and of full rank, which in its turn corresponds to the case that z_t is stationary rather than a multivariate unit root process. Since the usual regularity condition for maximum likelihood estimation do not apply in this case, the likelihood ratio tests involved have nonstandard limiting null distributions. Moreover, given the number r of cointegrating vectors, Johansen also proposes a likelihood ratio test of parametric restrictions on β of the form $\beta = H\varphi$, where H is a given $q \times s$ matrix of rank $s \leq r$ and φ is an unrestricted $s \times r$ matrix. For example, in the case $r = 1$, $q = 2$, one might wish to test whether β^T is proportional to $(1,-1) = H^T$. The likelihood ratio test statistic

$$-2\ln[\sup_{\varphi}\hat{L}(r,H\varphi)/\hat{L}(r,\hat{\beta}_r)] \tag{13}$$

involved has a limiting χ^2 null distribution with $r(q-s)$ degrees of freedom.

4.2 The lambda-max and trace tests

I will now illustrate how Johansen's cointegration tests are conducted for the case where the data generating process is a Gaussian VECM of the form (1) with $d_t = 1$ and $p = 2$, where z_t is observable for $t = -1,0,..,n$:

$$\Delta z_t = \pi_0 + \Pi_1 \Delta z_{t-1} + \alpha\beta^T z_{t-2} + e_t, \quad e_t \sim i.i.d.\ N_q(0,\Sigma). \tag{14}$$

Given β, α and Σ, the maximum likelihood estimates of π_0 and Π_1 can be obtained simply by regressing $\Delta z_t - \alpha\beta^T z_{t-2}$ on an intercept 1 and Δz_{t-1}, using OLS. The residuals of this regression are $\hat{R}_{1,t} - \alpha\beta^T\hat{R}_{2,t}$, where $\hat{R}_{1,t}$ is the residual of the regression of Δz_t on 1 and Δz_{t-1}, and $\hat{R}_{2,t}$ is the residual of the regression of z_{t-2} on 1 and Δz_{t-1}. Now the log-likelihood function with π_0 and Π_1 concentrated out is of the form

$$-\ 0.5\,n\ \ln(\det\Sigma) - 0.5\Sigma_{t=1}^n (\hat{R}_{1,t} - \alpha\beta^T\hat{R}_{2,t})^T\Sigma^{-1}(\hat{R}_{1,t} - \alpha\beta^T\hat{R}_{2,t}) + \text{rest}, \tag{15}$$

where "rest" stands for the terms that do not depend on parameters. Similarly, we can concentrate α out, given β and Σ, by regressing $\hat{R}_{1,t}$ on $\beta^T\hat{R}_{2,t}$, which yields the estimate

$$\hat{\alpha}(\beta) = \hat{S}_{1,2}\beta[\beta^T\hat{S}_{2,2}\beta]^{-1}, \tag{16}$$

where $\hat{S}_{i,j} = (1/n)\Sigma_{t=1}^n \hat{R}_{i,t}\hat{R}_{j,t}^T, \quad i,j = 1,2$. Next, concentrate Σ out, given β, by substituting the well-known maximum likelihood estimator of the variance matrix of a normal distribution with zero mean vector:

$$\hat{\Sigma}(\beta) = (1/n)\Sigma_{t=1}^n\left(\hat{R}_{1,t} - \hat{\alpha}(\beta)\beta^T\hat{R}_{2,t}\right)\left(\hat{R}_{1,t} - \hat{\alpha}(\beta)\beta^T\hat{R}_{2,t}\right)^T = \hat{S}_{1,1} - \hat{S}_{1,2}\beta\left(\beta^T\hat{S}_{2,2}\beta\right)^{-1}\beta^T\hat{S}_{2,1}. \tag{17}$$

Thus, the concentrated log-likelihood now becomes

$$\ln\left(\hat{L}(r,\beta)\right) = -0.5\,n\,\ln\left(\det\hat{\Sigma}(\beta)\right) + \text{rest}, \tag{18}$$

hence the maximum likelihood estimator of β is found by solving the minimization problem

$$\min\ \det\left(\hat{S}_{1,1} - \hat{S}_{1,2}\beta(\beta^T\hat{S}_{2,2}\beta)^{-1}\beta^T\hat{S}_{2,1}\right), \tag{19}$$

where the minimum is taken over all $q \times r$ matrices β. Using the matrix equalities

$$\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} = \begin{pmatrix} A & O \\ B^T & I \end{pmatrix}\begin{pmatrix} I & A^{-1}B \\ O & C - B^T A^{-1}B \end{pmatrix} = \begin{pmatrix} I & B \\ O & C \end{pmatrix}\begin{pmatrix} A - BC^{-1}B^T & O \\ C^{-1}B^T & I \end{pmatrix},$$

where A and C are nonsingular square matrices, it is a standard exercise to verify that the minimization problem (19) is equivalent to

$$\min\ \det\left(\beta^T\hat{S}_{2,2}\beta - \beta^T\hat{S}_{2,1}\hat{S}_{1,1}^{-1}\hat{S}_{1,2}\beta\right)\det(\hat{S}_{1,1})/\det\left(\beta^T\hat{S}_{2,2}\beta\right). \tag{20}$$

Note that the solution involved is not unique, as we may freely multiply β by a conformable nonsingular matrix. It is now quite easy to recognize the minimization problem (20) as a generalized eigenvalue problem: let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_q$ be the ordered solutions of the generalized eigenvalue problem $\det(\lambda\hat{S}_{2,2} - \hat{S}_{2,1}\hat{S}_{1,1}^{-1}\hat{S}_{1,2}) = 0$, let $\hat{B} = (\hat{b}_1,...,\hat{b}_q)$ be the matrix of corresponding eigenvectors, normalized such that $\hat{B}^T\hat{S}_{2,2}\hat{B} = I_q$, and choose $\beta = \hat{B}\xi$, where ξ is a $q \times r$ matrix normalized such that $\xi^T\xi = I_r$. Then the minimization problem (20) becomes

$$\begin{aligned}
&\min_{\xi^T\xi = I_r} \det\left(\xi^T\hat{B}^T\hat{S}_{2,2}\hat{B}\xi - \xi^T\hat{B}^T\hat{S}_{2,1}\hat{S}_{1,1}^{-1}\hat{S}_{1,2}\hat{B}\xi\right)\det(\hat{S}_{1,1})/\det\left(\xi^T\hat{B}^T\hat{S}_{2,2}\hat{B}\xi\right) \\
&= \min_{\xi^T\xi = I_r} \det\left(\xi^T\xi - \xi^T\hat{B}^T\hat{S}_{2,1}\hat{S}_{1,1}^{-1}\hat{S}_{1,2}\hat{B}\xi\right)\det(\hat{S}_{1,1})/\det\left(\xi^T\xi\right) \\
&= \min_{\xi^T\xi = I_r} \det\left(I_r - \xi^T\hat{B}^T\hat{S}_{2,1}\hat{S}_{1,1}^{-1}\hat{S}_{1,2}\hat{B}\xi\right)\det(\hat{S}_{1,1})/\det\left(I_r\right) \\
&= \min_{\xi^T\xi = I_r} \det\left(I_r - \xi^T\hat{\Lambda}\xi\right)\det(\hat{S}_{1,1})
\end{aligned} \tag{21}$$

