Cointegration. Overview and Development

Søren Johansen

Department of Applied Mathematics and Statistics, University of Copenhagen
sjo@math.ku.dk

Summary. This article presents a survey of the analysis of cointegration using the vector autoregressive model. After a few illustrative economic examples, the three model based approaches to the analysis of cointegration are discussed. The vector autoregressive model is defined and the moving average representation of the solution, the Granger representation, is given. Next the interpretation of the model and its parameters and likelihood based inference follows using reduced rank regression. The asymptotic analysis includes the distribution of the Gaussian maximum likelihood estimators, the rank test, and test for hypotheses on the cointegrating vectors. Finally, some applications and extensions of the basic model are mentioned and the survey concludes with some open problems.

1 Introduction

Granger [13] coined the term cointegration as a formulation of the phenomenon that nonstationary processes can have linear combinations that are stationary. It was his investigations of the relation between cointegration and error correction that brought modeling of vector autoregressions with unit roots and cointegration to the center of attention in applied and theoretical econometrics; see Engle and Granger [10].

During the last 20 years, many have contributed to the development of theory and applications of cointegration. The account given here focuses on theory, more precisely on likelihood based theory for the vector autoregressive model and its extensions; see [21]. By building a statistical model as a framework for inference, one has to make explicit assumptions about the model used and hence has a possibility of checking the assumptions made.

1.1 Two examples of cointegration

As a first simple economic example of the main idea in cointegration, consider the exchange rate series, e_t, between Australian and US dollars and four time series p_t^{au}, $p_t^{us}, i_t^{au}, i_t^{us}$: log consumer price and five year treasury bond rates in Australia and US. If the quarterly series from 1972:1 to 1991:1 are plotted, they clearly show nonstationary behavior, and we discuss in the following a method of modeling such nonstationary time series. As a simple example of an economic hypothesis consider Purchasing Power Parity (PPP), which asserts that $e_t = p_t^{us} - p_t^{au}$. This identity is not found in the data, so a more realistic formulation is that $ppp_t = e_t - p_t^{us} + p_t^{au}$ is a stationary process, possibly with mean zero. Thus we formulate the economic relation, as expressed by PPP, as a stationary relation among nonstationary processes. The purpose of modeling could be to test the null hypothesis that ppp_t is stationary, or in other words that $(e_t, p_t^{us}, p_t^{au} i_t^{au}, i_t^{us})$ cointegrate with $(1, -1, 1, 0, 0)'$ as a cointegration vector. If that is not found, an outcome could be to suggest other cointegration relations, which have a better chance of capturing comovements of the five processes in the information set. For a discussion of the finding that real exchange rate, ppp_t, and the spread, $i_t^{au} - i_t^{us}$, are cointegrated $I(1)$ processes so that a linear combination $ppp_t - c\,(i_t^{au} - i_t^{us})$ is stationary, see Juselius and MacDonald [31].

Another example is one of the first applications of the idea of cointegration in finance; see Campbell and Shiller [8]. They considered a present value model for the price of a stock Y_t at the end of period t and the dividend y_t paid during period t. They assume that there is a vector autoregressive model describing the data which contain Y_t and y_t and may contain values of other financial assets. The expectations hypothesis is expressed as

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

where c and θ are positive constants and the discount factor δ is between 0 and 1. The notation $E_t y_{t+i}$ means model based conditional expectations of y_{t+i} given information in the data at the end of period t. By subtracting θy_t, the model is written as

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

It is seen that when the processes y_t and Y_t are nonstationary and their differences stationary, the present value model implies that the right hand side and hence the left hand side are stationary. Thus there is cointegration between Y_t and y_t with a cointegration vector $\beta' = (1, -\theta, 0, \ldots, 0)$; see section 6.1 for a discussion of rational expectations and cointegration.

There are at present three different ways of modeling the cointegration idea in a parametric statistical framework. To illustrate the ideas they are formulated in the simplest possible case, leaving out deterministic terms.

