Modelling of cointegration in the vector autoregressive model

Søren Johansen*

Department of Economics, European University Institute, Via dei Roccettini 9, 50016, San Domenico di Fiesole, Florence, Italy

A survey is given of some results obtained for the cointegrated VAR. The Granger representation theorem is discussed and the notions of cointegration and common trends are defined. The statistical model for cointegrated I(1) variables is defined, and it is shown how hypotheses on the cointegrating relations can be estimated under suitable identification conditions. The asymptotic theory is briefly mentioned and a few economic applications of the cointegration model are indicated. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

The notion of cointegration has become one of the more important concepts in time series econometrics since the papers by Granger (1983) and Engle and Granger (1987). The topic of cointegration has found widespread applications in the analysis of economic data as published in the econometric literature. The special issues of the Oxford Bulletin of Economics and Statistics, 54 (3) (1992) and Journal of Policy Modelling, 14 (3,4) (1992) contain many papers where the method is applied and extended. The book of readings by Engle and Granger (1991) contains a collection of papers that have been important for the development of the topic. Many text books contain the basic aspects of cointegration: see for instance Reinsel (1991), Lütkepohl (1993), Hamilton (1994), or Cuthbertson et al. (1992). The books by Banerjee et al. (1993), Johansen (1996) and Hansen and Johansen (1998), are systematic treatments of the topic of cointegration.

The present paper gives a brief overview of the theory of cointegration in the vector autoregressive model as developed in Johansen and Juselius (1990, 1992, 1994). The theory is developed as a careful study of the mathematical structure of...
the error correction model and its solution, followed by an analysis of the Gaussian likelihood function that allows one to derive estimators and test statistics. The asymptotic analysis is only briefly indicated, and a few extensions of the theory are pointed out.

Economic insight is used in formulating the problem of interest, and therefore in the choice of variables, as well as in the discussion of which economic relations we expect to find. The statistical model is then used as a description of the non-stationary statistical variation of the data. The cointegrating relations are used as a tool for discussing the existence of long-run economic relations and the various hypotheses are then tested in view of the statistical variation of the data. The interpretation of the cointegrating relations require a thorough understanding of the underlying economic problem, and the purpose of the statistical modelling is to provide a platform on which to discuss the economic questions of interest. We sometimes find that economic theory is rejected by the data. This can be because the theory is not developed enough but of course also because the choice of variables for testing the theory may be inadequate.

1.1. Basic definitions

In this section we give the basic definitions and a discussion of the concepts, and we start by defining the class of stationary and non-stationary processes we want to investigate: Let $\mathbf{e}_i$ denote a doubly infinite sequence of $n$-dimensional i.i.d. stochastic variables with a mean of zero and finite variance. From these we construct a linear process $X_t = \sum_{i=0}^{\infty} C_i e_{t-i}$ where the coefficient matrices $C_i$ decrease exponentially fast, so that the series converges almost surely. This implies that the power series

$$C(z) = \sum_{i=0}^{\infty} C_i z^i$$

is convergent for $|z| < 1 + \delta$, for some $\delta > 0$. For the analysis of the likelihood function we need a further condition that the $\mathbf{e}$ values are Gaussian. For the asymptotic analysis, however, this condition is not needed, and we only need conditions under which the Central Limit Theorem holds, and for which we get convergence to certain stochastic integrals; see Section 4. We do not go into details with the asymptotics and the probability assumptions in this presentation.

In the following we define the concept of $I(0)$ and $I(1)$. The purpose is to define classes of non-stationary processes, $I(1)$, which become stationary after differencing, and a class of stationary processes, $I(0)$, which become non-stationary when summed, thus mimicking the relation between a random walk and its increments.

**Definition 1** A linear $n$-dimensional process $X_t = \sum_{i=0}^{\infty} C_i e_{t-i}$ is called integrated of order zero, $I(0)$ if $\sum_{i=0}^{\infty} C_i \neq 0$.

Using the concept of $I(0)$ we now define the main concept for the analysis of cointegration, namely integration of order 1, $I(1)$.
Definition 2 An n-dimensional stochastic process $X_t$ is called integrated of order 1, $I(1)$, if $\Delta X_t$ is $I(0)$.

