

## STATISTICAL ANALYSIS OF COINTEGRATION VECTORS

Søren JOHANSEN\*

*University of Copenhagen, DK-2100 Copenhagen, Denmark*

Received September 1987, final version received January 1988

We consider a nonstationary vector autoregressive process which is integrated of order 1, and generated by i.i.d. Gaussian errors. We then derive the maximum likelihood estimator of the space of cointegration vectors and the likelihood ratio test of the hypothesis that it has a given number of dimensions. Further we test linear hypotheses about the cointegration vectors.

The asymptotic distribution of these test statistics are found and the first is described by a natural multivariate version of the usual test for unit root in an autoregressive process, and the other is a  $\chi^2$  test.

### 1. Introduction

The idea of using cointegration vectors in the study of nonstationary time series comes from the work of Granger (1981), Granger and Weiss (1983), Granger and Engle (1985), and Engle and Granger (1987). The connection with error correcting models has been investigated by a number of authors; see Davidson (1986), Stock (1987), and Johansen (1988) among others.

Granger and Engle (1987) suggest estimating the cointegration relations using regression, and these estimators have been investigated by Stock (1987), Phillips (1985), Phillips and Durlauf (1986), Phillips and Park (1986a, b, 1987), Phillips and Ouliaris (1986, 1987), Stock and Watson (1987), and Sims, Stock and Watson (1986). The purpose of this paper is to derive maximum likelihood estimators of the cointegration vectors for an autoregressive process with independent Gaussian errors, and to derive a likelihood ratio test for the hypothesis that there is a given number of these. A similar approach has been taken by Ahn and Reinsel (1987).

This program will not only give good estimates and test statistics in the Gaussian case, but will also yield estimators and tests, the properties of which can be investigated under various other assumptions about the underlying data generating process. The reason for expecting the estimators to behave better

\*The simulations were carefully performed by Marc Andersen with the support of the Danish Social Science Research Council. The author is very grateful to the referee whose critique of the first version greatly helped improve the presentation.

than the regression estimates is that they take into account the error structure of the underlying process, which the regression estimates do not.

The processes we shall consider are defined from a sequence  $\{\varepsilon_t\}$  of i.i.d.  $p$ -dimensional Gaussian random variables with mean zero and variance matrix  $\Lambda$ . We shall define the process  $X_t$  by

$$X_t = \Pi_1 X_{t-1} + \cdots + \Pi_k X_{t-k} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (1)$$

for given values of  $X_{-k+1}, \dots, X_0$ . We shall work in the conditional distribution given the starting values, since we shall allow the process  $X_t$  to be nonstationary. We define the matrix polynomial

$$A(z) = I - \Pi_1 z - \cdots - \Pi_k z^k,$$

and we shall be concerned with the situation where the determinant  $|A(z)|$  has roots at  $z = 1$ . The general structure of such processes and the relation to error correction models was studied in the above references.

We shall in this paper mainly consider a very simple case where  $X_t$  is integrated of order 1, such that  $\Delta X_t$  is stationary, and where the impact matrix

$$A(z)|_{z=1} = \Pi = I - \Pi_1 - \cdots - \Pi_k$$

has rank  $r < p$ . If we express this as

$$\Pi = \alpha\beta', \quad (2)$$

for suitable  $p \times r$  matrices  $\alpha$  and  $\beta$ , then we shall assume that, although  $\Delta X_t$  is stationary and  $X_t$  is nonstationary as a vector process, the linear combinations given by  $\beta' X_t$  are stationary. In the terminology of Granger this means that the vector process  $X_t$  is cointegrated with cointegration vectors  $\beta$ . The space spanned by  $\beta$  is the space spanned by the rows of the matrix  $\Pi$ , which we shall call the cointegration space.

In this paper we shall derive the likelihood ratio test for the hypothesis given by (2) and derive the maximum likelihood estimator of the cointegration space. Then we shall find the likelihood ratio test of the hypothesis that the cointegration space is restricted to lie in a certain subspace, representing the linear restrictions that one may want to impose on the cointegration vectors.

The results we obtain can briefly be described as follows: the estimation of  $\beta$  is performed by first regressing  $\Delta X_t$  and  $X_{t-k}$  on the lagged differences. From the residuals of these  $2p$  regressions we calculate a  $2p \times 2p$  matrix of product moments. We can now show that the estimate of  $\beta$  is the empirical canonical variates of  $X_{t-k}$  with respect to  $\Delta X_t$  corrected for the lagged differences.

The likelihood ratio test is now a function of certain eigenvalues of the product moment matrix corresponding to the smallest squared canonical correlations. The test of the linear restrictions involve yet another set of eigenvalues of a reduced product moment matrix. The asymptotic distribution of the first test statistic involves an integral of a multivariate Brownian motion with respect to itself, and turns out to depend on just one parameter, namely the dimension of the process, and can hence be tabulated by simulation or approximated by a  $\chi^2$  distribution. The second test statistic is asymptotically distributed as  $\chi^2$  with the proper degrees of freedom. It is also shown that the maximum likelihood estimator of  $\beta$  suitably normalised is asymptotically distributed as a mixture of Gaussian variables.

## 2. Maximum likelihood estimation of cointegration vectors and likelihood ratio tests of hypotheses about cointegration vectors

We want to estimate the space spanned by  $\beta$  from observations  $X_t$ ,  $t = -k + 1, \dots, T$ . For any  $r \leq p$  we formulate the model as the hypothesis

$$H_0: \text{rank}(\Pi) \leq r \quad \text{or} \quad \Pi = \alpha\beta', \quad (3)$$

where  $\alpha$  and  $\beta$  are  $p \times r$  matrices.

Note that there are no other constraints on  $\Pi_1, \dots, \Pi_k$  than (3). Hence a wide class containing stationary as well as nonstationary processes is considered.

The parameters  $\alpha$  and  $\beta$  can not be estimated since they form an over-parametrisation of the model, but one can estimate the space spanned by  $\beta$ .

We can now formulate the main result about the estimation of  $\text{sp}(\beta)$  and the test of the hypothesis (3).

*Theorem 1.* *The maximum likelihood estimator of the space spanned by  $\beta$  is the space spanned by the  $r$  canonical variates corresponding to the  $r$  largest squared canonical correlations between the residuals of  $X_{t-k}$  and  $\Delta X_t$  corrected for the effect of the lagged differences of the  $X$  process.*

*The likelihood ratio test statistic for the hypothesis that there are at most  $r$  cointegration vectors is*

$$-2\ln(Q) = -T \sum_{i=r+1}^p \ln(1 - \hat{\lambda}_i), \quad (4)$$

where  $\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_p$  are the  $p - r$  smallest squared canonical correlations.

*Proof.* Before studying the likelihood function it is convenient to reparametrise the model (1) such that the parameter of interest  $\Pi$  enters explicitly. We

write

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \cdots + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t, \quad (5)$$

where

$$\Gamma_i = -I + \Pi_1 + \cdots + \Pi_i, \quad i = 1, \dots, k.$$

Then  $\Pi = -\Gamma_k$ , and whereas (3) gives a nonlinear constraint on the coefficients  $\Pi_1, \dots, \Pi_k$ , the parameters  $(\Gamma_1, \dots, \Gamma_{k-1}, \alpha, \beta, \Lambda)$  have no constraints imposed. In this way the impact matrix  $\Pi$  is found as the coefficient of the lagged levels in a nonlinear least squares regression of  $\Delta X_t$  on lagged differences and lagged levels. The maximisation over the parameters  $\Gamma_1, \dots, \Gamma_{k-1}$  is easy since it just leads to an ordinary least squares regression of  $\Delta X_t + \alpha\beta'X_{t-k}$  on the lagged differences. Let us do this by first regressing  $\Delta X_t$  on the lagged differences giving the residuals  $R_{0t}$  and then regressing  $X_{t-k}$  on the lagged differences giving the residuals  $R_{kt}$ . After having performed these regressions the concentrated likelihood function becomes proportional to

$$\begin{aligned} L(\alpha, \beta, \Lambda) \\ = |\Lambda|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (R_{0t} + \alpha\beta'R_{kt})' \Lambda^{-1} (R_{0t} + \alpha\beta'R_{kt}) \right\}. \end{aligned}$$