1.2 Three ways of modeling cointegration

The regression formulation

The multivariate process $x_t = (x'_{1t}, x'_{2t})'$ of dimension $p = p_1 + p_2$ is given by the regression equations

$$x_{1t} = \gamma' x_{2t} + u_{1t},$$
$$\Delta x_{2t} = u_{2t},$$

where $u_t = (u'_{1t}, u'_{2t})'$ is a linear invertible process defined by i.i.d. errors ε_t with mean zero and finite variance. The assumptions behind this model imply that x_{2t} is nonstationary and not cointegrated, and hence the cointegration rank, p_1, is known so that models for different ranks are not nested. The first estimation method used in this model is least squares regression, Engle and Granger [10], which is shown to give a superconsistent estimator by Stock [46]. This estimation method gives rise to residual based tests for cointegration. It was shown by Phillips and Hansen [42] that a modification of the regression estimator, involving a correction using the long-run variance of the process u_t, would give useful methods for inference for coefficients of cointegration relations; see also Phillips [41].

The autoregressive formulation

The autoregressive formulation is given by

$$\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t,$$

where ε_t are i.i.d. errors with mean zero and finite variance, and α and β are $p \times r$ matrices of rank r. Under the condition that Δx_t is stationary, the solution is

$$x_t = C \sum_{i=1}^{t} \varepsilon_i + \sum_{i=0}^{\infty} C_i \varepsilon_{t-i} + A, \tag{1}$$

where $C = \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp$ and $\beta' A = 0$. Here $\beta_\perp$ is a full rank $p \times (p - r)$ matrix so that $\beta' \beta_\perp = 0$. This formulation allows for modeling of both the long-run relations, $\beta' x$, and the adjustment, or feedback α, towards the attractor set $\{x : \beta' x = 0\}$ defined by the long-run relations. Models for different cointegration ranks are nested and the rank can be analyzed by likelihood ratio tests. Thus the model allows for a more detailed description of the data than the regression model. Methods usually applied for the analysis are derived from the Gaussian likelihood function, which are discussed here; see also [18, 21], and Ahn and Reinsel [1].

The unobserved component formulation

Let x_t be given by

$$x_t = \xi\eta' \sum_{i=1}^{t} \varepsilon_i + v_t,$$

where v_t is a linear process, typically independent of the process ε_t, which is i.i.d. with mean zero and finite variance.

In this formulation too, hypotheses of different ranks are nested. The parameters are linked to the autoregressive formulation by $\xi = \beta_\perp$ and $\eta = \alpha_\perp$, even though the linear process in (1) depends on the random walk part, so the unobserved components model and the autoregressive model are not the same. Thus both adjustment and cointegration can be discussed in this formulation, and hypotheses on the rank can be tested. Rather than testing for unit roots one tests for stationarity, which is sometimes a more natural formulation. Estimation is usually performed by the Kalman filter, and asymptotic theory of the rank tests has been worked out by Nyblom and Harvey [37].

1.3 The model analyzed in this article

In this article cointegration is modelled by the vector autoregressive model for the $p-$dimensional process x_t

$$\Delta x_t = \alpha\,(\beta' x_{t-1} + \Upsilon D_t) + \sum_{i=1}^{k-1} \Gamma_i\, \Delta x_{t-i} + \Phi d_t + \varepsilon_t, \tag{2}$$

where ε_t are i.i.d. with mean zero and variance Ω, and D_t and d_t are deterministic terms, like constant, trend, seasonal- or intervention dummies. The matrices α and β are $p \times r$ where $0 \leq r \leq p$. The parametrization of the deterministic term $\alpha\Upsilon D_t + \Phi d_t$, is discussed in section 2.2. Under suitable conditions, see again section 2.2, the processes $\beta' x_t$ and Δx_t are stationary around their means, and (2) can be formulated as

$$\Delta x_t - E(\Delta x_t) = \alpha\,(\beta' x_{t-1} - E(\beta' x_{t-1})) + \sum_{i=1}^{k-1} \Gamma_i\,(\Delta x_{t-i} - E(\Delta x_{t-i})) + \varepsilon_t.$$

This shows how the change of the process reacts to feedback from disequilibrium errors $\beta' x_{t-1} - E(\beta' x_{t-1})$ and $\Delta x_{t-i} - E(\Delta x_{t-i})$, via the short-run adjustment coefficients α and Γ_i. The equation $\beta' x_t - E(\beta' x_t) = 0$ defines the long-run relations between the processes.