where $\hat{\Lambda} = \mathrm{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_q)$. Clearly, the solution of (21) is $\xi^T = (I_r, O)$, hence the maximum likelihood estimator $\hat{\beta}_r$ of β, given the number r of cointegrating vectors, is equal to the matrix of the first r columns of $\hat{B}$: $\hat{\beta}_r = (\hat{b}_1, ..., \hat{b}_r)$. Moreover, $\det[\hat{\Sigma}(\hat{\beta}_r)] = \det(I_r - \hat{\Lambda}_r)\det(\hat{S}_{1,1})$, where $\hat{\Lambda}_r = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_r)$, so that the maximum log-likelihood given r becomes:

$$\ln\left(\hat{L}(r,\hat{\beta}_r)\right) = -0.5\, n\, \Sigma_{i=1}^{r}\ln(1-\hat{\lambda}_i) - .5\, n \ln[\det(\hat{S}_{1,1})] + \mathit{rest}. \tag{22}$$

Thus, the likelihood ratio test $-2\ln(\hat{L}(r,\hat{\beta}_r)/\hat{L}(r+1,\hat{\beta}_{r+1}))$ of the null hypothesis that there are r cointegrated vectors against the alternative that there are r+1 cointegrating vectors becomes $-n\ln(1-\hat{\lambda}_{r+1}) \approx n\hat{\lambda}_{r+1}$, and the likelihood ratio test $-2\ln(\hat{L}(r,\hat{\beta}_r)/\hat{L}(q,\hat{\beta}_q))$ of the same null hypothesis against the alternative that there are q cointegrated vectors becomes $-n\Sigma_{i=1+r}^{q}\ln(1-\hat{\lambda}_i) \approx n\Sigma_{i=r+1}^{q}\hat{\lambda}_i$. Johansen (1988, 1991) proves that under the null of r cointegrating vectors, $(\hat{\lambda}_1, ..., \hat{\lambda}_r)^T$ converges in probability to a vector of constants between zero and one, and $n(\hat{\lambda}_{r+1}, ..., \hat{\lambda}_q)^T$ converges in distribution to the vector of ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_{q-r}$ of a stochastic a.s. positive definite $(q-r) \times (q-r)$ matrix which components are functionals of a $q-r$-variate standard Brownian motion. Therefore, the likelihood ratio test $-2\ln(\hat{L}(r,\hat{\beta}_r)/\hat{L}(r+1,\hat{\beta}_{r+1})) \approx n\hat{\lambda}_{r+1}$ is called the *lambda-max* test, and the likelihood ratio test

$-2\ln(\hat{L}(r,\hat{\beta}_r)/\hat{L}(q,\hat{\beta}_q)) \approx n\Sigma_{i=r+1}^{q}\hat{\lambda}_i$ is called the *trace* test.

4.3 Testing parametric restrictions on the cointegrating vectors

Similarly to (22) it can be shown that under the null hypothesis $\beta = H\varphi$, where H is a given $q \times s$ matrix of rank $s \leq r$ with r given, and φ an unrestricted $s \times r$ matrix, the log-likelihood is $-.5n\sum_{i=1}^{s}\ln[1 - \tilde{\lambda}_i] - .5n\ln[\det(H^T\hat{S}_{1,1}H)]$ + rest, where $\tilde{\lambda}_1 \geq .... \geq \tilde{\lambda}_s$ are the solutions of the generalized eigenvalue problem $\det(\lambda H^T\hat{S}_{2,2}H - H^T\hat{S}_{2,1}\hat{S}_{1,1}^{-1}\hat{S}_{1,2}H) = 0$ and the rest term is the same as in (22). Thus, the likelihood ratio test statistic involved is:

$$-2\ln(LR) = n\sum_{i=1}^{s}\ln(1 - \tilde{\lambda}_i) - n\sum_{i=1}^{r}\ln(1 - \hat{\lambda}_i) + n\ln[\det(H^T\hat{S}_{1,1}H)] - n\ln[\det(\hat{S}_{1,1})].$$

Johansen (1988, 1991) proved that this likelihood ratio test has a χ^2 null distribution with $r(q-s)$ degrees of freedom.

4.4 Cointegrating restrictions on the intercept parameters

The null distributions of the lambda-max and trace tests in the above case depends on whether the vector π_0 of intercept parameters in model (14) can be written as

$$\pi_0 = \alpha\delta, \quad with\ \delta \in \mathbb{R}^r, \tag{23}$$

or not. If so, the VECM (14) becomes $\Delta z_t = \Pi_1\Delta z_{t-1} + \alpha[\delta + \beta^T z_{t-2}] + e_t$.

Proposition 1: *Under the restriction* (23) , $\delta + \beta^T z_{t-2}$ *is a zero-mean stationary process*, *hence* Δz_t *is then a zero-mean stationary process, so that* z_t *itself is a multivariate unit root process* ***without drift***.

Proof: Recall that the general VECM

$$\Delta z_t = \pi_0 + \sum_{j=1}^{p-1}\Pi_j\Delta z_{t-j} + \alpha\beta^T z_{t-p} + e_t \tag{24}$$

is derived from

$$\Delta z_t = \mu + C(L)v_t = \mu + C(1)v_t + w_t - w_{t-1}, \tag{25}$$

where v_t is a bivariate zero-mean white noise process with unit variance matrix, $w_t = D(L)v_t$ with $D(L) = (1-L)^{-1}(C(L)-C(1))$, and μ is the non-zero vector of drift parameters. Moreover, $\beta^T C(1) = 0^T$ and $\beta^T D(1) \neq 0^T$. As to the latter, I will impose the stronger condition that

$$\det(D(1)) \neq 0. \tag{26}$$

Backwards substitution of (25) yields

$$z_t = \mu.t + C(1)\Sigma_{j=1}^{t} v_j + w_t + z_0 - w_0, \tag{27}$$

hence

$$\beta^T z_t = \beta^T \mu.t + \beta^T w_t + \beta^T (z_0 - w_0). \tag{28}$$

The last term in (28) acts as constant, and because $\beta^T z_t$ is stationary around a constant we must have that

$$\beta^T \mu = 0. \tag{29}$$

Thus,

$$\beta^T z_{t-p} = \beta^T w_{t-p} + \beta^T (z_0 - w_0) = \beta^T L^p D(L) v_t + c, \textit{ where } c = \beta^T (z_0 - w_0). \tag{30}$$