For fixed  $\beta$  we can maximise over  $\alpha$  and  $\Lambda$  by a usual regression of  $R_{0t}$  on  $-\beta'R_{kt}$  which gives the well-known result

$$\hat{\alpha}(\beta) = -S_{0k}\beta(\beta'S_{kk}\beta)^{-1}, \quad (6)$$

and

$$\hat{\Lambda}(\beta) = S_{00} - S_{0k}\beta(\beta'S_{kk}\beta)^{-1}\beta'S_{k0}, \quad (7)$$

where we have defined product moment matrices of the residuals as

$$S_{ij} = T^{-1} \sum_{t=1}^T R_{it}R'_{jt}, \quad i, j = 0, k. \quad (8)$$

The likelihood profile now becomes proportional to

$$|\hat{\Lambda}(\beta)|^{-T/2},$$

and it remains to solve the minimisation problem

$$\min |S_{00} - S_{0k}\beta(\beta'S_{kk}\beta)^{-1}\beta'S_{k0}|,$$

where the minimisation is over all  $p \times r$  matrices  $\beta$ . The well-known matrix

relation [see Rao (1973)]

$$\begin{aligned} \begin{vmatrix} S_{00} & S_{0k}\beta \\ \beta'S_{k0} & \beta'S_{kk}\beta \end{vmatrix} &= |S_{00}| |\beta'S_{kk}\beta - \beta'S_{k0}S_{00}^{-1}S_{0k}\beta| \\ &= |\beta'S_{kk}\beta| |S_{00} - S_{0k}\beta(\beta'S_{kk}\beta)^{-1}\beta'S_{k0}| \end{aligned}$$

shows that we shall minimise

$$|\beta'S_{kk}\beta - \beta'S_{k0}S_{00}^{-1}S_{0k}\beta| / |\beta'S_{kk}\beta|$$

with respect to the matrix  $\beta$ .

We now let  $D$  denote the diagonal matrix of ordered eigenvalues  $\lambda_1 > \dots > \lambda_p$  of  $S_{k0}S_{00}^{-1}S_{0k}$  with respect to  $S_{kk}$ , i.e., the solutions to the equation

$$|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}| = 0, \tag{9}$$

and  $E$  the matrix of the corresponding eigenvectors, then

$$S_{kk}ED = S_{k0}S_{00}^{-1}S_{0k}E,$$

where  $E$  is normalised such that

$$E'S_{kk}E = I.$$

Now choose  $\beta = E\xi$  where  $\xi$  is  $p \times r$ , then we shall minimise

$$|\xi'\xi - \xi'D\xi| / |\xi'\xi|.$$

This can be accomplished by choosing  $\xi$  to be the first  $r$  unit vectors or by choosing  $\hat{\beta}$  to be the first  $r$  eigenvectors of  $S_{k0}S_{00}^{-1}S_{0k}$  with respect to  $S_{kk}$ , i.e., the first  $r$  columns of  $E$ . These are called the canonical variates and the eigenvalues are the squared canonical correlations of  $R_k$  with respect to  $R_0$ . For the details of these calculations the reader is referred to Anderson (1984, ch. 12). This type of analysis is also called reduced rank regression [see Ahn and Reinsel (1987) and Velu, Reinsel and Wichern (1986)]. Note that all possible choices of the optimal  $\beta$  can be found from  $\hat{\beta}$  by  $\beta = \hat{\beta}\rho$  for  $\rho$  an  $r \times r$  matrix of full rank. The eigenvectors are normalised by the condition  $\hat{\beta}'S_{kk}\hat{\beta} = I$  such that the estimates of the other parameters are given by

$$\hat{\alpha} = -S_{0k}\hat{\beta}(\hat{\beta}'S_{kk}\hat{\beta})^{-1} = -S_{0k}\hat{\beta}, \tag{10}$$

which clearly depends on the choice of the optimising  $\beta$ , whereas

$$\hat{\Pi} = -S_{0k}\hat{\beta}(\hat{\beta}'S_{kk}\hat{\beta})^{-1}\hat{\beta}' = -S_{0k}\hat{\beta}\hat{\beta}', \tag{11}$$

and

$$\hat{\Lambda} = S_{00} - S_{0k} \hat{\beta} \hat{\beta}' S_{k0} = S_{00} - \hat{\alpha} \hat{\alpha}', \quad (12)$$

and the maximised likelihood as given by

$$L_{\max}^{-2/T} = |S_{00}| \prod_{i=1}^r (1 - \hat{\lambda}_i), \quad (13)$$

do not depend on the choice of optimising  $\beta$ .

With these results it is easy to find the estimates of  $\Pi$  and  $\Lambda$  without the constraint (3). These follow from (6) and (7) for  $r = p$  and  $\beta = I$  and give in particular the maximised likelihood function without the constraint (3):

$$L_{\max}^{-2/T} = |S_{00}| \prod_{i=1}^p (1 - \hat{\lambda}_i). \quad (14)$$

If we now want a test that there are at most  $r$  cointegrating vectors, then the likelihood ratio test statistic is the ratio of (13) and (14) which can be expressed as (4), where  $\hat{\lambda}_{r+1} > \dots > \hat{\lambda}_p$  are the  $p - r$  smallest eigenvalues. This completes the proof of Theorem 1.

Notice how this analysis allows one to calculate all  $p$  eigenvalues and eigenvectors at once, and then make inference about the number of important cointegration relations, by testing how many of the  $\lambda$ 's are zero.

Next we shall investigate the test of a linear hypothesis about  $\beta$ . In the case we have  $r = 1$ , i.e., only one cointegration vector, it seems natural to test that certain variables do not enter into the cointegration vector, or that certain linear constraints are satisfied, for instance that the variables  $X_{1t}$  and  $X_{2t}$  only enter through their difference  $X_{1t} - X_{2t}$ . If  $r \geq 2$ , then a hypothesis of interest could be that the variables  $X_{1t}$  and  $X_{2t}$  enter through their difference only in all the cointegration vectors, since if two different linear combinations would occur then any coefficients to  $X_{1t}$  and  $X_{2t}$  would be possible. Thus it seems that some natural hypotheses on  $\beta$  can be formulated as

$$H_1: \beta = H\varphi, \quad (15)$$

where  $H(p \times s)$  is a known matrix of full rank  $s$  and  $\varphi(s \times r)$  is a matrix of unknown parameters. We assume that  $r \leq s \leq p$ . If  $s = p$ , then no restrictions are placed upon the choice of cointegration vectors, and if  $s = r$ , then the cointegration space is fully specified.

*Theorem 2. The maximum likelihood estimator of the cointegration space, under the assumption that it is restricted to  $\text{sp}(H)$ , is given as the space spanned*

by the canonical variates corresponding to the  $r$  largest squared canonical correlations between the residuals of  $H'X_{t-k}$  and  $\Delta X_t$ , corrected for the lagged differences of  $X_t$ .

The likelihood ratio test now becomes

$$-2 \ln(Q) = T \sum_{i=1}^r \ln\left\{\frac{(1 - \lambda_i^*)}{(1 - \hat{\lambda}_i)}\right\}, \quad (16)$$

where  $\lambda_1^*, \dots, \lambda_r^*$  are the  $r$  largest squared canonical correlations.