The simplest example of an $I(1)$ process is a random walk, but any process of the form

$$X_t = C \sum_{i=1}^{\ell} \epsilon_i + \sum_{i=0}^{m} C_i \epsilon_{t-i}$$

is also an $I(1)$ process, at least if $C \neq 0$. Note that an $I(1)$ process is non-stationary, but that the non-stationarity can be removed by differencing.

We next give the definition of cointegration:

Definition 3 If $X_t$ is integrated of order 1 but some linear combination, $\beta' X_t$, $\beta \neq 0$ can be made stationary by a suitable choice of $\beta' X_t$, $X_t$ is called cointegrated and $\beta$ is the cointegrating vector. The number of linearly independent cointegrating vectors is called the cointegrating rank, and the space spanned by the cointegrating vectors is the cointegration space.

The idea behind cointegration is that sometimes the lack of stationarity of a multidimensional process is caused by common stochastic trends, which can be eliminated by taking suitable linear combinations of the process, thereby making the linear combination stationary.

In economics and other applications of statistics the autoregressive processes have long been applied to describe stationary phenomena and the idea of explaining the process by its past values has been very useful for prediction. If, however, we want to find relations between simultaneous values of the variables in order to understand the interactions of the economy one would get a lot more information by relating the value of a variable to the value of other variables at the same time point rather than relating it to its own past. One can say that if we want to discuss relations between variables, then one should take combinations of simultaneous values and if we want to discuss dynamic development of the variables we should investigate the dependence on the past.

The reason that cointegration has been so popular in econometrics is that classical macro-economic models are often formulated as simultaneous linear relations between variables following the Cowles Commission tradition. The theory of such equations was developed for stationary processes despite the fact that many (or even most) economic variables are non-stationary. If we think of the classical economic relations as long-run relations one can easily imagine that such relations can be stationary even if the variables themselves are non-stationary. Cointegration is the mathematical formulation of this phenomenon, and we shall treat it in the framework of the vector autoregressive model in the next section.

2. Granger’s representation theorem

The main content of this section is how to express the stochastic properties of
the solution of the autoregressive equations under various assumptions on the parameters. The equations are given in the reduced error correction form

\[ \Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{\infty} \Gamma_i \Delta X_{t-i} + \Phi D_t + \epsilon_t, \quad t = 1, \ldots, T, \]

where \( D_t \) are deterministic dummies and the \( \epsilon_t \) are i.i.d. \( N(0, \Omega) \). The equations determine the process \( X_t \) as a function of initial values \( X_0, \ldots, X_{-k} \), the \( \epsilon_t \) and the dummies \( D_t \) which can contain a constant, a linear term or seasonal dummies.

We let \( A(z) \) denote the \( n \times n \) matrix polynomial derived from Eq. 1

\[ A(z) = (1 - z)I - \Pi z - \sum_{i=1}^{\infty} \Gamma_i z^i (1 - z), \]

and let \( |A(z)| \) denote the determinant, and \( \text{adj}A(z) \) the adjoint matrix, so that

\[ A^{-1}(z) = \frac{\text{adj}(A(z))}{|A(z)|}. \]

The main assumption on the polynomial \( A(z) \) is

**Assumption 1** The polynomial \( A(z) \) satisfies the condition.

\[ |A(z)| = 0, \quad \text{implies either } |z| > 1 \text{ or } z = 1. \]

Thus, the coefficients should be chosen so that the roots of \( |A(z)| = 0 \) are either unit roots or stationarity roots.

For any \( n \times m \) matrix \( a \) of full rank \( m < n \), we denote by \( a_{\perp} \) an \( n \times (n - m) \) matrix of rank \( n - m \) such that \( da_{\perp} = 0 \). For notational convenience we let \( a_{\perp} = I \) if \( a = 0 \), and if \( a = I \) we define \( a_{\perp} = 0 \).

First, we give the classical result about the representation of stationary solutions.