Substituting (25) and (30) in (24) yields:

$$\left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)\mu + \left(I - \Sigma_{j=1}^{p-1} \Pi_j L^j\right) C(L) v_t = \pi_0 + \alpha\beta^T L^p D(L) v_t + \alpha c + e_t. \tag{31}$$

This result implies that

$$e_t = C(0) v_t \tag{32}$$

(why?), where

$$\det(C(0)) \neq 0, \tag{33}$$

$$\mu = \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} (\pi_0 + \alpha c) \tag{34}$$

and

$$\left(I - \Sigma_{j=1}^{p-1} \Pi_j L^j\right) C(L) = \alpha\beta^T L^p D(L) + C(0) \tag{35}$$

The latter result implies that

$$C(1) = \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} \alpha\beta^T D(1) + \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} C(0) \tag{36}$$

hence

$$\beta^T C(1) = \beta^T \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} \alpha\beta^T D(1) + \beta^T \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} C(0) = 0 \tag{37}$$

and thus

$$\beta^T \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} \alpha = \frac{-1}{\beta^T D(1) D(1)^T \beta} \beta^T \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} C(0) D(1)^T \beta \tag{38}$$

Due to (26) and (33), the right-hand side of (38) is non-zero:

$$\beta^T \left(I - \Sigma_{j=1}^{p-1} \Pi_j\right)^{-1} \alpha \neq 0. \tag{39}$$

Now suppose that (23) is true. Then (34) becomes

$$\mu = (\delta + c)\left(I - \Sigma_{j=1}^{p-1}\Pi_j\right)^{-1}\alpha. \tag{40}$$

But it follows now from (29) that

$$0 = \beta^T\mu = (\delta + c)\beta^T\left(I - \Sigma_{j=1}^{p-1}\Pi_j\right)^{-1}\alpha, \tag{41}$$

which by (39) implies that $c = -\delta$. Substituting this solution in (40) yields $\mu = 0$. Therefore, imposing the cointegrating restriction (23) in VECM (24) removes the drift! Q.E.D.

Consequently, the restriction (23) should only be imposed if the time series involved seem to run parallel, without drift, like in Figure 2:

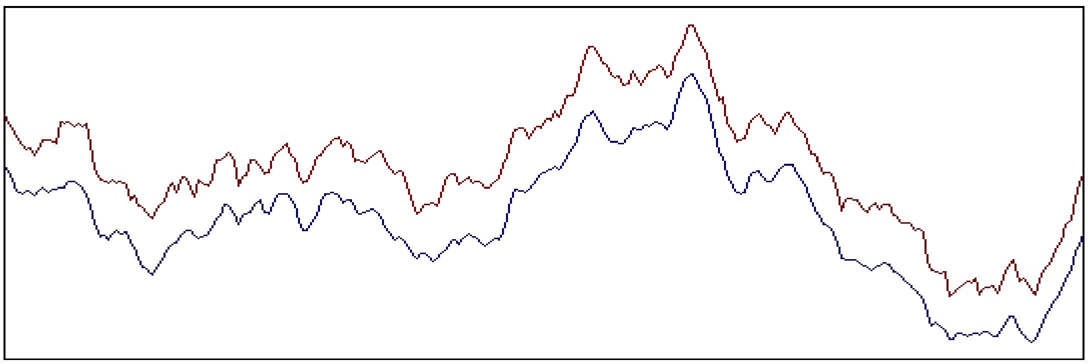

Figure 2 VECM with cointegrating restrictions on the intercept parameters

Without the restriction (23), the vector π_0 of intercept parameters produces common drift in the time series, like in Figure 3:

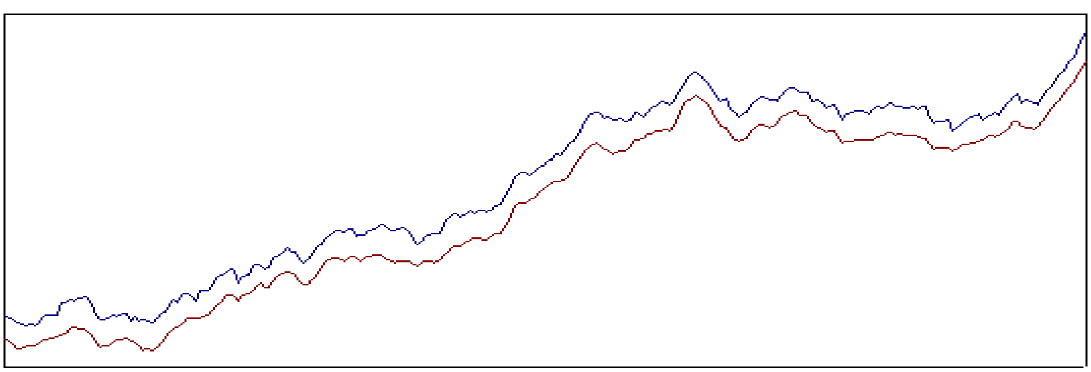

Figure 3 VECM without cointegrating restrictions on the intercept parameters

If we believe that the cointegrating restriction (23) on the intercept parameters holds, we can impose it as follows. First, concentrate Π_1 out, by regressing $\Delta z_t - \alpha(\delta + \beta^T z_{t-2})$ on Δz_{t-1}. The residuals of this regression are $\hat{R}_{1,t} - \alpha(\delta,\beta^T)\hat{R}_{2,t}$, where $\hat{R}_{1,t}$ is now the residual of the regression of Δz_t on Δz_{t-1} alone, and $\hat{R}_{2,t}$ is now the residual of the regression of $(1, z_{t-2})^T$ on Δz_{t-1}. Then proceed as before, with β replaced by $\beta_* = (\delta,\beta^T)^T$. Note that in this case the size of the matrix $\hat{S}_{2,2}$ is now $(q+1) \times (q+1)$, and the sizes of the matrices $\hat{S}_{1,2}$ and $\hat{S}_{2,1}$ are now $q \times (q+1)$ and $(q+1) \times q$, respectively. The limiting null distributions of the lambda-max and trace tests however are different from the ones before. Thus, there are three cases with different null distributions:

(i) The cointegrating restrictions (23) on the intercept parameters do not hold and are not imposed;

(ii) The cointegrating restrictions (23) on the intercept parameters hold but are not imposed;

(iii) The cointegrating restrictions (23) on the intercept parameters hold and are imposed.

Recall that case (i) corresponds to Figure 3, and the cases (ii) and (iii) correspond to Figure 2.

The problem with testing parametric restrictions on the cointegrating vectors in case (iii) is that we cannot confine our attention to restrictions of the form $\beta = H\varphi$ only, but that we have to include δ as well. Thus, we can only test restrictions of the form $\beta_* = (\delta,\beta^T)^T = H\varphi$, where H is now a given $(q+1)\times s$ matrix with rank $s \le r$, and φ is a conformable matrix of free parameters. However, the parameter vector δ is in general of no (economic) interest, so that one has to re-estimate the model without imposing the restriction (23) in order to test restrictions on β only.

