Søren Johansen

The Power Function of the Likelihood Ratio Test for Cointegration

Preprint
September
1989

Institute of Mathematical Statistics
University of Copenhagen
Søren Johansen

THE POWER FUNCTION OF THE LIKELIHOOD RATIO TEST
FOR COINTEGRATION

Preprint 1989 No. 8

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

September 1989
The distribution of the likelihood ratio test for cointegration under local alternatives is found to be given by a suitable functional of a p-r dimensional Ornstein-Uhlenbeck process. The results are related to those of Phillips (1988), who considered near integrated processes, and derived the power function of a test for unit roots based on a regression estimate. The power function is investigated numerically for a one-dimensional alternative.
1. Introduction

The concept of cointegration was introduced by Granger (1981) in order to define the notion of a stable economic relation among non stable economic variables.

More precisely he considered non stationary economic variables, i.e. a non stationary vector process, and defined a cointegrating relation as a linear combination of the components with the property that it determined a stationary stochastic process. This makes precise one of the many meanings of the notion of stability and as such it can be investigated and tested by statistical techniques. see Engle and Granger (1987), Phillips and Park (1986) and Phillips and Ouliaris (1987). Many papers have since then been devoted to finding properties of various regression and eigenvector estimates for cointegration vectors.

This paper deals with vector autoregressive processes with Gaussian errors where maximum likelihood estimators and likelihood ratio tests can be found, see Johansen (1988b). There are now a number of papers that describe this method in detail, and apply it to economic problems, see Johansen and Juselius (1989), Juselius (1989), Kunst (1988) and (1989), Kunst and Neusser (1988), Lütkepohl and Reimers (1989), Hoffman and Rasche (1989), Hall (1988), Garbers (1989). The method will be described briefly below.

Consider therefore the p-dimensional vector autoregressive process $X_t$, $t=1,...,T$, defined by the equations:

\[(1.1) \quad X_t = \Pi_1 X_{t-1} + \ldots + \Pi_k X_{t-k} + \epsilon_t, \quad t = 1,...,T.\]
where \( \epsilon_1, \ldots, \epsilon_T \) are independent Gaussian variables with mean zero and variance matrix \( A \) and \( X_0, \ldots, X_{-k+1} \) are fixed. The parameters are the \( p \times p \) matrices \( (\Pi_1, \ldots, \Pi_k, A) \). Let \( \Pi = -I + \Pi_1 + \ldots + \Pi_k \) be the total impact matrix, and consider the hypothesis of the existence of (at most) \( r \) cointegration vectors formulated as

\[
H: \quad \Pi = a\beta^t,
\]

where \( a \) and \( \beta \) are \( p \times r \) matrices.

The maximum likelihood estimation and likelihood ratio test of this model were investigated in Johansen (1988), and can be described as follows. First the model (1.1) is rewritten as

\[
\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + a\beta'X_{t-k} + \epsilon_t, \quad t = 1, \ldots, T,
\]

where \( \Gamma_i = -I + \Pi_i + \ldots + \Pi_i, \ i = 1, \ldots, k-1. \) In this model the parameters \( (\Gamma_1, \ldots, \Gamma_{k-1}, a, \beta, A) \) are variation independent and one can easily maximize with respect to the parameters \( \Gamma_1, \ldots, \Gamma_{k-1} \) by regressing \( \Delta X_t \) and \( X_{t-k} \) on the lagged differences, and obtain residuals \( R_{0t} \) and \( R_{kt} \).

Then we define

\[
S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it}R_{jt}, \quad i, j = 0, k,
\]

and find

\[
\hat{a}(\beta) = S_{0k}^{-1}(\beta' \sigma_{kk}' \beta)^{-1}.
\]

Finally \( \hat{\beta} \) is found as the eigenvectors corresponding to the \( r \) largest eigenvalues of the equation

\[
(1.4) \quad |\lambda S_{kk} - S_{0k}^{-1}S_{kk}'S_{0k}^{-1}| = 0.
\]

\(^1\)One can also rewrite the model as

\[
\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i^* \Delta X_{t-i} + \epsilon_t
\]

with \( \Gamma_i^* = \Gamma_i - \Pi = -\Pi_{i+1} - \ldots - \Pi_k \). It may be convenient to see the error correction term with one lag, but the analysis remains the same.
giving the maximized likelihood function
\[ L_{\text{max}}^{-2/T} = |S_{00}| \prod_{i=1}^{r} (1 - \hat{\lambda}_i). \]

Hence the likelihood ratio test for the hypothesis \( H \) of (at most) \( r \) cointegrating relations is given by
\[
(1.5) \quad -2 \ln(Q) = T \sum_{i=r+1}^{p} \ln(1 - \hat{\lambda}_i).
\]

The limiting distribution of this statistic is non standard and has been tabulated by simulation in Johansen (1988). The theory for the model which allows a constant term is a bit more involved, and is given in Johansen (1989).