**Theorem 4** If \( X_t \) is given by Eq. 1 and if Assumption 1 holds, then \( X_{t-I}(X_t) \) can be given an initial distribution such that it becomes \( I(0) \) if and only if \( A(1) = -\Pi \) has full rank, that is, \( |A(z)| \) has no unit roots. In this case \( X_t \) can be given the representation

\[ X_t = C_i \epsilon_{t-i} + C_i \Phi D_{t-i} \]

where the coefficients are given by

\[ C(z) = \sum_{i=0}^{\infty} C_i z^i = A(z)^{-1}, \quad |z| < 1 + \delta \text{ for some } \delta > 0. \]

This result shows that if \( |A(z)| \) has all roots outside the unit disk then the process generated by Eq. 1 is stationary or rather can be made stationary by a
suitable choice of the initial distribution. Thus, we have to allow other roots of |A(z)| for X to be non-stationary.

If unit roots are allowed we can prove Granger’s representation theorem.

**Theorem 5** If $X_t$ is given by Eq. (1) and if Assumption 1 holds, then $X_t$ is I(1) if and only if

$$\Pi = \alpha \beta',$$

where $\alpha$, $\beta$ ($n \times r$) are of full rank $r < n$, and

$$\alpha'_L \left( I - \sum_{i=1}^{k} \Gamma_i \right) \beta_L \text{ has full rank.}$$

(3)

In this case, $\Delta X_t = E(\Delta X_t)$ and $\beta'X_t - E(\beta'X_t)$ can be given initial distributions such that they become I(0), and the process $X_t$ has the representation:

$$X_t = C \sum_{i=1}^{t} (e_i + D_i) + C(L)(e_t + D_t) + A, \quad t = 1, \ldots, T,$$

(4)

where

$$C = \beta_L \left( \alpha'_L \left( I - \sum_{i=1}^{k} \Gamma_i \right) \beta_L \right)^{-1} \alpha'_L,$$

(5)

and $A$ depends on initial conditions such that $\beta' A = 0$.

Thus, the cointegrating vectors are $\beta$ and the common trends are $\alpha'_L \sum_{i=1}^{t} e_i$.

The representation theorem 5 is the common trends representation of the solution of the autoregressive model and shows that the non-stationarity in the variables is created by the cumulated unanticipated shocks in the process, but not all these shocks appear. They are multiplied by the matrix $\alpha'_L$ which shows that only $n - r$ random walks give rise to the non-stationarity. Since the matrix $C$ contains the factor $\beta_L$, we find $\beta' C = 0$ such that the linear combinations $\beta' X_t$ are not influenced by the random walks and become stationary.

The processes generated by Eq. (1) contain deterministic terms. It follows from Granger’s representation theorem that a constant term in the equations will generate a linear term in the process, but only in the non-stationary part of the process, that is, the process $\beta' X_t$ has no trend.

It is an important property that cointegration is invariant to the extension of the information set, that is, if more variables are included in the analysis we will still find the cointegrating vectors expanded by a zero for the new variables, but the common trends change character completely since what is unanticipated for the small system may not be unanticipated for the large system.

It is obvious that what is sometimes called the permanent shocks are the shocks $\alpha'_L e_t$ since they cumulate in the system. We propose to call the shocks $\alpha' \Omega^{-1} e_t$ the
transitory shocks. The reason will be apparent in the discussion of the asymptotic
distribution of \( \beta \), here we just note that the definition implies that the transitory
shocks are independent of the permanent shocks.

3. The statistical model for \( I(1) \) processes

If model (1) describes an \( I(1) \) process having cointegration we should restrict the
parameters as given by conditions (2), (3), and Assumption 1. Assumption 1, which
says that the roots are outside the unit disk or at 1, is very difficult to handle
analytically. Fortunately it rarely turns out that the roots are inside the unit disk,
and if they are, it is more important to know where they are than to force them to
the boundary of the unit disk. Hence we do not restrict the parameters in the
model by Assumption 1, but check that it is satisfied by the estimates. Condition (3)
is easily satisfied, since matrices with full rank are dense in the space of all
matrices, thus even without the restriction that \( \alpha' (I - \sum_{i=1}^{r} \Gamma_i) \beta \) has full rank
the estimator derived has full rank with probability one. Thus only condition (2),
\( \Pi = \alpha \beta' \) is included in the formulation of the model.