4.5 Further extensions

Along the same lines as above one may include seasonal dummy variables in the VECM, provided they are taken in deviation from their sample means so that they become orthogonal to the intercept, without affecting the null distributions of the lambda-max and trace tests. Moreover, recently Johansen (1994) considered also the case where a time trend is included in the VECM, i.e.,

$$\Delta z_t = \pi_{0,0} + \pi_{0,1} t + \sum_{j=1}^{p-1} \Pi_j \Delta z_{t-j} + \alpha\beta^T z_{t-p} + e_t. \tag{42}$$

In this case cointegrating restrictions on the trend parameters take the form

$$\pi_{0,1} = \alpha\gamma, \tag{43}$$

so that then

$$\Delta z_t = \pi_{0,0} + \sum_{j=1}^{p-1}\Pi_j\Delta z_{t-j} + \alpha[\gamma.t + \beta^T z_{t-p}] + e_t. \tag{44}$$

Similar to Proposition 1 it can be shown that under the conditions (43), $\gamma.t + \beta^T z_{t-p}$ is zero-mean stationary, hence $\beta^T z_t$ is trend stationary, and z_t is a multivariate unit root **with drift** process. In this case the time series pattern is as in Figure 4:

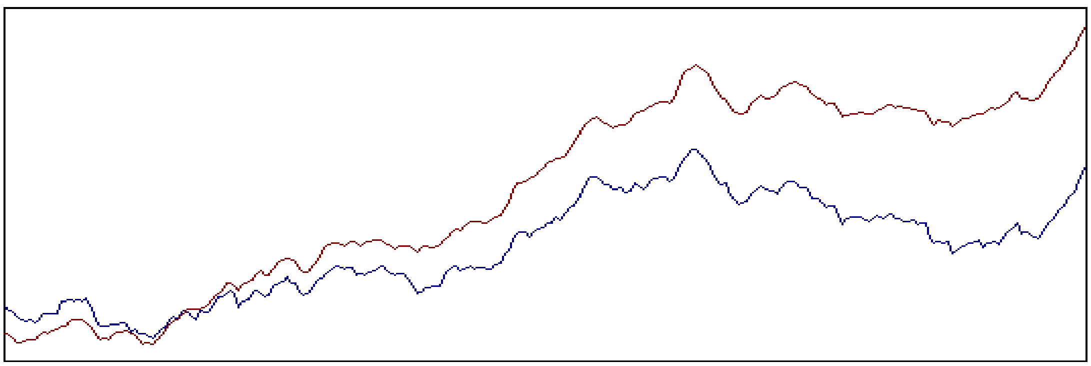

Figure 4 VECM with cointegrating restrictions on the trend parameters imposed

Without the restriction (43) z_t has linear drift and thus a quadratic time trend, which is unlikely in practice. Therefore, if the time series have drift and veer apart, VECM (44) is the appropriate model. Note that the time series in Figure 1 veer apart, which corresponds to VECM (44).

Again, the null distributions of the lambda-max and trace tests differ from the cases with an intercept only, and between the cases where cointegrating restrictions on the trend parameters are imposed or not.

5. Nonparametric cointegration analysis

5.1 Nonparametric tests of the number of cointegrating vectors

The basic ideas behind my nonparametric cointegration approach is that the difference in asymptotic behavior of certain weighted means of z_t and Δz_t under cointegration can be exploited to construct cointegration tests. In particular, these weighted means can be used to construct two random matrices such that cointegration tests can be based on their generalized eigenvalues, similarly to Johansen's approach. I will only outline the main ideas; for the details and the proofs I refer to Bierens (1997a) and the separate appendix to that paper.

Denote the partial sums associated to z_t and Δz_t by $S_n^z(x) = 0$ if $x \in [0,n^{-1}]$, $S_n^z(x) = \sum_{t=1}^{[xn]} z_t$ if $x \in (n^{-1},1]$, and $S_n^{\Delta z}(x) = 0$ if $x \in [0,n^{-1}]$, $S_n^{\Delta z}(x) = \sum_{t=1}^{[xn]} \Delta z_t$ if $x \in (n^{-1},1]$, respectively. Then it is not hard to prove that under Assumption 1,

$$\begin{pmatrix} S_n^z(x)/(n\sqrt{n}) \\ S_n^{\Delta z}(x)/\sqrt{n} \end{pmatrix} \Rightarrow \begin{pmatrix} C(1)\int_0^x W(y)dy \\ C(1)W(x) \end{pmatrix}, \tag{45}$$

where $W(.)$ is a q-variate standard Wiener process, and "$\Rightarrow$" means weak convergence. Cf. Billingsley (1968). The latter symbol will also be used to indicate convergence in distribution and convergence in probability, as these concepts are special cases of weak convergence.

Next, consider the following class of weighted means of z_t and Δz_t:

$$M_n^z(F) = \frac{1}{n}\sum_{t=1}^{n} F(t/n)z_t, \quad M_n^{\Delta z}(F) = \frac{1}{n}\sum_{t=1}^{n} F(t/n)\Delta z_t, \tag{46}$$

where F is a continuously differentiable function on the unit interval [0,1] with derivative f. Then it is pretty straightforward to verify from (45) and Lemma 9.6.3 in Bierens (1994, p.200) that

$$\begin{aligned} \begin{pmatrix} M_n^z(F)/\sqrt{n} \\ M_n^{\Delta z}(F)\sqrt{n} \end{pmatrix} &= F(1)\begin{pmatrix} S_n^z(1)/(n\sqrt{n}) \\ S_n^{\Delta z}(1)/\sqrt{n} \end{pmatrix} - \int f(x)\begin{pmatrix} S_n^z(x)/(n\sqrt{n}) \\ S_n^{\Delta z}(x)/\sqrt{n} \end{pmatrix} dx \Rightarrow \\ &\begin{pmatrix} C(1)\int F(x)W(x)dx \\ C(1)(F(1)W(1) - \int f(x)W(x)dx) \end{pmatrix} \sim N_{2q}(0,\ (C(1)C(1)^T)\otimes\Sigma_F), \end{aligned} \tag{47}$$

where

$$\Sigma_F = \begin{pmatrix} \int\int F(x)F(y)\min(x,y)dx\,dy & \frac{1}{2}\left(\int F(x)dx\right)^2 \\ \frac{1}{2}\left(\int F(x)dx\right)^2 & \int F(x)^2 dx \end{pmatrix}. \tag{48}$$

(The integrals in (47), Σ_F and below are taken over the unit interval, unless otherwise indicated). Note that if we choose F such that

$$\int F(x)dx = 0 \tag{49}$$

then Σ_F becomes a diagonal matrix, so that then the two components on the right-hand side of (47) are independent normally distributed:

Lemma 1: *Under Assumption 1 and condition* (49),

$$\begin{pmatrix} M_n^z(F)/\sqrt{n} \\ M_n^{\Delta z}(F)\sqrt{n} \end{pmatrix} \Rightarrow \begin{pmatrix} C(1)X_F\sqrt{\int\int F(x)F(y)\min(x,y)\,dx\,dy} \\ C(1)Y_F\sqrt{\int F(x)^2dx} \end{pmatrix}, \tag{50}$$

where X_F and Y_F are independent q-variate standard normally distributed random vectors depending on F in the following way:

$$X_F = \frac{\int F(x)W(x)dx}{\sqrt{\int\int F(x)F(y)\min(x,y)dxdy}}, \quad Y_F = \frac{F(1)W(1) - \int f(x)W(x)dx}{\sqrt{\int F(x)^2dx}}. \tag{51}$$

Note that in the case of cointegration the matrix $C(1)C(1)^T$ is singular, so that the limiting normal distribution at the right-hand side of (50) is singular, hence for any cointegrating vector ξ we have $\xi^T M_n^z(F)/\sqrt{n} \Rightarrow 0$ and $\xi^T M_n^{\Delta z}(F)\sqrt{n} \Rightarrow 0$. This suggests that for cointegrating vectors ξ the rates of convergence of $\xi^T M_n^z(F)$ and $\xi^T M_n^{\Delta z}(F)$ will be different from the case in Lemma 1:

Lemma 2: *Let Assumption 1 and condition* (49) *hold. If z_t is cointegrated then for each matrix* $\Xi = (\xi_1, ..., \xi_r)$ *of cointegrating vectors ξ_i,*

$$\begin{pmatrix} \Xi^T M_n^z(F)\sqrt{n} \\ \Xi^T M_n^{\Delta z}(F)\,n \end{pmatrix} \Rightarrow \begin{pmatrix} \Xi^T D(1)Y_F\sqrt{\int F(x)^2dx} \\ F(1)\Xi^T D_* Z \end{pmatrix}, \tag{52}$$

where Y_F and Z are independent q-variate standard normally distributed, with Y_F defined by (51) *and* $D_* = [\Sigma_{j=0}^n D_j D_j^T]^{1/2}$. [*c.f.* (5)].

Comparing Lemmas 1 and 2 we see that the asymptotic behavior, in particular the absolute and relative rates of convergence, of the statistics (46) differ substantially according to whether z_t is cointegrated or not. These differences can now be exploited in constructing nonparametric cointegration tests, as follows.

Choose a sequence F_k, $k = 1,2,...,m$, with $m \geq q$, of continuously differentiable real functions on [0,1] with derivatives f_k satisfying condition (49), i.e., $\int F_k(x)dx = 0$ for $k = 1,...,m$, so that the random vectors

$$X_k = \frac{\int F_k(x)W(x)dx}{\sqrt{\int\int F_k(x)F_k(y)\min(x,y)dxdy}}, \quad Y_k = \frac{F_k(1)W(1) - \int f_k(x)W(x)dx}{\sqrt{\int F_k(x)^2dx}} \tag{53}$$

[cf. (51)] are mutually independent, together with conditions ensuring that these random vectors are also independent for $k = 1,2,....$ Such functions F_k do exist. For example, let

$$F_k(x) = \cos(2k\pi x), \quad k = 1,2,3,.... \tag{54}$$

Actually, this choice of F_k is "optimal" in the sense that it maximizes a lower bound of the power function of the nonparametric cointegration test.

Next, construct the random matrices $\hat{A}_m = \sum_{k=1}^m a_{n,k}a_{n,k}^T$ and $\hat{B}_m = \sum_{k=1}^m b_{n,k}b_{n,k}^T$, where

$$a_{n,k} = \frac{M_n^z(F_k(\cdot))}{\sqrt{\int\int F_k(x)F_k(y)\min(x,y)dxdy}}, \quad b_{n,k} = \frac{\sqrt{n}\,M_n^{\Delta z}(F_k(\cdot))}{\sqrt{\int F_k(x)^2dx}}. \tag{55}$$

Moreover, denote

$$\gamma_k = \frac{\sqrt{\int F_k(x)^2dx}}{\sqrt{\int\int F_k(x)F_k(y)\min(x,y)dxdy}}, \quad \delta_k = \frac{F_k(1)}{\sqrt{\int F_k(x)^2dx}}. \tag{56}$$

Then it follows from Lemmas 1-2:

Lemma 3: *Let rank $C(1) = q - r$, let R_{q-r} be the matrix of orthonormal eigenvectors of $C(1)C(1)^T$ corresponding to the $q-r$ positive eigenvalues, let R_r be the matrix of orthonormal eigenvectors corresponding to the r zero eigenvalues, and denote $R = (R_{q-r},R_r)$. Then under Assumption 1*:

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

$$\Rightarrow \begin{pmatrix} R_{q-r}^T C(1)\sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r} & R_{q-r}^T C(1)\sum_{k=1}^{m} \gamma_k X_k Y_k^T D(1)^T R_r \\ R_r^T D(1)\sum_{k=1}^{m} \gamma_k Y_k X_k^T C(1)^T R_{q-r} & R_r^T D(1)\sum_{k=1}^{m} \gamma_k^2 Y_k Y_k^T D(1)^T R_r \end{pmatrix} \quad (57)$$

and

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

$$\Rightarrow \begin{pmatrix} R_{q-r}^T C(1)\sum_{k=1}^{m} Y_k Y_k^T C(1)^T R_{q-r} & R_{q-r}^T C(1)\sum_{k=1}^{m} \delta_k Y_k Z^T D_*^T R_r \\ R_r^T D_* \sum_{k=1}^{m} \delta_k Z Y_k^T C(1)^T R_{q-r} & R_r^T D_* \sum_{k=1}^{m} \delta_k^2 Z Z^T D_*^T R_r \end{pmatrix}, \quad (58)$$

where the X_i's, the Y_j's and Z are independent q-variate standard normally distributed.

Now at first sight one might think of employing these results for constructing cointegration tests by using the solutions of the generalized eigenvalue problem $\det(\hat{A}_m - \lambda\hat{B}_m) = 0$, similarly to Johansen's approach. However, the problem is that under cointegration both matrices converge in distribution to singular matrices. In deriving the limiting distribution of the generalized eigenvalues, Johansen used a result of Andersen, Brons and Jensen (1983) saying that the ordered solutions of the generalized eigenvalue problem $\det(P_n - \lambda Q_n) = 0$, where P_n and Q_n are stochastic matrices converging jointly in distribution to P_* and Q_*, say, converge in distribution to the ordered solutions of the generalized eigenvalue problem $\det(P_* - \lambda Q_*) = 0$, provided Q_* is a.s. nonsingular. Due to the latter condition, this result cannot be used to derive the limiting distribution of the ordered solutions of the generalized eigenvalue problem $\det(\hat{A}_m - \lambda\hat{B}_m) = 0$. However, the following trick will cure the problem.