It is the purpose of this paper to derive the power function of the above likelihood ratio test, applying the results of Phillips (1988).

The alternative we are interested in is clearly that there are one or more extra cointegrating relations than assumed under \( H \). If we investigate the power of the test under the assumption that there is in fact another cointegration relation with some non zero loadings, then it is not difficult to see, that the power tends to one. More interesting is it to consider local alternatives of the form
\[
(1.6) \quad H_\pi: \quad \Pi_T = a \beta' + a_1 \beta'_1 / T
\]
where \( a_1 \) and \( \beta_1 \) are \( p \times s \) matrices. Under the alternative \( H_\pi \) we are thus allowing \( s \) extra cointegration vectors to enter the model with small weight \( a_1 / T \). These extra linear components of \( X_t \) are "near integrated" in the terminology of Phillips (1988), and the basic mathematical results are directly taken from that paper and adopted to this slightly different framework of analysing the multivariate Gaussian distribution.
In the next section we describe the asymptotic properties of near
integrated series, and apply these to find an asymptotic expression for
the power function. In section 3 we investigate by simulation the power
function for the local alternative of one extra cointegration vector with
a small loading. The appendix contains the proofs of the mathematical
results.

2. Asymptotic properties of the process $X_t$ and the power function under
local alternatives.

In order to discuss the asymptotic properties we shall recall some of
the properties of the process $X_t$ as given in Theorem 3.1 (Grangers
representation theorem) Johansen (1989). We assume that rank($\alpha$) =
rank($\beta$) = $r$ and let $\alpha_\perp$ and $\beta_\perp$ denote $p \times (p-r)$ matrices orthogonal to $\alpha$ and
$\beta$ and of full rank. The condition of "balance", see Johansen (1988a),
expressed as a relation between the impact matrix $\Pi = \alpha \beta'$ and the mean
lag matrix $\Psi = \sum_{i=1}^{k} i \Pi_i$, is given by

$$\text{rank}(\alpha_\perp \Psi \beta_\perp) = p-r.$$  

Under this condition we have the moving average representation $\Delta X_t =
C(L) \varepsilon_t$ and the expression

$$C(1) = \beta_\perp (\alpha_\perp \Psi \beta_\perp)^{-1} \alpha_\perp = C,$$

say. From this it is seen that $\Delta X_t$ is stationary, $\beta'X_t$ is stationary and
$X_t$ is non stationary.

In order to separate the stationary part of the process from the near
integrated and integrated components under the model $H_\perp$ we introduce the
$r$-dimensional process $Y_t^{(T)} = (\beta'\beta)^{-1} \beta'X_t^{(T)}$ and the $(p-r)$-dimensional
process $Z_t^{(T)} = (\beta_\perp \beta_\perp)^{-1} \beta_\perp X_t^{(T)}$, such that
\( (2.3) \quad X_t^{(T)} = \beta y_t^{(T)} + \beta_1 z_t^{(T)} \).

**THEOREM 1.** Under the local alternatives

\[ H_T : \Pi_T = a_1^T + a_1^1 \beta_1^T , \]

and under the condition (2.1) and

\( (2.4) \quad \text{rank}\{ (a_1^1 \beta_1^1) \} = p - r \)

the process \( X_t^{(T)} \) converges weakly to the process \( X_t' \), and the process \( Z_t^{(T)} \) converges weakly to the \( (p-r) \)-dimensional process \( K_t' \), which satisfies the stochastic differential equation

\( (2.5) \quad - \int_0^t b' K_u du + K_t = B_t \).

Here \( B_t \) is a \( (p-r) \)-dimensional Brownian motion, and \( a = a_1^1 a_1^1 \) and \( b = \beta_1^1 \).