*Proof.* It is apparent from the derivation of  $\hat{\beta}$  that if  $\beta = H\varphi$  is fixed, then regression of  $R_{0t}$  on  $-\varphi'H'R_{kt}$  is still a simple linear regression and the analysis is as before with  $R_{kt}$  replaced by  $H'R_{kt}$ . Thus the matrix  $\varphi$  can be estimated as the eigenvectors corresponding to the  $r$  largest eigenvalues of  $H'S_{k0}S_{00}^{-1}S_{0k}H$  with respect to  $H'S_{kk}H$ , i.e., the solution to

$$|\lambda H'S_{kk}H - H'S_{k0}S_{00}^{-1}S_{0k}H| = 0.$$

Let the  $s$  eigenvalues be denoted by  $\lambda_i^*$ ,  $i = 1, \dots, s$ . Then the likelihood ratio test of  $H_1$  in  $H_0$  can be found from two expressions like (13) and is given by (16), which completes the proof of Theorem 2.

In the next section we shall find the asymptotic distribution of the test statistics (4) and (16) and show that the cointegration space, the impact matrix  $\Pi$  and the variance matrix  $\Lambda$  are estimated consistently.

### 3. Asymptotic properties of the estimators and the test statistics

In order to derive properties of the estimators we need to impose more precise conditions on the parameters of the model, such that they correspond to the situation we have in mind, namely of a process that is integrated of order 1, but still has  $r$  cointegration vectors  $\beta$ .

First of all we want all roots of  $|A(z)| = 0$  to satisfy  $|z| > 1$  or possibly  $z = 1$ . This implies that the nonstationarity of the process can be removed by differencing. Next we shall assume that  $X_t$  is integrated of order 1, i.e., that  $\Delta X_t$  is stationary and that the hypothesis (3) is satisfied by some  $\alpha$  and  $\beta$  of full rank  $r$ . Correspondingly we can express  $\Delta X_t$  in terms of the  $\varepsilon$ 's by its moving average representation,

$$\Delta X_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j},$$

for some exponentially decreasing coefficients  $C_j$ . Under suitable conditions on these coefficients it is known that this equation determines an error correction

model of the form (5), where  $\Gamma_k X_{t-k} = -\Pi X_{t-k}$  represents the error correction term containing the stationary components of  $X_{t-k}$ , i.e.,  $\beta' X_{t-k}$ . Moreover the null space for  $C' = \sum_{j=0}^{\infty} C_j'$  given by  $\{\xi | C'\xi = 0\}$  is exactly the range space of  $\Gamma_k'$ , i.e., the space spanned by the columns in  $\beta$  and vice versa. We thus have the following representations:

$$\Pi = \alpha\beta' \quad \text{and} \quad C = \gamma\tau\delta', \tag{17}$$

where  $\tau$  is  $(p-r) \times (p-r)$ ,  $\gamma$  and  $\delta$  are  $p \times (p-r)$ , and all three are of full rank, and  $\gamma'\beta = \delta'\alpha = 0$ . We shall later choose  $\delta$  in a convenient way [see the references to Granger (1981), Granger and Engle (1985), Engle and Granger (1987), or Johansen (1988) for the details of these results].

We shall now formulate the main results of this section in the form of two theorems which deal with the test statistics derived in the previous section. First we have a result about the test statistic (4) and the estimators derived in (11) and (12).

*Theorem 3. Under the hypothesis that there are  $r$  cointegrating vectors the estimate of the cointegration space as well as  $\Pi$  and  $\Lambda$  are consistent, and the likelihood ratio test statistic of this hypothesis is asymptotically distributed as*

$$\text{tr} \left\{ \int_0^1 dB B' \left[ \int_0^1 B B' du \right]^{-1} \int_0^1 B dB' \right\}, \tag{18}$$

where  $B$  is a  $(p-r)$ -dimensional Brownian motion with covariance matrix  $I$ .

In order to understand the structure of this limit distribution one should notice that if  $B$  is a Brownian motion with  $I$  as the covariance matrix, then the stochastic integral  $\int_0^t B dB'$  is a matrix-valued martingale, with quadratic variation process

$$\int_0^t \text{var}(B dB') = \int_0^t B B' du \otimes I,$$

where the integral  $\int_0^t B B' du$  is an ordinary integral of the continuous matrix-valued process  $B B'$ . With this notation the limit distribution in Theorem 3 can be considered a multivariate version of the square of a martingale  $\int_0^t B dB'$  divided by its variance process  $\int_0^t B B' du$ . Notice that for  $r = p - 1$ , i.e., for testing  $p - 1$  cointegration relations one obtains the limit distribution with a one-dimensional Brownian motion, i.e.,

$$\left( \int_0^1 B dB \right)^2 / \int_0^1 B^2 du = \left( (B(1)^2 - 1)/2 \right)^2 / \int_0^1 B^2 du,$$



Table 1

The quantiles in the distribution of the test statistic,

$$\text{tr} \left\{ \int_0^1 dB B' \left( \int_0^1 B(u) B(u)' du \right)^{-1} \int_0^1 B dB' \right\},$$

where  $B$  is an  $m$ -dimensional Brownian motion with covariance matrix  $I$ .

$m$	2.5%	5%	10%	50%	90%	95%	97.5%
1	0.0	0.0	0.0	0.6	2.9	4.2	5.3
2	1.6	1.9	2.5	5.4	10.3	12.0	13.9
3	7.0	7.8	8.8	14.0	21.2	23.8	26.1
4	16.0	17.4	19.2	26.3	35.6	38.6	41.2
5	28.3	30.4	32.8	42.1	53.6	57.2	60.3

which is the square of the usual 'unit root' distribution [see Dickey and Fuller (1979)]. A similar reduction is found by Phillips and Ouliaris (1987) in their work on tests for cointegration based on residuals. The distribution of the test statistic (18) is found by simulation and given in table 1.

A surprisingly accurate description of the results in table 1 is obtained by approximating the distributions by  $c\chi^2(f)$  for suitable values of  $c$  and  $f$ . By equating the mean of the distributions based on 10,000 observations to those of  $c\chi^2$  with  $f = 2m^2$  degrees of freedom, we obtain values of  $c$ , and it turns out that we can use the empirical relation

$$c = 0.85 - 0.58/f.$$

Notice that the hypothesis of  $r$  cointegrating relations reduces the number of parameters in the  $\Pi$  matrix from  $p^2$  to  $rp + r(p - r)$ , thus one could expect  $(p - r)^2$  degrees of freedom if the usual asymptotics would hold. In the case of nonstationary processes it is known that this does not hold but a very good approximation is given by the above choice of  $2(p - r)^2$  degrees of freedom.

Next we shall consider the test of the restriction (15) where linear constraints are imposed on  $\beta$ .

*Theorem 4. The likelihood ratio test of the hypothesis*

$$H_1: \beta = H\varphi$$

*of restricting the  $r$ -dimensional cointegration space to an  $s$ -dimensional subspace of  $R^p$  is asymptotically distributed as  $\chi^2$  with  $r(p - s)$  degrees of freedom.*

We shall now give the proof of these theorems, through a series of intermediate results. We shall first give some expressions for variances and their

limits, then show how the algorithm for deriving the maximum likelihood estimator can be followed by a probabilistic analysis ending up with the asymptotic properties of the estimator and the test statistics.