Observe that part (57) of Lemma 3 implies

$$\frac{R^T \hat{A}_m^{-1} R}{n^2} \Rightarrow \begin{pmatrix} O & O \\ O & V_{r,m}^{-1} \end{pmatrix}, \quad (59)$$

where

$$V_{r,m} = R_r^T D(1) \sum_{k=1}^{m} \gamma_k^2 Y_k Y_k^T D(1)^T R_r - \left(R_r^T D(1) \sum_{k=1}^{m} \gamma_k Y_k X_k^T C(1)^T R_{q-r} \right)$$
$$\times \left(R_{q-r}^T C(1) \sum_{k=1}^{m} X_k X_k^T C(1)^T R_{q-r} \right)^{-1} \left(R_{q-r}^T C(1) \sum_{k=1}^{m} \gamma_k X_k Y_k^T D(1)^T R_r \right). \quad (60)$$

Note that by Assumption 2, this matrix is a.s. nonsingular. Hence $R^T(\hat{B}_m + n^{-2}\hat{A}_m^{-1})R$ converges in distribution to a nonsingular block-diagonal matrix. Now using the result of Andersen, Brons and Jensen (1983) it follows straightforwardly:

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

$$\det\left(\hat{A}_m - \lambda(\hat{B}_m + n^{-2}\hat{A}_m^{-1})\right) = 0, \quad (61)$$

and let $\lambda_{1,m} \geq ... \geq \lambda_{q-r,m}$ *be the ordered solution of the generalized eigenvalue problem* $\det(\sum_{k=1}^{m} X_k^* X_k^{*T} - \lambda \sum_{k=1}^{m} Y_k^* Y_k^{*T}) = 0$, *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 1-2,* $(\hat{\lambda}_{1,m},..,\hat{\lambda}_{q,m}) \Rightarrow (\lambda_{1,m},...,\lambda_{q-r,m},0,..,0)$.

This result suggests to use $\hat{\lambda}_{q-r,m}$ as a test statistic for testing the null hypothesis that there are r cointegrating vectors against the alternative that there are r+1 cointegrating vectors. The test involved is a left-sided test: the null is rejected if $\hat{\lambda}_{q-r,m}$ is smaller than a critical value. See Bierens (1997, Table 2) for the critical values involved.

The power of the test involved depends on the choice of m as well as on the choice of the functions F_k. As mentioned before, the choice (54) is "optimal" in that it maximizes a lower bound of the power function of the test. However, the asymptotically equivalent functions $F_k(x) = \cos(2k\pi(x - 0.5/n))$ will do an even better job because then the test becomes invariant for drift in the multivariate unit root process z_t, i.e. the case where $z_t = z_{t-1} + c + u_t$, where c is a vector of drift parameters and u_t is a zero mean stationary process satisfying Assumption 1. The matrices $\hat{A}_m$ and $\hat{B}_m$ then become

$$\hat{A}_m = \frac{8\pi^2}{n} \sum_{k=1}^{m} k^2 \left(\frac{1}{n} \sum_{t=1}^{n} \cos(2k\pi(t-0.5)/n) z_t \right) \left(\frac{1}{n} \sum_{t=1}^{n} \cos(2k\pi(t-0.5)/n) z_t \right)^T \quad (62)$$

$$\hat{B}_m = 2n\sum_{k=1}^{m}\left(\frac{1}{n}\sum_{t=1}^{n}\cos(2k\pi(t-0.5)/n)\Delta z_t\right)\left(\frac{1}{n}\sum_{t=1}^{n}\cos(2k\pi(t-0.5)/n)\Delta z_t\right)^T . \qquad (63)$$

The same lower bound of the power function of the test mentioned before depends on *m*, hence maximizing this lower bound w.r.t. *m* would yield a sensible choice for *m*. The resulting values for *m* for significance levels $s\times5\%$, $s = 1,2$, and $0 \leq r \leq 4$, $1 \leq q \leq 5$, can be expressed by the formula:

$$m = q + I(q \geq s + 1)I(r = 0), \qquad (64)$$

where $I(.)$ is the indicator function.

5.2 Testing linear restrictions on the cointegrating vectors

Once the number *r* of cointegrating vectors is established, and $0 < r < q$, one may wish to verify whether there exist cointegrating vectors β satisfying the linear restriction $H_0:\ \beta = H\varphi,\quad \varphi \in \mathbb{R}^s$, where *H* is a given $q \times s$ matrix with full column rank $s \leq r$ and φ is arbitrary. Thus, the null hypothesis is that the space spanned by the columns of the matrix *H* is contained in the space of cointegrating vectors. For example, in the case $q = 3$ we may wish to test whether there exists a cointegrating vector $\beta = (\beta_1, \beta_2, \beta_3)^T$ such that $\beta_1 + \beta_2 = 0$ and $\beta_3 = 0$, so that then $H = (1,-1,0)^T$.

At first sight one might think of mimicking Johansen's tests for these 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 null distribution. Therefore I propose two test, the trace test and the lambda-max test, on the basis of the matrix $\hat{A}_m$ only. The recipe for the lambda-max test is as follows. Choose $m = 2q$. The lambda-max test is based on the maximum solution, say $\lambda_{\max}(H)$, of the generalized eigenvalue problem

$$\det\left[H^T\hat{A}_m H - \lambda H^T(\hat{A}_m + n^{-2}\hat{A}_m^{-1})^{-1}H\right] = 0. \qquad (65)$$

The test statistic involved is $n^2\lambda_{\max}(H)$, and we reject the null hypothesis if $n^2\lambda_{\max}(H)$ is larger than a critical value. See Bierens (1997, Table 4). The trace test statistic is n^2 times the sum of the solutions of (65), and the critical values involved are given in Bierens (1997, Table 3).

The choice of $m = 2q$ is somewhat heuristic: a lower bound of the power function is monotonic increasing in m, but too large an m may mess up the size of the test. Since this lower bound of the power function is almost flat for $m > 2q$, I recommend the "rule of thumb" $m = 2q$.

5.3 Consistent estimation of a basis of the space of cointegrating vectors

Given that there are r linear independent (but unknown) cointegrating vectors $\xi_1,...,\xi_r$, one can consistently estimate a basis of the space of cointegrated vectors as follows. Choose again $m = 2q$, and let H be the matrix of eigenvectors corresponding to the r smallest eigenvalues of the generalized eigenvalue problem

$$\det\left[\hat{A}_m - \lambda(\hat{A}_m + n^{-2}\hat{A}_m^{-1})^{-1}\right] = 0, \qquad (66)$$

where H is standardized such that

$$\hat{H}^T\left(\hat{A}_m + n^{-2}\hat{A}_m^{-1}\right)^{-1}\hat{H} = I_r. \qquad (67)$$

Then $\hat{H} = (\xi_1,\dots,\xi_r)\hat{\Gamma}_r + O_p(1/n)$, where $\hat{\Gamma}_r$ is $r \times r$ with $\text{rank}(\hat{\Gamma}_r) = r$. Since the cointegrating vectors $\xi_1,...,\xi_r$ can be chosen orthonormal, we can interpret this result also in terms of projections: The distances between the columns of H and their corresponding projections on the space of cointegrating vectors vanish at order $O_p(1/n)$.