The \( (p-r) \times (p-r) \) matrix \( ab' \) has rank \( s \) and the solution to (2.5) is the Ornstein-Uhlenbeck process in \( p - r \) dimensions which can be expressed as

\( (2.6) \quad K_t = \int_0^t \exp(ab'(t-u))dB. \)

The proof of this result will be given in the appendix. We shall apply the result in the proof of the main result about the power function:
THEOREM 2. The asymptotic distribution of the likelihood ratio test
- \(2\ln Q\) for the hypothesis \(H: \Pi = \alpha \beta'\) is, under the local alternative

\[ H_T : \Pi = \alpha \beta' + \alpha_1 \beta_1' / T, \]
given by

\[(2.7) \quad \text{tr}\{\int (dK)K' (\int KK' du)^{-1} \int K(dK)\}', \]

where \(K\) is given by the \((p - r)\)-dimensional Ornstein-Uhlenbeck process \((2.6)\).

The proof is given in the appendix. If the \((p-r)\)-dimensional process
\(K_t\) has coordinates \(\{K_i(t), i = 1, \ldots, p-r\}\), then the notation \(\int (dK)K'\)
stands for a \((p-r)\times(p-r)\) matrix of stochastic integrals with elements

\[\int_0^1 dK_i K_j = \int_0^1 K_j dK_i.\]

The matrix \(\int KK' du\) has elements \(\int_0^1 K_i(u) K_j(u) du\).

This result is closely related to the result by Phillips (1988) p.1031 on the power function for a certain test statistic for the hypothesis \(\Pi = 0\) based on a normalization of a regression estimate of \(\Pi\). In this situation, where we want to test \(\Pi = 0\), the present test which exploits the Gaussian distribution can not be distinguished by its asymptotic properties for local alternatives from the general test given by Phillips.

It is seen that the asymptotic power function for local alternatives depends on the parameters only through \(a = a' \alpha_1\) and \(b = \beta' \beta_1\). Thus it depends on how the extra loadings (\(\alpha_1\)) and cointegrating relations (\(\beta_1\)) are related to the \(\alpha\) and \(\beta\) assumed under \(H\). A tabulation of the power function thus involves \(2(p-r)\times s\) parameters. It turns out, however, that fewer parameters will do. In order to see this, we shall exploit the invariance of the multivariate Brownian motion under rotation and the
invariance of the test statistic.

Let $O$ denote a $(p-r) \times (p-r)$ orthonormal matrix. Then clearly $B_t$ and $OB_t$ have the same distribution and hence multiplying (2.5) by $O$ we see that $OK_t$ solves the equation (2.5) for the Brownian motion $OB_t$ with the $p \times s$ matrices $Oa$ and $Ob$. The test statistic is independent of this transformation and thus the power function is the same for any alternatives $ab'$ and $a_1b_1'$ such that $Oab'O' = a_1b_1'$. We can simplify still further by choosing new coordinates using the orthonormal vectors

\[
e_1 = b(b'b)^{-1/2}
\]
\[
e_2 = (a - b(b'b)^{-1}b'a)(a'a - a'b(b'b)^{-1}b'a)^{-1/2}
\]

and finally $e_3$ such that $(e_1, e_2, e_3)$ are orthonormal. We define $K_{1t} = e_1K_t$ and $B_{1t} = e_1B_t$ for $i = 1, 2, 3$. Then the equation (2.5) is equivalent to

\[
(2.7) \quad - (b'b)^{-1/2}b'a(b'b)^{1/2} \int_0^t K_{1u} \, du + K_{1t} = B_{1t},
\]
\[
(2.8) \quad - (a'a - a'b(b'b)^{-1}b'a)^{1/2}(b'b)^{1/2} \int_0^t K_{1u} \, du + K_{2t} = B_{2t},
\]
\[
(2.9) \quad K_{3t} = B_{3t}.
\]

This shows that the power function only depends on $a_1$ and $b_1$ through the matrices $a'a$, $b'b$ and $a'b$. For $s = 1$ this result simplifies still further.

**COROLLARY.** Under the local alternative $II = a\beta' + a_1\beta_1' T$, where $a_1$ and $\beta_1$ are $p \times 1$ vectors, the power function for the likelihood ratio test for cointegration depends on $a_1$ and $\beta_1$ only through the quantities $f$ and $g$ given by

\[
(2.10) \quad f = a'b = \beta_1'C a_1 < 0,
\]
and

\[(2.11) \quad g^2 = a'ab'b - (a'b)^2 \]

\[= (a'\alpha_1'(a'\Lambda\alpha_1')^{-1}a'\alpha_1')(\beta_1'\Psi\alpha_1') - (\beta_1'\alpha_1')^2, \]

where \( C = \beta_1'(\alpha'\Psi\beta_1')^{-1}a_1' \), see (2.2).