Definition 6 The reduced form error correction model \( H_r \) is described by the equations

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

where \( \alpha \) and \( \beta \) are \( n \times r \), and \( \epsilon_t, \ldots, \epsilon_T \) are independent Gaussian \( N(0, \Omega) \), and the
variables \( D_t \) are deterministic terms. The parameters \( \alpha, \beta, \Gamma_1, \ldots, \Gamma_r, \Phi, \Omega \) are freely varying.

We allow here the deterministic dummies \( D_t \) to enter the model, and assume
that the errors are Gaussian in order to be able to work with a likelihood function.
Note that in model \( H_r \) the parameters \( \alpha \) and \( \beta \) are not identified, since
\( \Pi = \alpha \beta' = \alpha \xi^{-1} (\beta \xi')' \) for any \( r \times r \) matrix \( \xi \) of full rank, but that one can
estimate the spaces spanned by \( \alpha \) and \( \beta \), respectively, and the parameters in \( \beta \) can
be estimated if they are identified or normalised suitably.

Thus, cointegration analysis is formulated as the problem of making inference on
the cointegration space, \( \text{sp}(\beta) \), and the adjustment space, \( \text{sp}(\alpha) \). If we want to
estimate individual coefficients it is necessary to normalise \( \beta \) or impose restrictions
so that the parameters become identified.

The above allows one to formulate a nested sequence of hypotheses

\[
H_0 \subset \cdots \subset H_r \subset \cdots \subset H_n,
\]

and the test of \( H_r \) in \( H_n \), is then the test that there are (at most) \( r \) cointegrating
relations. Thus, \( H_0 \) is just a vector autoregressive model for \( X_t \) in differences and
\( H_n \) the unrestricted autoregressive model for \( X_t \) in levels, and the models in
between, $H_1, \ldots, H_{n-1}$, give the possibility to exploit the information in the reduced rank matrix $P$, and contain information about the long-run relations in the economy.

Thus, instead of analysing non-stationary processes by differencing them to obtain stationarity and then analyse the differences by an autoregressive model, we choose to leave the variables in levels and draw inference from the cointegrating relations.

Note that the model we get by fitting a vector autoregressive model to the differences is just $H_0$, the adequacy of which can be tested if we start with the general model $H_n$, by testing $H_0$ in $H_n$.

3.1. Hypotheses on cointegrating relations

Once the cointegrating rank has been determined we can test hypotheses about the coefficients $\alpha$ and $\beta$, and we next give examples of such hypotheses.

In order to make the discussion of hypotheses more concrete we consider the example of five series: the log consumer price index in Australia and the US, $p_1$ and $p_2$, and the log exchange rate, $\text{exch}$, as well as the bond rate in both countries $i_1$ and $i_2$. The data is analysed in Johansen (1996) from the point of view of a cointegration analysis.

The hypothesis that only relative prices enter the cointegrating relations, can be expressed as the hypothesis that the coefficients to $p_1$ and $p_2$ sum to zero, or as the restriction $(1, 1, 0, 0, 0)\beta = 0$. This is the same restriction on all cointegrating relations which can also be expressed as a direct parameterization

$$ \beta = H\varphi, $$

where $H = (1, 1, 0, 0, 0)'$ is known and $\varphi (4 \times r)$ is unknown. This hypothesis on $\beta$ does not depend on $\beta$ being identified uniquely, since it is the same set of restrictions on all the relations. If $\beta$ satisfies Eq. (7) then so does $\beta\xi$ for any matrix $\xi (r \times r)$. Hence, Eq. (7) is a testable hypothesis on the cointegrating space, despite the fact that $\beta$ is not identified.

The hypothesis that some cointegrating vectors are known, like $(1, -1, -1, 0, 0)$ corresponding to $\text{PPP}$, or $(0, 0, 0, 1, -1)$, corresponding to $\text{UIP}$, can be formulated as

$$ \beta = (b, \psi), $$

where $b (p \times r_1)$ is known and $\psi (p \times r_2)$ is unknown, $r_1 + r_2 = r$. In particular it means that the test that an individual variable is stationary can be expressed in the form Eq. (8) for $b$ equal to a unit vector. Thus the stationarity of a single component of $X_t$ is a special case of cointegration.