We can represent  $X_t$  as  $X_t = \sum_{j=1}^t \Delta X_j$ , where  $X_0$  is a constant which we shall take to be zero to simplify the notation. We shall describe the stationary process  $\Delta X_t$  by its covariance function

$$\psi(i) = \text{var}(\Delta X_t, \Delta X_{t+i}),$$

and we define the matrices

$$\mu_{ij} = \psi(i-j) = E(\Delta X_{t-i} \Delta X_{t-j}'), \quad i, j = 0, \dots, k-1,$$

$$\mu_{ki} = \sum_{j=k-i}^{\infty} \psi(j), \quad i = 0, \dots, k-1,$$

and

$$\mu_{kk} = - \sum_{j=-\infty}^{\infty} |j| \psi(j).$$

Finally define

$$\psi = \sum_{j=-\infty}^{\infty} \psi(j).$$

Note the following relations:

$$\psi(i) = \sum_{j=0}^{\infty} C_j \Lambda C_{j+i}',$$

$$\psi = \sum_{j=0}^{\infty} C_j \Lambda \sum_{j=0}^{\infty} C_j' = C \Lambda C',$$

$$\text{var}(X_{t-k}) = \sum_{j=-t+k}^{t-k} (t-k-|j|) \psi(j),$$

$$\text{cov}(X_{t-k}, \Delta X_{t-i}) = \sum_{j=k-i}^{t-i} \psi(j),$$

which show that

$$\text{var}(X_T/T^{1/2}) \rightarrow \sum_{i=-\infty}^{\infty} \psi(i) = \psi,$$

and

$$\text{cov}(X_{T-k}, \Delta X_{T-i}) \rightarrow \sum_{j=k-i}^{\infty} \psi(j) = \mu_{ki},$$

whereas the relation

$$\text{var}(\beta' X_{T-k}) = (T-k) \sum_{j=-T+k}^{T-k} \beta' \psi(j) \beta - \sum_{j=-T+k}^{T-k} |j| \beta' \psi(j) \beta$$

shows that

$$\text{var}(\beta' X_{T-k}) \rightarrow \beta' \mu_{kk} \beta,$$

since  $\beta' C = 0$  implies that  $\beta' \psi = 0$ , such that the first term vanishes in the limit. Note that the nonstationary part of  $X_t$  makes the variance matrix tend to infinity, except for the directions given by the vectors in  $\beta$ , since  $\beta' X_t$  is stationary.

The calculations involved in the maximum likelihood estimation all center around the product moment matrices

$$M_{ij} = T^{-1} \sum_{t=1}^T \Delta X_{t-i} \Delta X'_{t-j}, \quad i, j = 0, \dots, k-1,$$

$$M_{ki} = T^{-1} \sum_{t=1}^T X_{t-k} \Delta X'_{t-i}, \quad i = 0, \dots, k-1,$$

$$M_{kk} = T^{-1} \sum_{t=1}^T X_{t-k} X'_{t-k}.$$

We shall first give the asymptotic behaviour of these matrices, then find the asymptotic properties of  $S_{ij}$  and finally apply these results to the estimators and the test statistic. The methods are inspired by Phillips (1985) even though I shall stick to the Gaussian case, which make the arguments somewhat simpler.

The following lemma can be derived by the results in Phillips and Durlauf (1986). We let  $W$  be a Brownian motion in  $p$  dimensions with covariance matrix  $\Lambda$ .

*Lemma 1.* As  $T \rightarrow \infty$ , we have

$$T^{-1/2}X_{[Tt]} \xrightarrow{w} CW(t), \quad (19)$$

$$M_{ij} \xrightarrow{\text{a.s.}} \mu_{ij}, \quad i, j = 0, \dots, k-1, \quad (20)$$

$$M_{ki} \xrightarrow{w} C \int_0^1 W dW' C' + \mu_{ki}, \quad i = 0, \dots, k-1, \quad (21)$$

$$\beta' M_{kk} \beta \xrightarrow{\text{a.s.}} \beta' \mu_{kk} \beta, \quad (22)$$

$$T^{-1}M_{kk} \xrightarrow{w} C \int_0^1 W(u)W'(u) du C'. \quad (23)$$

Note that for any  $\xi \in R^p$ ,  $\xi' M_{kk} \xi$  is of the order in probability of  $T$  unless  $\xi$  is in the space spanned by  $\beta$ , in which case it is convergent. Note also that the stochastic integrals enter as limits of the nonstationary part of the process  $X_t$ , and that they disappear when multiplied by  $\beta$ , since  $\beta' C = 0$ .

We shall apply the results to find the asymptotic properties of  $S_{ij}$ ,  $i, j = 0, k$ , see (8). These can be expressed in terms of the  $M_{ij}$ 's as follows:

$$S_{ij} = M_{ij} - M_{i*} M_{**}^{-1} M_{*j}, \quad i, j = 0, k,$$

where

$$M_{**} = \{M_{ij}, i, j = 1, \dots, k-1\},$$

$$M_{k*} = \{M_{ki}, i = 1, \dots, k-1\},$$

$$M_{0*} = \{M_{0i}, i = 1, \dots, k-1\}.$$

A similar notation is introduced for the  $\mu_{ij}$ 's. It is convenient to have the notation

$$\Sigma_{ij} = \mu_{ij} - \mu_{i*} \mu_{**}^{-1} \mu_{*j}, \quad i, j = 0, k.$$

We now get:

*Lemma 2.* The following relations hold

$$\Sigma_{00} = \Gamma_k \Sigma_{k0} + \Lambda, \quad (24)$$

$$\Sigma_{0k} \Gamma_k' = \Gamma_k \Sigma_{kk} \Gamma_k', \quad (25)$$

and hence, since  $\Gamma_k = -\alpha\beta'$ .

$$\Sigma_{00} = \alpha(\beta'\Sigma_{kk}\beta)\alpha' + \Lambda \quad \text{and} \quad \alpha = -\Sigma_{0k}\beta(\beta'\Sigma_{kk}\beta)^{-1}. \quad (26)$$

*Proof.* From the defining eq. (5) for the process  $X_t$  we find the equations

$$M_{k,i} = \Gamma_1 M_{1i} + \cdots + \Gamma_{k-1} M_{k-1,i} + \Gamma_k M_{ki} + T^{-1} \sum_{t=1}^T \varepsilon_t \Delta X'_{t-i}, \quad (27)$$

$i = 0, 1, \dots, k-1$ , and

$$M_{0k} = \Gamma_1 M_{1k} + \cdots + \Gamma_{k-1} M_{k-1,k} + \Gamma_k M_{kk} + T^{-1} \sum_{t=1}^T \varepsilon_t X'_{t-k}. \quad (28)$$

Now let  $T \rightarrow \infty$ , then we get the equations

$$\mu_{00} = \Gamma_1 \mu_{10} + \cdots + \Gamma_{k-1} \mu_{k-1,0} + \Gamma_k \mu_{k0} + \Lambda, \quad (29)$$

$$\mu_{0i} = \Gamma_1 \mu_{1i} + \cdots + \Gamma_{k-1} \mu_{k-1,i} + \Gamma_k \mu_{ki}, \quad i = 1, \dots, k-1, \quad (30)$$

$$\mu_{0k}\beta = \Gamma_1 \mu_{1k}\beta + \cdots + \Gamma_{k-1} \mu_{k-1,k}\beta + \Gamma_k \mu_{kk}\beta. \quad (31)$$

If we solve the eq. (30) for the matrices  $\Gamma_*$  and insert into (29) and (31), we get (24), (25), and (26).

We shall now find the asymptotic properties of  $S_{ij}$ .

*Lemma 3.* For  $T \rightarrow \infty$ , it holds that, if  $\delta$  is chosen such that  $\delta'\alpha = 0$ , then

$$S_{00} \xrightarrow{\text{a.s.}} \Sigma_{00}, \quad (32)$$

$$\delta'S_{0k} \xrightarrow{w} \delta' \int_0^1 dW W' C', \quad (33)$$

$$\beta'S_{k0} \xrightarrow{\text{a.s.}} \beta'\Sigma_{k0}, \quad (34)$$

$$T^{-1}S_{kk} \xrightarrow{w} C \int_0^1 W(u) W'(u) du C'. \quad (35)$$

$$\beta'S_{kk}\beta \xrightarrow{\text{a.s.}} \beta'\Sigma_{kk}\beta. \quad (36)$$

*Proof.* All relations follow from Lemma 1 except the second. If we solve for  $\Gamma_*$  in the eq. (27), insert the solution into (28), and use the definition of  $S_{ij}$  in

terms of the  $M$ 's, then we get

$$S_{0k} = T^{-1} \sum_{i=1}^T \varepsilon_i X'_{t-k} + \Gamma_k S_{kk} - \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} T^{-1} \sum_{i=1}^T \varepsilon_i \Delta X'_{t-i} M^{ij} M_{jk}. \tag{37}$$

The last term goes a.s. to zero as  $T \rightarrow \infty$ , since  $\varepsilon_t$  and  $\Delta X'_{t-i}$  are stationary and uncorrelated. The second term vanishes when multiplied by  $\delta'$ , since  $\delta' \Gamma_k = -\delta' \alpha \beta' = 0$ , and the first term converges to the integral as stated.