Note that the expressions above for \( C \) as well as \( \alpha_1'(a'\Lambda\alpha_1')^{-1}a_1' \) do not depend on the particular choice for \( \alpha_1 \) and \( \beta_1 \). Note also that for \( p - r = 1 \) the power function depends only on \( f \), since (2.7) above describe the process \( K_t \).

The proof of the corollary is given in the appendix.

In order to interpret these results we consider the hypothesis \( H_T \) for the simple example we get from (1.3) by letting \( k = 1 \), that is we consider the system

\[ \Delta X^{(T)}_t = T^{-1}\alpha_1\beta_1'\xi^{(T)}_{t-1} + \epsilon_t. \]

Thus we have left out the short term dynamics and assumed that \( \alpha = \beta = 0 \).

We then want to investigate if we can find a cointegration vector \( \beta_1 \) in the data. In this case \( \Pi = 0 \), \( \alpha_1 = \beta_1 = I \) and \( \Psi = I \). Thus \( C = I \), \( a = \alpha_1 \) and \( b = \beta_1 \), such that \( f = a_1'\beta_1 \) and \( g^2 = a_1'\alpha_1\beta_1'\beta_1 - (a_1'\beta_1)^2 \).

We can express the results as follows: The power for finding a stationary relation \( \beta_1 \) with loadings \( T^{-1}\alpha_1 \) depends on the position of the vectors \( \alpha_1 \) and \( \beta_1 \) through the angle \( (f/g) \) between them and the area \( (g) \) spanned by them, as well as the dimension \( (p) \) of the space in which we are looking for them.
3. Numerical investigation of the power function under the alternative of one extra cointegration vector.

For \( s = 1 \) the equations (2.7), (2.8) and (2.9) are of the form

\[
\begin{align*}
(3.1) & \quad t \int_{0}^{t} K_1 u \text{d}u + K_1 t = B_1 t' \\
(3.2) & \quad t \int_{0}^{t} K_2 u \text{d}u + K_2 t = B_2 t' \\
(3.3) & \quad K_3 t = B_3 t'.
\end{align*}
\]

The distribution of the functional (2.7) is too complicated to find analytically but can be found by simulation. In order to simulate the processes we shall apply the discrete version of these equations

\[
\begin{align*}
(3.4) & \quad K_{1t} = (1 + f/T)K_{1t-1} + \varepsilon_{1t}' \\
(3.5) & \quad K_{2t} = K_{2t-1} + (g/T)K_{1t-1} + \varepsilon_{2t}' \\
(3.6) & \quad K_{3t} = K_{3t-1} + \varepsilon_{3t}'
\end{align*}
\]

\( t = 1, \ldots, T \) starting with \( K_0 = 0 \).

Note that the actual expression for the limit distribution is not so important. What is being applied is that the limit distribution exists and does not involve the parameters, but for \( f \) and \( g \). Thus in order to simulate the system we simply choose the simplest possible system compatible with the given \( f \) and \( g \).

We then easily solve the equations (3.4), (3.5) and (3.6) recursively and form the \( T \times (p-r) \) matrix \( M \) with elements \( M_{ti} = K_{it} \). Then we calculate \( \Delta M \) and \( M_{-1} \) i.e. the differences and lagged variables respectively and find the test statistic

\[
\text{Test} = \text{tr}\{\Delta M'M_{-1}(M_{-1}'M_{-1})^{-1}M_{-1}'AM\}.
\]
The number of observations $T$ has to be chosen so large that the approximation of the random walk to the Brownian motion is sufficiently good. We have chosen $T = 400$.

We find the results for stationary alternatives and $p-r = 1, 2$, and $3$ in Table I and Figures 1, 2 and 3 based on $2000 - 5000$ simulations and $T = 400$. It is seen that, not surprisingly, the power decreases as the dimension increases, i.e. if there are many dimensions to hide in, i.e. it is difficult to find the cointegrating vector, if it has a small loading. It was found that the non stationary alternatives (not shown) are readily picked up by the test with large power. The test appears unbiased as $f$ and $g$ move away from 0.

One can interpret the coefficient $1 + f/T = 1 + \alpha_1'\beta_1/T$ as the autoregressive parameter, see (3.4), in the stationary (or near integrated) relation we are trying to find. Hence we see from Table I, that if for instance we have $T = 100$ observations and expect an autoregressive coefficient of say $0.79$ then we have $f = 100 \times (0.79 - 1.00) = -21$. Now the power of finding such a process depends on the relation between the loadings and the cointegration vector. If $\alpha_1$ and $\beta_1$ are proportional, such that $g = 0$, then, if $p = 1$, we have a probability of $0.998$ of rejecting the hypothesis of non stationarity, and hence of finding a stationary relation. If, however, the system is of dimension $p = 3$, then the probability of rejecting the non stationarity hypothesis is only $0.346$. For a given angle that is for fixed $f/g$ it is seen from the tables that the larger the vectors the easier it is to find them.
4. Appendix.