There are many surveys of the theory of cointegration; see for instance Watson [48] or Johansen [26]. The topic has become part of most textbooks in econometrics; see among others Banerjee, Dolado, Galbraith and Hendry [4], Hamilton [14], Hendry [17] and Lütkepohl [34]. For a general account of the methodology of the cointegrated vector autoregressive model, see Juselius [32].

2 Integration, cointegration and Granger's Representation Theorem

The basic definitions of integration and cointegration are given together with a moving average representation of the solution of the error correction model (2). This solution reveals the stochastic properties of the solution. Finally the interpretation of cointegration relations is discussed.

2.1 Definition of integration and cointegration

The vector autoregressive model for the $p-$dimensional process x_t given by (2) is a dynamic stochastic model for all components of x_t. By recursive substitution, the equations define x_t as function of initial values, $x_0, \ldots, x_{-k+1}$, errors $\varepsilon_1, \ldots, \varepsilon_t$, deterministic terms, and parameters. Properties of the solution of these equations are studied through the characteristic polynomial

$$\Psi(z) = (1-z)I_p - \Pi z - (1-z)\sum_{i=1}^{k-1} \Gamma_i z^i \tag{3}$$

with determinant $|\Psi(z)|$. The function $C(z) = \Psi(z)^{-1}$ has poles at the roots of the polynomial $|\Psi(z)|$ and the position of the poles determines the stochastic properties of the solution of (2). First a well known result is mentioned; see Anderson [3].

Theorem 1. *If $|\Psi(z)| = 0$ implies that $|z| > 1$, then α and β have full rank p, and the coefficients of $\Psi^{-1}(z) = \sum_{i=0}^{\infty} C_i z^i$ are exponentially decreasing. Let $\mu_t = \sum_{i=0}^{\infty} C_i(\alpha \Upsilon D_{t-i} + \Phi d_{t-i})$. Then the distribution of the initial values of x_t can be chosen so that $x_t - \mu_t$ is stationary. Moreover, x_t has the moving average representation*

$$x_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i} + \mu_t. \tag{4}$$

Thus the exponentially decreasing coefficients are found by simply inverting the characteristic polynomial if the roots are outside the unit disk. If this condition fails, the equations generate nonstationary processes of various types, and the coefficients are not exponentially decreasing. Still, the coefficients of $C(z)$ determine the stochastic properties of the solution of (2), as is discussed in section 2.2. A process of the form (4) is a linear process and forms the basis for the definitions of integration and cointegration.

Definition 1. *The process x_t is integrated of order 1, $I(1)$, if $\Delta x_t - E(\Delta x_t)$ is a linear process, with $C(1) = \sum_{i=0}^{\infty} C_i \neq 0$. If there is a vector $\beta \neq 0$ so that $\beta' x_t$ is stationary around its mean, then x_t is cointegrated with cointegration vector β. The number of linearly independent cointegration vectors is the cointegration rank.*

Example 1. A bivariate process is given for $t = 1, \ldots, T$ by the equations

$$\begin{aligned} \Delta x_{1t} &= \alpha_1(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t}, \\ \Delta x_{2t} &= \alpha_2(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t}. \end{aligned}$$

Subtracting the equations, we find that the process $y_t = x_{1t} - x_{2t}$ is autoregressive and stationary if $|1+\alpha_1 - \alpha_2| < 1$ and the initial value is given by its invariant distribution. Similarly we find that $S_t = \alpha_2 x_{1t} - \alpha_1 x_{2t}$ is a random walk, so that

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \frac{1}{\alpha_2 - \alpha_1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} S_t - \frac{1}{\alpha_2 - \alpha_1} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} y_t.$$

This shows, that when $|1 + \alpha_1 - \alpha_2| < 1$, x_t is $I(1)$, $x_{1t} - x_{2t}$ is stationary, and $\alpha_2 x_{1t} - \alpha_1 x_{2t}$ is a random walk, so that x_t is a cointegrated $I(1)$ process with cointegration vector $\beta' = (1, -1)$. We call S_t a common stochastic trend and α the adjustment coefficients.