*Lemma 4. The ordered eigenvalues of the equation*

$$|\lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}| = 0 \tag{38}$$

*converge in probability to  $(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$ , where  $\lambda_1, \dots, \lambda_r$  are the ordered eigenvalues of the equation*

$$|\lambda \beta' \Sigma_{kk} \beta - \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta| = 0. \tag{39}$$

*Proof.* We want to express the problem in the coordinates given by the  $p$  vectors in  $\beta$  and  $\gamma$ , where  $\gamma$  is of full rank  $p - r$  and  $\gamma' \beta = 0$ . This can be done by multiplying (9) by  $|(\beta, \gamma)'$  and  $|(\beta, \gamma)|$  from the left and right, then the eigenvalues solve the equation

$$\left| \lambda \begin{bmatrix} \beta' S_{kk} \beta & \beta' S_{kk} \gamma \\ \gamma' S_{kk} \beta & \gamma' S_{kk} \gamma \end{bmatrix} - \begin{bmatrix} \beta' S_{k0} S_{00}^{-1} S_{0k} \beta & \beta' S_{k0} S_{00}^{-1} S_{0k} \gamma \\ \gamma' S_{k0} S_{00}^{-1} S_{0k} \beta & \gamma' S_{k0} S_{00}^{-1} S_{0k} \gamma \end{bmatrix} \right| = 0.$$

We define  $A_T = (\gamma' S_{kk} \gamma)^{-1/2}$  so that by (35),  $A_T \xrightarrow{P} 0$ . Then the eigenvalues have to satisfy the equation

$$\left| \lambda \begin{bmatrix} \beta' S_{kk} \beta & \beta' S_{kk} \gamma A_T \\ A_T' \gamma' S_{kk} \beta & I \end{bmatrix} - \begin{bmatrix} \beta' S_{k0} S_{00}^{-1} S_{0k} \beta & \beta' S_{k0} S_{00}^{-1} S_{0k} \gamma A_T \\ A_T' \gamma' S_{k0} S_{00}^{-1} S_{0k} \beta & A_T' \gamma' S_{k0} S_{00}^{-1} S_{0k} \gamma A_T \end{bmatrix} \right| = 0.$$

The coefficient matrices converge in probability by the results in Lemma 3, and the limiting equation is

$$\left| \lambda \begin{bmatrix} \beta' \Sigma_{kk} \beta & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta & 0 \\ 0 & 0 \end{bmatrix} \right| = 0,$$

or

$$|\lambda \beta' \Sigma_{kk} \beta - \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta| |\lambda I| = 0,$$

where  $I$  is an identity matrix of dimension  $p - r$ , which means that the equation has  $p - r$  roots at  $\lambda = 0$ . It is known that the ordered eigenvalues are continuous functions of the coefficient matrices [see Andersson, Brøns and Jensen (1983)], and hence the statement of Lemma 4 is proved.

We shall need one more technical lemma before we can prove the more useful results. We decompose the eigenvectors  $\hat{\beta}$  as follows:  $\hat{\beta}_i = \beta \hat{x}_i + \gamma \hat{y}_i$ , where  $\hat{x}_i = (\beta' \beta)^{-1} \beta' \hat{\beta}_i$  and  $\hat{y}_i = (\gamma' \gamma)^{-1} \gamma' \hat{\beta}_i$ . Let  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_r) = (\beta' \beta)^{-1} \beta' \hat{\beta}$ . Note that although we have proved that the eigenvalues  $\lambda_i$  are convergent, the same cannot hold for the eigenvectors  $\hat{\beta}_i$  or  $\hat{x}_i$ , since if the limiting eq. (39) has multiple roots then the eigenvectors are not uniquely defined. This complicates the formulation below somewhat. Let

$$S(\lambda) = \lambda S_{kk} - S_{k0} S_{00}^{-1} S_{0k}.$$

*Lemma 5.* For  $i = 1, \dots, r$ , we have

$$\gamma' S(\hat{\lambda}_i) \gamma / T \xrightarrow{w} \lambda_i \gamma' C \int_0^1 W W' du C' \gamma, \quad (40)$$

$$\hat{y}_i \in O_p(T^{-1}), \quad \hat{x} \in O_p(1), \quad \hat{x}^{-1} \in O_p(1), \quad (41)$$

$$\beta' S(\hat{\lambda}_i) \beta \hat{x}_i \in O_p(T^{-1}), \quad (42)$$

$$\gamma' S(\hat{\lambda}_i) \beta \hat{x}_i = -\gamma' \left[ T^{-1} \sum_{i=1}^T X_{i-k} \epsilon'_i \right] \Sigma_{00}^{-1} \Sigma_{0k} \beta \hat{x}_i + o_p(1). \quad (43)$$

*Proof.* The relation (40) follows directly from (35) in Lemma 3, since by Lemma 4,  $\hat{\lambda}_i \xrightarrow{p} \lambda_i > 0$ .

The normalisation  $\hat{\beta}' S_{kk} \hat{\beta} = I$  implies by an argument similar to that of Lemma 4 for the eigenvalues of  $S_{kk}$  that  $\hat{\beta}$  and hence  $\hat{x}$  is bounded in probability. The eigenvectors  $\hat{\beta}_i$  and eigenvalues  $\hat{\lambda}_i$  satisfy

$$\beta' S(\hat{\lambda}_i) \beta \hat{x}_i + \beta' S(\hat{\lambda}_i) \gamma \hat{y}_i = 0, \quad (44)$$

$$\gamma' S(\hat{\lambda}_i) \beta \hat{x}_i + \gamma' S(\hat{\lambda}_i) \gamma \hat{y}_i = 0. \quad (45)$$

Now (40) and (45) imply that  $\hat{y}_i \in O_p(T^{-1})$ , and hence from (44) we find that  $\beta'S(\hat{\lambda}_i)\beta\hat{x}_i \in O_p(T^{-1})$ , which shows (42), and also from the normalising condition, that  $\hat{x}'\beta'S_{kk}\beta\hat{x} \xrightarrow{P} I$ . Hence  $|\hat{x}|^2|\beta'S_{kk}\beta| \xrightarrow{P} 1$ , which implies that also  $\hat{x}^{-1}$  is bounded in probability, which proves (41). From (37) it follows that

$$\gamma'S_{k0} + \gamma'S_{kk}\beta\alpha' = \gamma'T^{-1} \sum_{t=1}^T X_{t-k}\varepsilon'_t + o_p(1).$$

Now replace  $\alpha = -\Sigma_{0k}\beta(\beta'\Sigma_{kk}\beta)^{-1}$  [see (26)] by the consistent estimate  $-S_{0k}\beta(\beta'S_{kk}\beta)^{-1}$ , then

$$\begin{aligned} \gamma'S(\hat{\lambda}_i)\beta\hat{x}_i &= \gamma'S_{kk}\beta(\beta'S_{kk}\beta)^{-1}(\hat{\lambda}_i\beta'S_{kk}\beta - \beta'S_{k0}S_{00}^{-1}S_{0k}\beta)\hat{x}_i \\ &\quad - \gamma'T^{-1} \sum_{t=1}^T X_{t-k}\varepsilon'_t\Sigma_{00}^{-1}\Sigma_{0k}\beta\hat{x}_i + o_p(1). \end{aligned}$$

The first term contains the factor  $\beta'S(\hat{\lambda}_i)\beta\hat{x}_i$  which tends to zero in probability by (42). This completes the proof of Lemma 5.