We have here collected the proofs of Theorem 1, Theorem 2 and the Corollary. We first define the variance and covariance matrices conditional on the lagged differences by

\[
\Sigma_{00} = \text{Var}(\Delta X_t | \Delta X_{t-1}, \ldots, \Delta X_{t-k+1}),
\]

\[
\beta' \Sigma_{kk} \beta = \text{Var}(\beta' X_{t-k} | \Delta X_{t-1}, \ldots, \Delta X_{t-k+1}),
\]

\[
\Sigma_{0k} \beta = \text{Cov}(\Delta X_t, \beta' X_{t-k} | \Delta X_{t-1}, \ldots, \Delta X_{t-k+1}).
\]

It follows from the representation (1.3) that, since \( e_t \) is independent of the past values of the process \( X_t \), one can find the relations

\[
\Sigma_{00} = \alpha \beta' \Sigma_{kk} \beta' + \Lambda,
\]

and

\[
\Sigma_{0k} \beta = \alpha \beta' \Sigma_{kk} \beta,
\]

which can be solved for \( \alpha \)

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

One can apply these identities to prove

**Lemma 1.** The following relations hold

\[
\Lambda^{-1} - \Lambda^{-1} \alpha (\alpha' \Lambda^{-1} \alpha)^{-1} \alpha' \Lambda^{-1} = \alpha_\perp (\alpha'_\perp \Lambda_\perp)\alpha^{-1}_\perp = \\
\alpha_\perp (\alpha' \Sigma_{00} \alpha_\perp)^{-1} \alpha'_\perp = \Sigma_{00}^{-1} - \Sigma_{00}^{-1} \alpha (\alpha' \Sigma_{00}^{-1} \alpha)^{-1} \alpha' \Sigma_{00}^{-1}
\]

**Proof.** The second equality sign holds by the relation (4.1) and the two other are just the well known relation between a block matrix \( M \) and its inverse

\[
M_{22} - M_{21}(M_{11})^{-1}M_{12} = M_{22}^{-1},
\]
expressed in the coordinates given by $\alpha$ and $\alpha_\perp$.

We now choose $\alpha_\perp$, such that

\begin{equation}
\alpha_\perp' \Lambda \alpha_\perp = I.
\end{equation}

This can always be done for if $\tilde{\alpha}_\perp$ is any choice of p-r vectors orthogonal to $\alpha$, then we just take $\alpha_\perp = \tilde{\alpha}_\perp (\tilde{\alpha}_\perp' \Lambda \tilde{\alpha}_\perp)^{-1/2}$.

We shall further choose $\beta_\perp$ such that

\begin{equation}
\beta_\perp' \psi \beta_\perp = I.
\end{equation}

This is always possible since for any choice of $\tilde{\beta}_\perp$ of p-r vectors orthogonal to $\beta$ we just take $\beta_\perp = \tilde{\beta}_\perp (\tilde{\beta}_\perp' \psi \tilde{\beta}_\perp)^{-1}$ which satisfies (4.6).

These choices will simplify the calculations later, but the results about the power function are of course formulated such that they do not depend on this particular choice.

We now give the proof of Theorem 1. The result follows from the results in Phillips (1988), and we shall here only sketch the calculations involved in the proof.

We write the model (1.1) under $H_T$ in the form

\begin{equation}
- (a\beta' + \alpha_\perp \beta_\perp' T) X^{(T)}_t + \psi \Delta X^{(T)}_t + \Pi_2(L) \Delta^2 X^{(T)}_t = \epsilon_t.
\end{equation}

The basic idea is to go to the limit in this stochastic difference equation and derive either another stochastic difference equation or a limiting stochastic differential equation. It turns out that only $\Pi$ and $\psi$ will be relevant in this last approximation, such that near integrated processes can be approximated by AR(1) processes, see (3.4).