5.4 Seasonal drift

The above results apply to multivariate unit root processes with constant drift, but not to processes with seasonal drift. In the latter case one should replace z_t in the matrices A_m and B_m by seasonal moving averages $\bar{z}_t = (1/s)\Sigma_{\tau=0}^{s-1}z_{t-\tau}$, where s is the number of seasons. With this modification, the nonparametric approach is applicable to time series with seasonal drift.

5.5 Concluding remarks

My nonparametric cointegration approach has some clear advances over Johansen's maximum likelihood approach, in particular that it does not require to specify a lag length p and the deterministic variables d_t of the VECM (1), and that the critical values are case independent. This will become more clear in the empirical example in section 6. However, there is also a disadvantage, namely that the nonparametric tests are not invariant for scale: Replacing z_t by $z_t^* = Qz_t$, where Q is a nonsingular matrix, the generalized eigenvalue problem (61) becomes

$\det[\hat{A}_m - \lambda(\hat{B}_m + n^{-2}(Q^TQ)^{-1}\hat{A}_m^{-1}(Q^TQ)^{-1})] = 0$, and similarly the matrix $\hat{A}_m^{-1}$ in (65), (66) and (67) changes accordingly to $(Q^TQ)^{-1}\hat{A}_m^{-1}(Q^TQ)^{-1}$. Of course, asymptotically this does not matter, but in small samples it clearly will. On the other hand, due to the fact that for k = 1,2,3....,

$$\sum_{t=1}^{n}\cos(2\pi k(t - 0.5)/n) = 0, \quad \sum_{t=1}^{n}t\cos(2\pi k(t - 0.5)/n) = 0, \tag{68}$$

the tests are invariant for location shifts in Δz_t. Therefore, if all the variables in z_t are in logs the units of measurement of the original variables do not matter, due to (68), but one should be cautious in conducting the nonparametric tests to vector time series processes z_t if not all components are in logs.

6. An empirical example

I will now apply my nonparametric and Johansen's likelihood ratio cointegration tests to the annual data on the logs of consumption and income in Sweden from 1861 to 1988. However, before conducting cointegration analysis, one should test first whether the time series involved are unit root processes or not. From Figure 1 it is obvious that the appropriate hypotheses to be tested are the unit root *with drift* hypothesis against *trend* stationarity. Therefore, I have conducted the Augmented Dickey-Fuller (ADF) t-test of the null hypothesis $\alpha = 0$ in the auxiliary regression

$$\Delta y_t = \alpha y_{t-1} + \sum_{j=1}^{p}\beta_j\Delta y_{t-j} + \gamma_0 + \gamma_1 t + \varepsilon_t$$

with p depending on the sample size n [see Said and Dickey (1984)], and the Phillips-Perron (1988) test Z_α of the unit root with drift hypothesis against the trend stationarity hypothesis. The truncation lag p of the Newey-West (1987) estimator of the long run variance of Δy_t employed by the Phillips-Perron test, as well as the ADF lag length p have been chosen: $p = [5n^{1/4}] = 16$ for $n = 128$. The result is that for both time series the unit root with drift hypothesis can not be rejected at the 10% significance level. Also, I have conducted the Bierens-Guo (1993) tests of the trend stationarity hypothesis against the unit root with drift hypothesis, and for both time series the null hypothesis of linear trend stationarity is rejected at the 5% significance level.

The results of the nonparametric cointegration tests, conducted at the 10% significance level, are:

H_0	H_1	*Test statistic*	10% *critical region*	*Conclusion*
$r = 0$	$r = 1$	0.00005	(0, 0.005)	*Reject* H_0
$r = 1$	$r = 2$	24.33266	(0, 0.111)	*Accept* H_0

Thus the conclusion is that there is one cointegrating vector: $r = 1$. The estimate $\hat{\beta}$ of the parameter β in the cointegrating vector $(1,-\beta)^T$ is: $\hat{\beta} = 0.9444$, and the null hypothesis $\beta = 1$ is not rejected at the 10% significance level. The latter hypothesis corresponds to the hypothesis that the long run marginal propensity to consume from income equals 1.

Next, I have conducted Johansen's tests on the basis of VECM (1) for $p = 1,..,6$, for the following five cases w.r.t the deterministic part $\Pi_0 d_t$:

1. $\Pi_0 d_t = \pi_0$, where π_0 is not proportional to α.
2. $\Pi_0 d_t = \pi_0$, where π_0 is proportional to α but this restriction is not imposed.
3. $\Pi_0 d_t = \pi_0$, where π_0 is proportional to α and this restriction is imposed.
4. $\Pi_0 d_t = \pi_0 + \pi_1 t$, where π_1 is proportional to α but this restriction is not imposed.
5. $\Pi_0 d_t = \pi_0 + \pi_1 t$, where π_1 is proportional to α and this restriction is imposed.

In view of Figure 1, the options 1, 2 and 3 are not suitable because they imply that the time series run parallel, whereas the two time series involved veer apart. Therefore, only the options 4 or 5 are applicable. Nevertheless, to demonstrate the sensitivity of the Johansen approach for the specification of the deterministic part of the VECM, I will try all five options. In the cases 3 and 5 with test result $r = 1$ I have also tested whether the imposed cointegrating restriction holds. Moreover, for the cases with test result $r = 1$ I have tested the hypothesis that the cointegrating vector $(1,-\beta)^T$ is equal to $(1,-1)^T$, so that $\beta = 1$. All tests are conducted at the 10% significance level. The results are presented in Tables 1-3.