We shall now choose  $\delta$  of dimension  $p \times (p - r)$  such that  $\delta'\alpha = 0$  [see (17)] in the following way. We let  $P_\alpha(\Lambda)$  denote the projection of  $R^p$  onto the column space spanned by  $\alpha$  with respect to the matrix  $\Lambda^{-1}$ , i.e.,

$$P_\alpha(\Lambda) = \alpha(\alpha'\Lambda^{-1}\alpha)^{-1}\alpha'\Lambda^{-1}.$$

We then choose  $\delta$  of full rank  $p - r$  to satisfy

$$\delta\delta' = \Lambda^{-1}(I - P_\alpha(\Lambda)).$$

Note that  $\delta'\alpha = 0$ , and that  $\delta'\Lambda\delta = I$  of dimension  $(p - r) \times (p - r)$ . Note also that  $P_\alpha(\Lambda) = P_\alpha(\Sigma_{00})$  since  $\Sigma_{00}$  is given by (26). This relation is well-known from the theory of random coefficient regression [see Rao (1965) or Johansen (1984)].

*Lemma 6.* For  $T \rightarrow \infty$ , we have that  $T\hat{\lambda}_{r+1}, \dots, T\hat{\lambda}_p$  converge in distribution to the ordered eigenvalues of the equation

$$\left| \lambda \int_0^1 BB' du - \int_0^1 B dB' \int_0^1 dB B' \right| = 0, \tag{46}$$

where  $B$  is a Brownian motion in  $p - r$  dimensions with covariance matrix  $I$ .



*Proof.* We shall consider the ordered eigenvalues of the equation

$$\left| \begin{bmatrix} \beta' S_{kk} \beta / T & \beta' S_{kk} \gamma / T \\ \gamma' S_{kk} \beta / T & \gamma' S_{kk} \gamma / T \end{bmatrix} - \mu \begin{bmatrix} \beta' S_{k0} S_{00}^{-1} S_{0k} \beta & \beta' S_{k0} S_{00}^{-1} S_{0k} \gamma \\ \gamma' S_{k0} S_{00}^{-1} S_{0k} \beta & \gamma' S_{k0} S_{00}^{-1} S_{0k} \gamma \end{bmatrix} \right| = 0.$$

For any value of  $T$  the ordered eigenvalues are

$$\hat{\mu}_1 = (T \hat{\lambda}_p)^{-1}, \dots, \hat{\mu}_p = (T \hat{\lambda}_1)^{-1}.$$

Since the ordered eigenvalues are continuous functions of the coefficients, we find from Lemma 3 that  $\hat{\mu}_1, \dots, \hat{\mu}_p$  converge in distribution to the ordered eigenvalues from the equation

$$\left| \begin{bmatrix} 0 & 0 \\ 0 & \gamma' C \int_0^1 W W' du C' \gamma \end{bmatrix} - \mu \begin{bmatrix} \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta & \beta' \Sigma_{k0} \Sigma_{00}^{-1} F \\ F' \Sigma_{00}^{-1} \Sigma_{0k} \beta & F' \Sigma_{00}^{-1} F \end{bmatrix} \right| = 0,$$

where  $F$  is the weak limit of  $S_{0k} \gamma$ .

This determinant can also be written as

$$\begin{aligned} & \left| \mu \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta \right| \left| \gamma' C \int_0^1 W W' du C' \gamma \right. \\ & \left. - \mu F' \left[ \Sigma_{00}^{-1} - \Sigma_{00}^{-1} \Sigma_{0k} \beta (\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)^{-1} \beta' \Sigma_{k0} \Sigma_{00}^{-1} \right] F \right|, \end{aligned} \quad (47)$$

which shows that in the limit there are  $r$  roots at zero. Now apply (24) and (25) of Lemma 2 and find that

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1} \Sigma_{0k} \beta (\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)^{-1} \beta' \Sigma_{k0} \Sigma_{00}^{-1}$$

equals

$$\Sigma_{00}^{-1} (I - P_\alpha(\Sigma_{00})) = \Lambda^{-1} (I - P_\alpha(\Lambda)) = \delta \delta',$$

and hence that the second factor of (47) is

$$\left| \gamma' C \int_0^1 W W' du C' \gamma - \mu F' \delta \delta' F \right|,$$

which by (33) equals

$$\left| \gamma' C \int_0^1 W W' du C' \gamma - \mu \gamma' C \int_0^1 W dW' \delta \delta' \int_0^1 dW W' C' \gamma \right|. \quad (48)$$

Thus the limiting distribution of the  $p - r$  largest  $\mu$ 's is given as the distribution of the ordered eigenvalues of (48).

The representation (17):  $C = \gamma \tau \delta'$  for some nonsingular matrix  $\tau$  now implies, since  $|\gamma' \gamma| \neq 0$  and  $|\tau| \neq 0$ , that

$$\left| \delta' \int_0^1 W W' du \delta - \mu \delta' \int_0^1 W dW' \delta \delta' \int_0^1 dW W' \delta \right| = 0. \quad (49)$$

Now  $B = \delta' W$  is a Brownian motion with variance  $\delta' \Lambda \delta = I$ , and the result of Lemma 6 is found by noting that the solutions of (46) are the reciprocal values of the solutions to the above equation.

*Corollary.* The test statistic for  $H_0: \Pi = \alpha \beta'$  given by (4) will converge in distribution to the sum of the eigenvalues given by (46), i.e., the limiting distribution is given by (18):

$$\text{tr} \left\{ \int_0^1 d B B' \left( \int_0^1 B B' du \right)^{-1} \int_0^1 B dB' \right\}.$$

*Proof.* We just expand the test statistic (4),

$$-2 \ln(Q) = -T \sum_{i=r+1}^p \ln(1 - \hat{\lambda}_i) = \sum_{i=r+1}^p T \hat{\lambda}_i + o_p(1),$$

and apply Lemma 6.

We can now complete the proof of Theorem 3.

The asymptotic distribution of the test statistic (4) follows from the above corollary, and the consistency of the cointegration space and the estimators of  $\Pi$  and  $\Lambda$  is proved as follows. We decompose  $\hat{\beta} = \beta \hat{x} + \gamma \hat{y}$ , and it then follows from (41) that  $\hat{\beta} \hat{x}^{-1} - \beta = \gamma \hat{y} \hat{x}^{-1} \in O_p(T^{-1})$ . Thus we have seen that the projection of  $\beta$  onto the orthogonal complement of  $\text{sp}(\beta)$  tends to zero in probability of the order of  $T^{-1}$ . In this sense the cointegration space is consistently estimated. From (11) we find

$$\hat{\Pi} = -S_{0k} \hat{\beta} (\hat{\beta} S_{kk} \hat{\beta})^{-1} \hat{\beta}' = -S_{0k} \hat{\beta} \hat{x}^{-1} \left( (\hat{\beta} \hat{x}^{-1})' S_{kk} \hat{\beta} \hat{x}^{-1} \right)^{-1} (\hat{\beta} \hat{x}^{-1})',$$

which converges in probability to

$$-\Sigma_{0k}\beta(\beta'\Sigma_{kk}\beta)^{-1}\beta' = \alpha\beta' = \Pi.$$

From (12) we get by a similar trick that

$$\begin{aligned} \hat{\Lambda} &= S_{00} - S_{0k}\hat{\beta}(\hat{\beta}'S_{kk}\hat{\beta})^{-1}\hat{\beta}'S_{k0} \\ &\xrightarrow{P} \Sigma_{00} - \Sigma_{0k}\beta(\beta'\Sigma_{kk}\beta)^{-1}\beta'\Sigma_{k0} = \Sigma_{00} - \alpha(\beta'\Sigma_{kk}\beta)\alpha', \end{aligned}$$

which by (26) equals  $\Lambda$ . This completes the proof of Theorem 3.