A more general linear hypothesis can, for $r = 2$ say, be formulated as

$$ \beta = (H_1\varphi_1, H_2\varphi_2), $$
where $H(n \times s)$ are known and $\varphi(s \times r)$ are unknown and $(r_1 + r_2 = r$; see Johansen and Juselius (1994).

An example of Eq. (9) is given by the hypothesis that $p_1$, $p_2$, and $e_{12}$ cointegrate and that the interest rates cointegrate. In this case we are looking for two relations of the form $(a, b, c, 0, 0)$ and $(0, 0, 0, d, e)$, which clearly form a set of uniquely identified equations even though they also need a careful economic interpretation. The hypothesis has the form Eq. (9) with

$$H_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$ 

Thus we are, in the econometric language, testing for the over-identifying restrictions that there is a cointegrating relation between the variables that has two zeros as coefficients to the interest rates and another one with zeros as coefficients for the prices and exchange rates. It is a common econometric formulation that one wants to identify linear relations of econometric relevance by linear restrictions on the coefficients, in particular zero restrictions.

Thus, linear restrictions are formulated on individual relations in the hope that they are sufficiently distinct so that identification is in fact possible.

3.2. Estimation of cointegrating relations and calculation of test statistics

This section contains a brief description of likelihood analysis of the cointegration model and then discusses how the estimation problem of the various hypotheses from Section 3.1 can be solved by analysing the Gaussian likelihood function.

3.2.1. Unrestricted maximum likelihood estimation

Model (6) gives rise to a reduced rank regression and the solution is available as an eigenvalue problem. It was solved by Anderson (1951) in the regression context and runs as follows:

First we eliminate the parameters $\Gamma_1, \ldots, \Gamma_k, \Phi$ by regressing $\Delta X_t$ and $X_{t-1}$ on $\Delta X_{t-1}, \ldots, \Delta X_{t-k}, D_t$. The residuals are $R_{0t}$ and $R_{1t}$, respectively. Next form the sums of squares and products

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R_{jt}', \quad i, j = 0, 1.$$ 

The likelihood function, maximized with respect to the parameters $\Gamma_1, \ldots, \Gamma_k, \Phi$, and $\Omega$ is given by

$$L_{max}^{2/T}(\beta) = (2\pi e)^{n/2}|S_{00}|^{1/2} \frac{|\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01}) \beta|}{|\beta'S_{11}\beta|}. $$
This is minimized with respect to $\beta$ by solving the eigenvalue problem

$$[\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}] = 0$$

(10)

The solution of this equation gives eigenvalues $1 > \lambda_1 > \ldots > \lambda_n > 0$ and eigenvectors $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_n)$, which satisfy

$$\lambda_i S_{11} \hat{v}_i = S_{10}S_{00}^{-1}S_{01} \hat{v}_i, \quad i = 1, \ldots, n,$$

and

$$\hat{V}'S_{11}\hat{V} = I.$$

A maximum likelihood estimator for $\beta$ is then given by

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r).$$

(11)

An estimator for $\alpha$ is then

$$\hat{\alpha} = S_{00}^{-1}\hat{\beta},$$

and the maximized likelihood function is given by

$$L_{\text{max}}^{-2/T} = (2\pi e)^n |S_{00}| \prod_{i=1}^{r} (1 - \lambda_i),$$

(12)

see Johansen and Juselius (1990) for details and applications. One can interpret $\lambda_i$ as a squared canonical correlation between $\Delta X_r$ and $X_{t-1}$ conditional on $\Delta X_{t-1}, \ldots, \Delta X_{t-r}$. Thus, the estimate of the ‘most stable’ relations between the levels are those that correlate most with the stationary process $\Delta X_r$ corrected for lagged differences and deterministic terms.

Since only $\text{sp}(\hat{\beta})$ is identifiable without further restrictions, one really estimates the cointegrating space as the space spanned by the first $r$ eigenvectors. This is seen by the fact that if $\hat{\beta}$ is given by Eq. (11) then $\hat{\beta}\xi$ also maximises the likelihood function for any choice of $\xi(r \times r)$ of full rank. The identification of $\beta$ as eigenvectors is convenient from a mathematical and numerical point of view but not necessarily from an economic point of view.