Table 1: Johansen's test results for the number r of cointegrating vectors

Case 1		Case 2		Case 3		Case 4		Case 5	
p	r	p	r	p	r	p	r	p	r
1	2	1	1	1	2	1	1	1	1
2	2	2	1	2	2	2	1	2	1
3	2	3	1	3	2	3	1	3	1
4	0 *or* 2	4	0 *or* 1	4	2	4	0	4	0
5	0	5	0	5	1	5	0	5	0
6	0 *or* 2	6	0	6	1	6	0	6	0

Table 2: $\hat{\beta}$ *and test of* H_0: $\beta=1$ *for* $r = 1$

Case 2			Case 3			Case 4			Case 5		
p	$\hat{\beta}$	H_0	p	$\hat{\beta}$	H_0	p	$\hat{\beta}$	H_0	p	$\hat{\beta}$	H_0
1	0.9448	*reject*	5	0.9420	*accept*	1	0.9245	*reject*	1	0.9245	*reject*
2	0.9367	*reject*	6	0.9442	*accept*	2	0.9115	*reject*	2	0.9115	*reject*
3	0.9397	*reject*				3	0.9165	*reject*	3	0.9165	*reject*

Table 3: Test of H_0:$\pi_i = \alpha.\gamma$ *for* $r = 1$

Case 3		Case 5	
p	H_0	p	H_0
5	*reject*	1	*reject*
6	*reject*	2	*accept*
		3	*reject*

The results in Table 1 where r takes two possible values are due to the fact that the lambda-max and trace tests gave different test results. The test result $r = 2$ would imply that both series are stationary, but the unit root and trend stationarity tests conducted on the single series indicate that they are unit root processes. For case 3 with $p = 5$ and 6 the imposed cointegrating restriction on π_0 is rejected, so that the result $r = 1$ in the cases 2 and 3 should be ignored. In case 5 with $p = 1$ and 3 the cointegrating restriction on the trend parameter vector π_1 is rejected, which would imply the presence of a linear time trend in the drift. Since this is implausible, because then the growth rates of consumption and income have a linear trend and therefore grow to infinity themselves, there is only one case with $r = 1$ left that make sense, namely case 5 with $p = 2$. For this case the estimated cointegrating vector is $(1,-\hat{\beta})^T$, where $\hat{\beta} = 0.9115$, and the hypothesis $\beta = 1$ is rejected. The latter result is probably more accurate than the corresponding result of the nonparametric test, because the nonparametric test of restrictions on the cointegrating vector seems less powerful than the corresponding Johansen test. See Bierens (1997a).

References

Anderson, S.A., H.K. Brons and S.T. Jensen (1983): "Distribution of Eigenvalues in Multivariate Statistical Analysis" *Annals of Statistics* **11**, pp. 392-415.

Bierens, H.J. (1994): *Topics in Advanced Econometrics*: *Estimation, Testing, and Specification of Cross-Section and Time Series Models*. Cambridge: Cambridge University Press.

Bierens, H.J. (1997a): "Nonparametric Cointegration Analysis", *Journal of Econometrics* 77, 379-404.

Bierens, H.J. (1997b): "Cointegration Analysis", in C. Heij, J.M. Schumacher, B. Hanzon and C. Praagman (Eds.), *System Dynamics in Economic and Financial Models*, John Wiley, 217-246.

Bierens, H.J. and S.Guo (1993): "Testing Stationarity and Trend Stationarity Against the Unit Root Hypothesis", *Econometric Reviews* **12**, pp. 1-32.

Billingsley, P. (1968): *Convergence of Probability Measures.* New York: John Wiley.

Boswijk, H.P. (1994): "Testing for an Unstable Root in Conditional and Structural Error Correction Models", *Journal of Econometrics* **63**, pp. 37-60.

Boswijk, H.P. (1995): "Efficient Inference on Cointegrating Parameters in Structural Error Correction Models", *Journal of Econometrics* **69**, pp. 133-158.

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Auto-regressive Times Series with a Unit Root", *Journal of the American Statistical Association* **74**, pp. 427-431.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Auto-regressive Time Series with a Unit Root", *Econometrica* **49**, pp. 1057-1072.

Engle, R.F. and C.W.J.Granger (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica* **55**, pp. 251-276.

Engle, R.F. and S.B. Yoo (1987), "Forecasting and Testing in Cointegrated Systems", *Journal of Econometrics* **35**, pp. 143-159

Engle, R.F. and S.B. Yoo (1991), "Cointegrated Economic Time Series: An Overview with New Results", in R.F. Engle and C.W.J. Granger (eds): *Long-Run Economic Relationships*, Oxford: Oxford University Press, pp. 237-266.

Friedman, M. (1957), *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*. New York: John Wiley.

Granger, C.W.J. (1981), "Some Properties of Time Series and Their Use in Econometric Model Specification", *Journal of Econometrics* **16**, 121-130.

Hamming, R.W. (1973), *Numerical Methods for Scientists and Engineers.* New York: Dover Publications.

Hall, P. and C.C. Heyde (1980), *Martingale Limit Theory and Its Applications*. San Diego: Academic Press.

Hansen, B.E. (1992), "Tests for Parameter Instability in Regressions with I(1) Processes", *Journal of Business and Economic Statistics* **10**, pp. 321-335.

Johansen, S. (1988), "Statistical Analysis of Cointegrated Vectors", *Journal of Economic Dynamics and Control* **12**, pp. 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrated Vectors in Gaussian Vector Autoregressive Models", *Econometrica* **59**, pp. 1551-1580.

Johansen, S. (1994), "The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables", *Econometric Reviews* **13**, pp. 205-229.

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**, pp. 169-210

Krantz, O. and L. Nilson (1975), *Swedish National Product 1861-1970: New Aspects on Methods and Measurements*, Lund: G.W.K. Gleerup/Liber Läromedel.

Melander, E., A. Vredin and A. Warne (1992), "Stochastic Trends and Economic Fluctuations in a Small Open Economy", *Journal of Applied Econometrics* **7**, pp. 369-394.

Newey, W.K. and K.D. West (1987), "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica* **55**, pp. 703-708.

Park, J.Y (1990), "Testing for Unit Root and Cointegration by Variable Addition", *Advances in Econometrics* **8**, pp. 107-133.

Perron, P. (1988): "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach", *Journal of Economic Dynamics and Control* **12**, pp. 297-332.

Perron, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica* **57**, pp. 1361-1402.

Perron, P. (1990): "Testing the Unit Root in a Time Series with a Changing Mean", *Journal of Business and Economic Statistics* **8**, pp. 153-162.

Phillips, P.C.B. (1987), "Time Series Regression With Unit Roots", *Econometrica* **55**, pp. 277-302.

Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems", *Econometrica* **59**, pp. 283-306.

Phillips, P.C.B. and B.Hansen (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes", *Review of Economic Studies* **57**, pp. 99-125

Phillips, P.C.B. and S.Ouliaris (1990): "Asymptotic Properties of Residual Based Tests for Cointegration", *Econometrica* **58**, pp. 165-193.

Phillips, P.C.B. and P.Perron (1988), "Testing for a Unit Roots in Time Series Regression", *Biometrika* **75**, pp. 335-346.

Phillips, P.C.B. and V.Solo (1992), "Asymptotics for Linear Processes", *Annals of Statistics* **20**, pp. 971-1001.

Said, S.E and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average of Unknown Order", *Biometrika* **71**, pp. 599-607.

Stock, J.H. and M.W. Watson (1988), "Testing for Common Trends", *Journal of the American Statistical Association* **83**, pp. 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**, pp. 113-144.