If we let $T \to \infty$, in (4.7) we get the equation

\begin{equation}
- a\beta' X_t + \psi \Delta X_t + \Pi_2(L) \Delta^2 X_t = \epsilon_t.
\end{equation}

which shows that $X^{(T)}_t$ converges weakly to $X_t$. If we multiply (4.7) by $\alpha_\perp'$ we get instead
\[ - T^{-1} \alpha \beta' X_t^{(T)} + \alpha' \Psi \Delta X_t^{(T)} + \alpha' \Pi_2(L) \Delta X_t^{(T)} = \alpha \epsilon_t. \]

Summing over \( t = 1, \ldots, [Tt] \), and dividing by \( T^{3/2} \) we obtain, apart from some initial values,

\[
\begin{align*}
(4.8) & \quad - a T^{-1} \sum_{t=1}^{[Tt]} \beta' X_t^{(T)}/T^{3/2} + a' \Psi X_t^{(T)}/T^{3/2} + \alpha' \Pi_2(L) \Delta X_t^{(T)}/T^{3/2} \\
& \quad = \sum_{t=1}^{[Tt]} a' \epsilon_t/T^{3/2}.
\end{align*}
\]

Now apply the decomposition (2.3) and use the result that \( \beta y(T)_{[Tt]} / T^{3/2} \rightarrow \beta' X_t \) since \( \beta' X_t \) is stationary. Hence we can replace \( X_t \) by \( \beta' Z_t \).

The right hand side converges to a Brownian motion with covariance \( \alpha' \beta' = I \), by the choice of \( \alpha', \) and the first term is \(-\alpha \beta' T^{-1} \sum_{t=1}^{[Tt]} Z_t^{(T)}/T^{3/2}\), i.e. a Riemann sum of the process \( Z_t^{(T)}/T^{3/2} \), which converges weakly to the integral \(-\alpha' \beta' \int_0^T du\), whereas the second term converges to \( a' \Psi \beta' K_t = K_t \) by the choice of \( \beta' \).

It is easily checked that the solution is given by (2.6). In order to see that the matrix \( \alpha' \beta' \) has rank \( s \) we have to apply the conditions (2.1) and (2.4) which guarantee that the process \( X_t \) is at most \( I(1) \) both under \( H \) and \( H_T \). Let \( b_1 = \beta' (\beta' \beta_1')^-1 \beta' \beta_1 \) and let \( b_2 \) be orthogonal to \( (\beta, b_1) \) such that \( (\beta, b_1, b_2) \) has full rank. Then \( b_1 = (b_1, b_2) \) and \( (\beta, b_1) = b_2 \). Similarly we define \( a_1 \) and \( a_2 \) from \( \alpha_1' \) and \( \alpha_1' \). The condition (2.1) can now be expressed as

\[
\text{rank } \begin{bmatrix} a_1' b_1 & a_1' b_2 \\ a_2' b_1 & a_2' b_2 \end{bmatrix} = p-r,
\]

whereas the condition (2.4) can be expressed as
These conditions clearly imply that
\[ A = a_1'b_1 - a_1'b_2(a_2'b_2)^{-1}a_2'b_1 \]
is of full rank s. This therefore also holds for \( a_1'b_1 \) and hence for \( a_1 \) and \( b_1 \). Now
\[
ba' = b_1'a_1a_1' = b_1'b_1a_1a_1' = \begin{bmatrix} b_1' \\ b_2' \end{bmatrix} \begin{bmatrix} b_1a_1a_1' & 0 \\ 0 & 0 \end{bmatrix}
\]
which has rank s.

Next we shall turn to the proof of Theorem 2. This follows closely the proof given in Johansen (1988) for the asymptotic distribution under \( H_0 \), and it will therefore not be given in full technical detail. The likelihood ratio test is calculated from (1.5) solving the eigenvalue problem (1.4). The basic idea in the proof is to study what happens as \( T \to \infty \) in the equation (1.4) under different normalizations, and then apply the continuity of the ordered eigenvalues as a function of the coefficients.

First multiply the matrix equation (1.4) by \((\beta, T^{-1/2} \beta')\) and its transposed and apply the stationarity of \( \beta'X_t \) to see that \( \beta'S_{kk}\beta \), \( \beta'S_{kk}\beta' \) are \( O_p(1) \). By Theorem 1, \( \beta'S_{kk}\beta'/T \) converges weakly to \( \beta'kk'd\beta' \), \( \phi \), say, and this shows that in the limit \( \lambda \) has to satisfy the equation
\[
|\lambda \begin{bmatrix} \beta'S_{kk}\beta & 0 \\ 0 & \phi \end{bmatrix} - \begin{bmatrix} \beta'S_{kk} & 0 \\ 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \beta'S_{kk} & 0 \\ 0 & 0 \end{bmatrix} \phi | = 0
\]
or
\[
|\lambda \beta'S_{kk}\beta - \beta'S_{kk}\beta' | |\lambda\phi| = 0,
\]
which has \( r \) positive solutions and \( p-r \) null solutions. Thus the \( r \) largest eigenvalues of (1.4) converge to those of
\[
|\lambda \beta'S_{kk}\beta - \beta'S_{kk}\beta'| = 0.
\]
and the p-r smallest tend to zero. The r largest eigenvalues correspond to the stationary components $\beta'X_t$, whereas the p-r smallest correspond to the near integrated as well as the non stationary components.