This solution provides the answer to estimation of all the models $H_r, r = 0, \ldots, n$. By comparing the likelihoods Eq. (12) one can test $H_r$ in $H_n$, i.e. test for $r$ cointegrating relations, by the likelihood ratio statistic

$$-2\ln Q(H_r | H_n) = -T \sum_{i=r+1}^{n} \ln(1 - \lambda_i).$$

(13)

The estimator Eq. (11) is an estimator of all cointegrating relations and it is sometimes convenient to normalize (or identify) the vectors by choosing a specific coordinate system in which to express the variables in order to facilitate the
interpretation and in order to be able to give an estimate of the variability of the coefficients. If \( c \) is any \( n \times r \) matrix, such that \( \beta'c \) has full rank, one can normalize \( \beta \) as

\[
\beta_c = \beta (c\beta)^{-1},
\]

which satisfies \( c\beta_c = I \) provided that \( |c\beta| \neq 0 \). A particular example is given by \( c' = (I,0) \) and \( \beta' = (\beta_1, \beta_2) \) where \( \beta_1 \) is \( r \times r \) of full rank, in which case \( \beta'c = \beta_1 \) and \( \beta'_c = (I, \beta_1^{-1}\beta_2) \) which corresponds to solving the cointegrating relations for the first \( r \) variables. The maximum likelihood estimator of \( \beta_c \) is then

\[
\hat{\beta}_c = \hat{\beta} (c\hat{\beta})^{-1}.
\]

This then gives the normalisation or just identification of \( \beta \) that allows one to give an estimate of the variability of the estimator of the individual coefficient of \( \beta \).

### 3.3. Estimation of \( \beta \) under restrictions

If one wants to estimate \( \beta \) under restrictions this can sometimes be done by the same analysis. Consider the hypothesis (7) where, \( \beta = H\varphi \). In this case

\[
\alpha \beta'X_t = \alpha \varphi'H'X_t,
\]

which shows that the cointegrating relations are found by reduced rank regression of \( \Delta X_t \) on \( H'X_{t-1} \) corrected for the lagged differences and \( D_t \), that is, by solving the eigenvalue problem

\[
|\lambda H'S_{11} - H'S_{10}S_{00}^{-1}S_{01}H| = 0.
\]  

Under hypothesis (8) there are some known cointegration relations and in this case \( \alpha \beta'X_t = \alpha_1 b'X_t + \alpha_2 \varphi'X_t \), which shows that the coefficient \( \alpha_1 \) to the observable \( b'X_{t-1} \) can be eliminated together with the parameters \( (\Gamma_1, \ldots, \Gamma_k, \Phi) \), so that the eigenvalue problem that has to be solved is

\[
|\lambda S_{11,b} - S_{10,b}S_{00,b}^{-1}S_{01,b}| = 0,
\]

where

\[
S_{ij,b} = S_{ij} - S_{i1,b}(b'S_{11,b})^{-1}b'S_{1,j}, \quad i,j = 0,1.
\]

The maximal value of the likelihood function is given by expressions similar to Eq. (12) and the test of hypotheses (7) and (8) then consists of comparing the \( r \) largest eigenvalues under the various restrictions, since the factor \( (2\pi)^r |S_{00}| \) cancels.

The hypothesis (9) is slightly more complicated, but can be solved by a switching algorithm, where each step involves an eigenvalue problem (see Johansen and Juselius, 1994).
Thus, it is seen that a number of interesting hypotheses can be solved provided one has an eigenvalue routine and the algorithm has been implemented in many statistical packages.

4. Asymptotic theory

This section contains a description of the asymptotic theory of test statistics and estimators, as well as a discussion of how the results can be applied to conduct inference about the cointegrating rank and the cointegrating vectors.

4.1. Test for cointegrating rank

The reason that inference for non-stationary processes is interesting and widely studied, is that it is non-standard, in the sense that estimators are not asymptotically Gaussian and test statistics are not in general asymptotically $\chi^2$. This was systematically explored by Dickey and Fuller (see Fuller, 1976) in testing for unit roots in univariate processes.