In order to prove the Theorem 4 we shall need an expansion of the likelihood function around a point where the maximum is attained. We formulate this as a lemma:

*Lemma 7. Let  $A$  and  $B$  be  $p \times p$  symmetric positive definite matrices, and define the function*

$$f(z) = \ln\{|z'Az|/|z'Bz|\},$$

where  $z$  is a  $p \times r$  matrix. If  $z$  is a point where the function attains its maximum or minimum, then for any  $p \times r$  matrix  $h$  we have

$$f(z+h) = f(z) + \text{tr}\{(z'Az)^{-1}h'Ah - (z'Bz)^{-1}h'Bh\} + O(h^3).$$

*Proof.* This is easily seen by expanding the terms in  $f$  using that the first derivative vanishes at  $z$ .

We shall now give the asymptotic distribution of the maximum likelihood estimator  $\hat{\beta}$  suitably normalised. This is not a very useful result in practise, since the normalisation depends on  $\beta$ , but it is convenient for deriving other results of interest.

*Lemma 8. The maximum likelihood estimator  $\hat{\beta}$  has the representation*

$$\begin{aligned} &T\left[\hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\beta - \beta\right] \\ &= \gamma(\gamma'S_{kk}\gamma/T)^{-1}\gamma'\left[T^{-1}\sum_{t=1}^T X_{t-k}\varepsilon'_t\right]\Sigma_{00}^{-1}\Sigma_{0k}\beta(\beta'\Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k}\beta)^{-1} \\ &\quad \times \beta'\Sigma_{kk}\beta + o_p(1), \end{aligned} \tag{50}$$

which converges in distribution to

$$\gamma \left( \int_0^1 UU' du \right)^{-1} \int_0^1 U dV',$$

where  $U$  and  $V$  are independent Brownian motions given by

$$U = \gamma' CW, \quad (51)$$

$$V = \beta' \Sigma_{kk} \beta (\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)^{-1} \beta' \Sigma_{k0} \Sigma_{00}^{-1} W. \quad (52)$$

The variance of  $V$  is given by

$$\text{var}(V) = \beta' \Sigma_{kk} \beta (\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)^{-1} \beta' \Sigma_{kk} \beta - \beta' \Sigma_{kk} \beta. \quad (53)$$

Note that the limiting distribution for fixed  $U$  is Gaussian with mean zero and variance

$$\gamma \int_0^1 UU' du \gamma' \otimes \text{var}(V).$$

Thus the limiting distribution is a mixture of Gaussian distributions. This will be used in the derivation of the limiting distributions of the test statistics [see Lemma 9].

*Proof.* From the decomposition  $\hat{\beta}_i = \beta \hat{x}_i + \gamma \hat{y}_i$  we have from (45)

$$T(\hat{\beta}_i - \beta \hat{x}_i) = T\gamma \hat{y}_i = -T\gamma (\gamma' S(\hat{\lambda}_i) \gamma)^{-1} (\gamma' S(\hat{\lambda}_i) \beta) \hat{x}_i.$$

It follows from (40) and (43) in Lemma 5 that this can be written as

$$\gamma (\gamma' S_{kk} \gamma / T)^{-1} \gamma' \left[ T^{-1} \sum_{i=1}^T X_{i-k} \varepsilon'_i \right] \Sigma_{00}^{-1} \Sigma_{0k} \beta \hat{x}_i \hat{\lambda}_i^{-1} + o_p(1).$$

Hence for  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_r) = (\beta' \beta)^{-1} \beta' \hat{\beta}$  and  $\hat{D}_r = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_r)$ ,

$$\begin{aligned} & T(\hat{\beta} - \beta \hat{x}) \\ &= \gamma (\gamma' S_{kk} \gamma / T)^{-1} \gamma' \left[ T^{-1} \sum_{i=1}^T X_{i-k} \varepsilon'_i \right] \Sigma_{00}^{-1} \Sigma_{0k} \beta \hat{x} \hat{D}_r^{-1} + o_p(1). \end{aligned} \quad (54)$$

Now multiply by  $\hat{x}^{-1}$  from the right and note that it follows from (42) that

$$\beta' S_{kk} \beta \hat{x} \hat{D}_r - \beta' S_{k0} S_{00}^{-1} S_{0k} \beta \hat{x} \in o_p(1),$$

which shows that

$$\hat{x} \hat{D}_r^{-1} \hat{x}^{-1} \xrightarrow{P} (\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)^{-1} \beta' \Sigma_{kk} \beta.$$

Now insert this into (54) and we obtain the representation (50), which converges as indicated with  $U$  and  $V$  defined as in (51) and (52). The variance matrix for  $V$  is calculated from (52) using the relation (26). Similarly the independence follows from (26).

*Lemma 9. The likelihood ratio test of the hypothesis of a completely specified  $\beta$  has an asymptotic representation of the form*

$$\begin{aligned} -2 \ln(Q) &= T \operatorname{tr} \left\{ \operatorname{var}(V)^{-1} (\hat{\beta} (\beta' \hat{\beta})^{-1} \beta' \beta - \beta)' \right. \\ &\quad \left. \times S_{kk} (\hat{\beta} (\beta' \hat{\beta})^{-1} \beta' \beta - \beta) \right\} + o_p(1), \end{aligned} \tag{55}$$

which converges weakly to

$$\operatorname{tr} \left\{ \operatorname{var}(V)^{-1} \int_0^1 dV U' \left( \int_0^1 U U' du \right)^{-1} \int_0^1 U dV' \right\}, \tag{56}$$

where  $U$  and  $V$  are given by (51) and (52).

The statistic (56) has a  $\chi^2$  distribution with  $r(p-r)$  degrees of freedom.

*Proof.* The likelihood ratio test statistic of a simple hypothesis about  $\beta$  has the form

$$\begin{aligned} T \{ \ln [ |\beta' (S_{kk} - S_{k0} S_{00}^{-1} S_{0k}) \beta| / |\beta' S_{kk} \beta| ] \\ - \ln [ |\hat{\beta}' (S_{kk} - S_{k0} S_{00}^{-1} S_{0k}) \hat{\beta}| / |\hat{\beta}' S_{kk} \hat{\beta}| ] \}. \end{aligned}$$

Now replace  $\hat{\beta}$  by  $\hat{\beta} \hat{x}^{-1} = \tilde{\beta}$ , say, which is also a maximum point for the likelihood function. Then the statistic takes the form  $T\{f(\beta) - f(\tilde{\beta})\}$  for  $A = S_{kk} - S_{k0} S_{00}^{-1} S_{0k}$  and  $B = S_{kk}$  [see Lemma 7]. By the result in Lemma 7 it

follows that this can be expressed as

$$\begin{aligned} & T \operatorname{tr}\left\{(\tilde{\beta}'A\tilde{\beta})^{-1}(\tilde{\beta}-\beta)'A(\tilde{\beta}-\beta) - (\tilde{\beta}'B\tilde{\beta})^{-1}(\tilde{\beta}-\beta)'B(\tilde{\beta}-\beta)\right\} \\ & \quad + O_p\left(T(\tilde{\beta}-\beta)^3\right) \\ & = T \operatorname{tr}\left\{\left[(\tilde{\beta}'A\tilde{\beta})^{-1} - (\tilde{\beta}'B\tilde{\beta})^{-1}\right](\tilde{\beta}-\beta)'S_{kk}(\tilde{\beta}-\beta)\right\} \\ & \quad - T \operatorname{tr}\left\{(\tilde{\beta}'A\tilde{\beta})^{-1}(\tilde{\beta}-\beta)'S_{k0}S_{00}^{-1}S_{0k}(\tilde{\beta}-\beta)\right\} + O_p\left(T(\tilde{\beta}-\beta)^3\right). \end{aligned}$$