Now normalize $\hat{\lambda}_i$ by T and define $\hat{\rho}_i = T\hat{\lambda}_i$. We multiply (1.4) from both sides by $(\beta, \beta_\perp)$ and let $T \to \infty$, then we get that the limiting value of $\hat{\rho}$ will satisfy the equation

$$\begin{vmatrix} 0 & 0 \\ 0 & \rho \phi \end{vmatrix} - \begin{bmatrix} \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta \\ F' \end{bmatrix} ' \begin{bmatrix} \Sigma_{00}^{-1} \beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta \\ F \end{bmatrix} \begin{bmatrix} 0 \\ \rho \phi \end{bmatrix} = 0,$$

where $F'$ is the weak limit of $\beta_\perp S_{k0}$. This can be written

$$|\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta| |\rho \phi - F'(\Sigma_{00}^{-1} - \Sigma_{k0}^{'0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)(\beta' \Sigma_{k0} \Sigma_{00}^{-1} \Sigma_{0k} \beta)^{-1} \beta' \Sigma_{k0} \Sigma_{00}^{-1}| F' = 0.$$

Since $\alpha = \Sigma_{k0}^{'0} \beta(\beta' \Sigma_{k0} \beta)^{-1}$ one can express the matrix between $F'$ and $F$ as

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1} a(\alpha' \Sigma_{00}^{-1} \alpha)^{-1} \alpha' \Sigma_{00}^{-1},$$

which by Lemma 1 equals

$$\alpha_\perp (\alpha_\perp' \lambda_\perp)^{-1} \alpha_\perp = \alpha_\perp \alpha_\perp'$$

by the definition of $\alpha_\perp$, see (4.5).

Thus $\rho$ is the solution to the equation

(4.9) \hspace{1cm} |\rho \phi - F' \alpha_\perp \alpha_\perp'= 0.$$

The equation (1.3) gives the following relation between the residuals and the residual sum of squares after the preliminary regressions.

$$S_{k0} - S_{kk}(\beta \alpha' + \beta_\perp \alpha_\perp T^{-1}) = T^{-1} \Sigma T_{kT} \epsilon_t.$$

It follows, that $\beta_\perp S_{k0} \alpha_\perp = F' \alpha_\perp$ has the same limit as

$$\beta_\perp S_{kk} \beta_\perp' \alpha_\perp T^{-1} + T^{-1} \Sigma T_{kT} \epsilon_t \alpha_\perp,$$

which by Theorem 1 converges to

$$\beta_\perp \beta_\perp' (JKK' du ba' + fK(db')'),$$

since $V(\epsilon_t) = \alpha_\perp' \lambda \alpha_\perp = I$, by the choice of $\alpha_\perp$.

Now apply (2.5) to obtain
\[ \beta_1 ' \beta_1 (\int_k (b' + jk'dub')\}) = \]
\[ \beta_1 ' \beta_1 \int_k (dK)' . \]

Thus inserting this result in (4.9) we find that the limiting value of \( \hat{\rho} \)
is a solution to the equation
\[ |\rho \int_k K' du - \int_k (dK)' \int_k (dK)' | = 0, \]
and hence the result (4.9) holds, since
\[ -2 \ln Q = - \sum_{i=r+1}^P \ln(1- \lambda_i) \simeq \sum_{i=r+1}^P \lambda_i = \sum_{i=r+1}^P \rho_i \approx \tr(\int_k (dK)'(\int_k K' du)^{-1} \int_k (dK)'). \]

Finally we shall give the proof of the Corollary. This result follows from the equations (2.7), (2.8) and (2.9), and the expressions for the coefficients \( f \) and \( g \) are easily deduced by applying the expression for \( C = \beta_1 (a_1 \psi_1)^{-1} \alpha_1 ' \) together with the identities \( \alpha_1 \Lambda \alpha_1 = I \) and \( \alpha_1 \psi_1 \psi_1 = I \). We find
\[ f = b'a = \beta_1 ' \beta_1 \alpha_1 \alpha_1 = \beta_1 ' \beta_1 (\alpha_1 \psi_1^\prime)_{-1} \alpha_1 ' \alpha_1 = \beta_1 ' \alpha_1 \]
which is non zero by the result of Theorem 1. Since we are interested in stationary alternatives only it follows from (2.7) that \( a'b < 0. \)