As an example consider the simple model of an autoregressive process of order one

\[ X_t = \rho X_{t-1} + \epsilon_t, \]

where $\epsilon_t$ are independent Gaussian variables with mean zero and variance $\sigma^2$. The null hypothesis of interest is that $\rho = 1$, which implies that $X_t$ is a random walk, that is, a non-stationary process. Dickey and Fuller found among other results that when $\rho = 1$, a non-standard limit distribution is obtained, and this can be expressed as

\[ T(\hat{\rho} - 1) = \frac{\sum_{t=1}^{T} X_{t-1} \epsilon_t}{\sum_{t=1}^{T} X_{t-1}^2} \Rightarrow \frac{\int_0^1 W(dW)}{\int_0^1 W^2 dW}, \]

where $W(u)$ is a univariate Brownian motion on $[0,1]$ with variance $\sigma^2$. The implication is that the likelihood ratio test statistic is asymptotically distributed as

\[ \left( \frac{\int_0^1 W(dW)}{\int_0^1 W^2 dW} \right)^2. \]

This distribution is often called the ‘unit root’ or Dickey–Fuller distribution and its multivariate version plays an important role in asymptotic inference for cointegration. We give the main results obtained for likelihood inference, and refer to Johansen (1988, 1991) and Ahn and Reinsel (1990) for the technical details.
Theorem 7 Under the model with $\Phi = 0$ and $r$ cointegrating relations the likelihood ratio statistic Eq. (13) satisfies

$$-2 \ln Q(H_0|H_n) \overset{d}{\rightarrow} \text{tr} \left( \int_0^1 (dB)B\left[\int_0^1 BB' \, du\right]^{-1}\int_0^1 B(dB)\right),$$

where the process $B$ is an $(n - r)$-dimensional Brownian motion with covariance matrix equal to $I$.

Thus, the limit distribution only depends on the number of common trends of the problem. It is seen that the distribution is a multivariate generalization of the unit root distribution. This is not surprising, since one can think of the test for $r = 1$ in the univariate model as a test for no cointegration, i.e. of $r = 0$, when $n = 1$, and $k = 1$.

Although the limit distribution given in Theorem 7 only depends on the degrees of freedom or the dimension of the Brownian motion, it turns out that if a constant term or a linear term is allowed in the model then the limit distribution changes. The various limit distributions are tabulated by simulation (see Johansen, 1996).

4.2. Test for restrictions on $\beta$

It is quite satisfactory, however, that the other test statistics described in Section 3 for hypotheses on $\alpha$ and $\beta$ all have asymptotic $\chi^2$ distributions. Thus, the only non-standard test is the test for cointegrating rank. The reason for this is that the asymptotic distribution of the estimator of $\beta$ is a mixed Gaussian distribution. We give the result for $\beta$, that is, $\beta$ normalized so that $c\beta = I$.

Theorem 8 The asymptotic distribution of $\hat{\beta}$ is given by

$$T\left(\hat{\beta} - \beta_0\right) \overset{d}{\rightarrow} (1 - \beta_0c')\beta_0 \left[\int_0^1 B_1B_1' \, du\right]^{-1}\int_0^1 B_1(dB_1),$$

where $B_1$ and $B_2$ are independent Brownian motions of dimension $n - r$ and $r$, respectively. The asymptotic conditional variance matrix is

$$(1 - \beta_0c')\beta_0 \left[\int_0^1 B_1B_1' \, du\right]^{-1}\beta_0' (I - c\beta_0') \otimes (\alpha_0'\Omega^{-1}\alpha_0)^{-1},$$

which is consistently estimated by

$$T\left(I - \hat{\beta}_0c\right)S_{ii}^{-1}(1 - c\hat{\beta}_0') \otimes \left(\hat{\alpha}_0'\hat{\Omega}^{-1}\hat{\alpha}_0\right)^{-1}.$$

For given value of $B_1$ the limit distribution of $\hat{\beta}$ is just a Gaussian distribution with mean zero and variance given by Eq. (17). It is this result that implies, by a simple conditioning argument, that the likelihood ratio test statistics for hypotheses about
restrictions on $\beta$ are asymptotically distributed as $\chi^2$ variables, which again makes inference about $\beta$ very simple if likelihood based methods are used.