Now use the result from Lemma 8 that  $\tilde{\beta}-\beta \in O_p(T^{-1})$  to see that the last two terms tend to zero in probability. We also get

$$\begin{aligned} & \left[\tilde{\beta}'(S_{kk} - S_{k0}S_{00}^{-1}S_{0k})\tilde{\beta}\right]^{-1} - \left[\tilde{\beta}'S_{kk}\tilde{\beta}\right]^{-1} \\ & \xrightarrow{P} \left[\beta'(\Sigma_{kk} - \Sigma_{k0}\Sigma_{00}^{-1}\Sigma_{0k})\beta\right]^{-1} - \left[\beta'\Sigma_{kk}\beta\right]^{-1} = \operatorname{var}(V)^{-1}. \end{aligned}$$

Finally we note that  $T(\tilde{\beta}-\beta)'S_{kk}(\tilde{\beta}-\beta)$  converges in distribution to

$$\int_0^1 dV U' \left( \int_0^1 U U' du \right)^{-1} \int_0^1 U dV'.$$

Now use the fact that for given value of  $U$  the  $(p-r) \times r$  matrix  $\int_0^1 U dV'$  is Gaussian with mean zero and variance matrix

$$\int_0^1 U U' du \otimes \operatorname{var}(V),$$

hence the distribution of (56) for fixed  $U$  is  $\chi^2$  with  $(p-r)r$  degrees of freedom. Since this result holds independently of the given value of  $U$  it also holds marginally. This basic independence and the conditioning argument is also used in the work of Phillips and Park (1986b) in the discussion of the regression estimates for integrated processes.

We can now finally give the proof of Theorem 4.

We want to test the hypothesis  $H_1: \beta = H\varphi$ .

We now choose  $\psi$  ( $s \times (s-r)$ ) to supplement  $\varphi$  ( $s \times r$ ) such that  $(\psi, \varphi)$  ( $s \times s$ ) has full rank. We can choose  $\psi$  such that  $H\psi = \gamma\eta$  for some  $\eta$  ( $(p-r) \times (s-r)$ ).

The test statistic (16) can be expressed as the difference of two test statistics we get by testing a simple hypothesis for  $\beta$ . Thus we can use the representation (55) and (50) for both statistics and we find that it has a weak limit, which can be expressed as

$$\begin{aligned} -2 \ln(Q) \xrightarrow{w} & \operatorname{tr} \left\{ \operatorname{var}(V)^{-1} \int_0^1 dV U' \left( \int_0^1 U U' d u \right)^{-1} \int_0^1 U dV' \right. \\ & \left. - \operatorname{var}(V_H)^{-1} \int_0^1 dV_H U_H' \left( \int_0^1 U_H U_H' d u \right)^{-1} \int_0^1 U_H dV_H' \right\}, \end{aligned}$$

where

$$U_H = \psi' H' C W = \eta' \gamma' C W = \eta' U,$$

and

$$V_H = \varphi' H' \Sigma_{kk} H \varphi \left( \varphi' H' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} H \varphi \right)^{-1} \varphi' H' \Sigma_{k0} \Sigma_{00}^{-1} W = V.$$

Thus

$$\begin{aligned} -2 \ln(Q) \xrightarrow{w} & \operatorname{tr} \left\{ \left[ \int_0^1 dV U' \left( \int_0^1 U U' d u \right)^{-1} \int_0^1 U dV' \right. \right. \\ & \left. \left. - \int_0^1 dV U' \eta \left( \eta' \int_0^1 U U' d u \eta \right)^{-1} \eta' \int_0^1 U dV' \right] \operatorname{var}(V)^{-1} \right\}. \end{aligned}$$

For fixed value of  $U$  the  $(p-r) \times r$  matrix

$$Y = \left( \int_0^1 U U' d u \right)^{-1/2} \int_0^1 U dV' \operatorname{var}(V)^{-1/2}$$

is Gaussian with variance matrix  $I \otimes I$ . Let  $\tilde{\eta} = \left( \int_0^1 U U' d u \right)^{-1/2} \eta$ , then the decomposition into independent components, given by

$$\operatorname{tr}\{Y'Y\} = \operatorname{tr}\{Y'(I - \tilde{\eta}(\tilde{\eta}'\tilde{\eta})^{-1}\tilde{\eta}')Y\} + \operatorname{tr}\{Y'\tilde{\eta}(\tilde{\eta}'\tilde{\eta})^{-1}\tilde{\eta}'Y\},$$

shows that each term is  $\chi^2$  distributed with degrees of freedom  $r(p-r)$ ,  $r(p-s)$  and  $r(s-r)$ , respectively. Thus the limiting distribution of  $-2 \ln(Q)$  is for fixed  $U$  a  $\chi^2$  distribution with  $(p-s)r$  degrees of freedom. Since this result does not depend on the value of  $U$  it holds marginally. Thus the proof of Theorem 4 is completed.

## References

- Ahn, S.K. and G.C. Reinsel, 1987, Estimation for partially nonstationary multivariate autoregressive models (University of Wisconsin, Madison, WI).
- Anderson, T.W., 1984, An introduction to multivariate statistical analysis (Wiley, New York).
- Andersson, S.A., H.K. Brøns and S.T. Jensen, 1983, Distribution of eigenvalues in multivariate statistical analysis, *Annals of Statistics* 11, 392–415.
- Davidson, J., 1986, Cointegration in linear dynamic systems, Mimeo. (London School of Economics, London).
- Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427–431.
- Engle, R.F. and C.W.J. Granger, 1987, Co-integration and error correction: Representation, estimation and testing, *Econometrica* 55, 251–276.
- Granger, C.J., 1981, Some properties of time series data and their use in econometric model specification, *Journal of Econometrics* 16, 121–130.
- Granger, C.W.J. and R.F. Engle, 1985, Dynamic model specification with equilibrium constraints, Mimeo. (University of California, San Diego, CA).
- Granger, C.W.J. and A.A. Weiss, 1983, Time series analysis of error correction models, in: S. Karlin, T. Amemiya and L.A. Goodman eds., *Studies in economic time series and multivariate statistics* (Academic Press, New York).
- Johansen, S., 1984, Functional relations, random coefficients, and non-linear regression with application to kinetic data, *Lecture notes in statistics* (Springer, New York).
- Johansen, S., 1988, The mathematical structure of error correction models, *Contemporary Mathematics*, in press.
- Phillips, P.C.B., 1985, Understanding spurious regression in econometrics, Cowles Foundation discussion paper no. 757.
- Phillips, P.C.B., 1987, Multiple regression with integrated time series, Cowles Foundation discussion paper no. 852.
- Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series regression with integrated processes, *Review of Economic Studies* 53, 473–495.
- Phillips, P.C.B. and S. Ouliaris, 1986, Testing for cointegration, Cowles Foundation discussion paper no. 809.
- Phillips, P.C.B. and S. Ouliaris, 1987, Asymptotic properties of residual based tests for cointegration, Cowles Foundation discussion paper no. 847.
- Phillips, P.C.B. and J.Y. Park, 1986a, Asymptotic equivalence of OLS and GLS in regression with integrated regressors, Cowles Foundation discussion paper no. 802.
- Phillips, P.C.B. and J.Y. Park, 1986b, Statistical inference in regressions with integrated processes: Part 1, Cowles Foundation discussion paper no. 811.
- Phillips, P.C.B. and J.Y. Park, 1987, Statistical inference in regressions with integrated processes: Part 2, Cowles Foundation discussion paper no. 819.
- Rao, C.R., 1965, The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves, *Biometrika* 52, 447–458.
- Rao, C.R. 1973, *Linear statistical inference and its applications*, 2nd ed. (Wiley, New York).
- Sims, A., J.H., Stock and M.W. Watson, 1986, Inference in linear time series models with some unit roots, Preprint.
- Stock, J.H., 1987, Asymptotic properties of least squares estimates of cointegration vectors, *Econometrica* 55, 1035–1056.
- Stock, J.H. and M.W. Watson, 1987, Testing for common trends, Working paper in econometrics (Hoover Institution, Stanford, CA).
- Velu, R.P., G.C. Reinsel and D.W. Wichern, 1986, Reduced rank models for multiple time series, *Biometrika* 73, 105–118.