Furthermore
\[ a'a = a_1 ' a_1 a_1 ' a_1 = (a_1 ' a_1 (a_1 \Lambda a_1)^{-1} a_1 ' a_1) \]
and
\[ b'b = \beta_1 ' \beta_1 \beta_1 ' \beta_1 = \beta_1 ' \beta_1 (\alpha_1 \psi_1)_{-1} (a_1 ' \Lambda a_1) (\beta_1 ' \psi_1 ' a_1)_{-1} \beta_1 ' \beta_1 = \beta_1 ' \Lambda \beta_1 \]

5. Acknowledgement.

The author thanks Neil Ericsson for some interesting discussions on the topic at an early stage of the work.
6. References


Table I
The asymptotic power of the likelihood ratio test at 5% for \( r \) cointegration vectors among \( p \) variables under the local alternative of one extra cointegration vector. The quantities \( f \) and \( g \) are defined in (2.10) and (2.11). The number of simulations are 2000–5000 and \( T = 400 \).

<table>
<thead>
<tr>
<th>( p - r = 1 )</th>
<th>( f )</th>
<th>(-3)</th>
<th>(-6)</th>
<th>(-9)</th>
<th>(-12)</th>
<th>(-15)</th>
<th>(-18)</th>
<th>(-21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>power</td>
<td>.052</td>
<td>.141</td>
<td>.350</td>
<td>.620</td>
<td>.820</td>
<td>.945</td>
<td>.987</td>
<td>.998</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( g )</th>
<th>( p - r = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.060</td>
</tr>
<tr>
<td>6</td>
<td>.615</td>
</tr>
<tr>
<td>12</td>
<td>.944</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( g )</th>
<th>( p - r = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.054</td>
</tr>
<tr>
<td>6</td>
<td>.510</td>
</tr>
<tr>
<td>12</td>
<td>.834</td>
</tr>
</tbody>
</table>
Legend to Figures.

Fig. 1.

The power function of the likelihood ratio test for cointegration, as a function of $f$, see (2.10), for $g = 0$, see (2.11), and different values of the dimension $p - r$.

Fig. 2.

The power function for the likelihood ratio test for cointegration as a function of $f$ and $g$, see (2.10) and (2.11) and $p-r = 2$.

Fig. 3

The power function for the likelihood ratio test for cointegration as a function of $f$ and $g$, see (2.10) and (2.11) and $p-r = 3$. 
Fig. 1

Power

1.00

0.05

0

-3

-6

-9

-12

-15

-18

-21

f

1

2

3

p-r
COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE +45 1353133.

No. 1  Jacobsen, Martin: Discrete Exponential Families: Deciding when the Maximum Likelihood Estimator Exists and Is Unique.

No. 2  Johansen, Søren and Jusélius, Katarina: Hypothesis Testing for Cointegration Vectors - with an Application to the Demand for Money in Denmark and Finland.


No. 4  Bertelsen, Aksel: On Non-Null Distributions Connected with Testing that a Real Normal Distribution Is Complex.

No. 5  Tjur, Tue: Statistical Tables for Personal Computer Users.

No. 6  Tjur, Tue: A New Upper Bound for the Efficiency of a Block Design.


No. 8  Brøns, Hans and Jensen, Søren Tolver: Maximum Likelihood Estimation in the Negative Binomial Distribution.

PREPRINTS 1989

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF
MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK,
TELEPHONE +45 1 35 31 33.

No. 1 Bertelsen, Aksel: Asymptotic Expansion of a Complex Hypergeometric Function.

No. 2 Davidsen, Michael and Jacobsen, Martin: Weak Convergence of Two sided
Stochastic Integrals, with an Application to Models for Left Truncated Survival Data.

No. 3 Johansen, Søren: Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.

No. 4 Johansen, Søren and Juselius, Katarina: The Full Information Maximum Likelihood Procedure for Inference on Cointegration – with Applications.

No. 5 Thompson, Steven K.: Adaptive Cluster Sampling.

No. 6 Thompson, Steven K.: Adaptive Cluster Sampling: Designs with Primary and Secondary Units.

No. 7 Thompson, Steven K.: Stratified Adaptive Cluster Sampling.

No. 8 Johansen, Søren: The Power Function of the Likelihood Ratio Test for Cointegration.