Despite the complicated formulation the result is surprisingly simple, since it only states that if $\beta$ is estimated as identified parameters then the asymptotic variance of $\hat{\beta}$ is given by the inverse information matrix, which is the Hessian used in the numerical maximization of a function. This result is exactly the same as the result that holds for inference in stationary processes. The only difference is the interpretation of Eq. (18), which for a stationary process would be an estimate of the asymptotic variance, but for $I(1)$ processes is a consistent estimator of the asymptotic conditional variance. The basic property, however, is the same in both cases, namely that it is the approximate scale parameter to use for normalizing the deviation $\hat{\beta} - \beta$.

Inference for the remaining parameters $\vartheta = (\alpha, \Gamma_1, \ldots, \Gamma_r, \Omega)$ is different. This is explained by Phillips (1991) and the idea is roughly the following. The second derivative of the log likelihood function with respect to $\beta$ tends to infinity as $T^2$, whereas the second derivative with respect to $\vartheta$ and the mixed derivatives tend to infinity like $T$. This means that $\hat{\beta} - \beta$ has to be normalized by $T$ and $\hat{\vartheta} - \vartheta$ by $T^{1/2}$. This on the other hand requires a normalization of the mixed derivatives by $T^{3/2}$ and makes them disappear in the limit. Thus, in the limit the information matrix, which is used to normalize $(\hat{\beta} - \beta, \hat{\vartheta} - \vartheta)$, is block diagonal with one block for $\beta$ and one block for the remaining parameters $\vartheta$.

5. Various applications of the cointegration model

The concept of cointegration and long-run relations can be found in many models of economic interest, since data is often stationary and the cointegrating relations correspond to the relations that have the smallest variance or which are the most stationary.

5.1. Rational expectations

A typical rational expectation model is given by the present value model in Campbell and Schiller (1987) which states that the present value of a variable $Y_t$ is a linear function of the discounted future values $y_t$.

$$Y_t = \gamma (1 - \delta) \sum_{j=0}^{\infty} \delta^j E_t y_{t+j} + c$$

If we let $X_t = (Y_t, y_t)$ we can write the equations on the form

$$E_t (c'X_{t+1} + c'X_t) + c = 0$$

If we assume that $X_t$ is given by autoregressive model
\[ \Delta X_t = \Pi X_{t-1} + \mu + \epsilon_t \]

the parameters must satisfy the restrictions

\[ c'_t \Pi = -(c_0 + c_t)' \quad \text{and} \quad c'_t \mu + c = 0, \]

which together with the cointegrating restriction \( \Pi = \alpha \beta' \) gives a set of restrictions on the parameters that can be tested. The idea is first to determine or test for the cointegrating rank, and next to estimate the model under the above restrictions and compare the likelihoods obtained using a likelihood ratio test, see Johansen and Swensen (1999).

5.2. Arbitrage pricing theory

The arbitrage pricing theory often describes a one period model for asset prices and derives a restriction on the mean return by assuming that there should be no arbitrage opportunity by creating a portfolio with positive excess mean return and no risk. The exact factor model does not allow this possibility but the lack of the restriction opens an approximate arbitrage opportunity by diversification over many assets.

If, however, we consider a multi-period model and instead of the asset returns consider the cumulated asset returns or the log prices, then these variables are found to be non-stationary. If we fit a cointegration model with a linear term restricted to the cointegrating relations we get a model for a rebalanced portfolio. The no arbitrage condition is then that any portfolio with linearly increasing mean and constant risk must have the same mean return (see Johansen and Lando, 1996). Thus, the APT hypothesis is a restriction on the deterministic terms and the cointegrating vectors.

5.3. Seasonal cointegration

A phenomenon that is not directly covered by the above model (1) is the seasonal variation of time series (see Hylleberg et al., 1990). It turns out that the basic result about inversion of matrix polynomials can be extended to cover this case as well, with the result that we get an error correction formulation and the possibility for testing for cointegrating rank at the various complex frequencies. The asymptotics are roughly the same as for the usual model but involves the complex Brownian motion, as a consequence of allowing roots at complex frequencies (see Johansen and Schaumburg, 1999).

References