Example 1 presents a special case of the Granger Representation Theorem, which gives the moving average representation of the solution of the error correction model.

2.2 The Granger Representation Theorem

If the characteristic polynomial $\Psi(z)$ defined in (3) has a unit root, then $\Psi(1) = -\Pi$ is singular, of rank $r < p$, and the process is not stationary. Next the $I(1)$ condition is formulated. It guarantees that the solution of (2) is a cointegrated $I(1)$ process. Let $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$ and denote for a $p \times m$ matrix a by $a_\perp$ a $p \times (p - m)$ matrix of rank $p - m$.

Condition 1. *(The $I(1)$ condition) The $I(1)$condition is satisfied if $|\Psi(z)| = 0$ implies that $|z| > 1$ or $z = 1$ and that*

$$|\alpha_\perp' \Gamma \beta_\perp| \neq 0. \tag{5}$$

Condition (5) is needed to avoid solutions that are integrated of order 2 or higher; see section 6. For a process with one lag $\Gamma = I_p$ and

$$\beta' x_t = (I_r + \beta' \alpha) \beta' x_{t-1} + \beta' \varepsilon_t.$$

In this case the $I(1)$ condition is equivalent to the condition that the absolute value of the eigenvalues of $I_r + \beta'\alpha$ are bounded by one, and in example 1 the condition is $|1 + \alpha_1 - \alpha_2| < 1$.

Theorem 2. (The Granger Representation Theorem) *Let $\Psi(z)$ be defined by* (3). *If $\Psi(z)$ has unit roots and the $I(1)$ condition is satisfied, then*

$$(1-z)\Psi(z)^{-1} = C(z) = \sum_{i=0}^{\infty} C_i z^i = C + (1-z)C^*(z) \tag{6}$$

converges for $|z| \leq 1 + \delta$ *for some* $\delta > 0$ *and*

$$C = \beta_\perp(\alpha_\perp' \Gamma \beta_\perp)^{-1}\alpha_\perp'. \tag{7}$$

The solution x_t *of equation* (2) *has the moving average representation*

$$x_t = C\sum_{i=1}^{t}(\varepsilon_i + \Phi d_i) + \sum_{i=0}^{\infty} C_i^*(\varepsilon_{t-i} + \Phi d_{t-i} + \alpha \Upsilon D_{t-i}) + A, \tag{8}$$

where A *depends on initial values, so that* $\beta' A = 0$.

This result implies that Δx_t and $\beta' x_t$ are stationary, so that x_t is a cointegrated $I(1)$ process with r cointegration vectors β and $p-r$ common stochastic trends $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$. The interpretation of this is that among p nonstationary processes the model (2) generates r stationary or stable relations and $p-r$ stochastic trends or driving trends, which create the nonstationarity.

The result (6) rests on the observation that the singularity of $\Psi(z)$ for $z = 1$ implies that $\Psi(z)^{-1}$ has a pole at $z = 1$. Condition (5) is a condition for this pole to be of order one. This is not proved here, see [27], but it is shown how this result can be applied to prove the representation result (8), which shows how coefficients of the inverse polynomial determine the properties of x_t.

We multiply $\Psi(L)x_t = \Phi d_t + \alpha \Upsilon D_t + \varepsilon_t$ by

$$(1-L)\Psi(L)^{-1} = C(L) = C + (1-L)C^*(L)$$

and find

$$\Delta x_t = (1-L)\Psi(L)^{-1}\Psi(L)x_t = (C + \Delta C^*(L))(\varepsilon_t + \alpha \Upsilon D_t + \Phi d_t).$$

Now define the stationary process $z_t = C^*(L)\varepsilon_t$ and the deterministic function $\mu_t = C^*(L)(\alpha \Upsilon D_t + \Phi d_t)$, and note that $C\alpha\Upsilon = 0$, so that

$$\Delta x_t = C(\varepsilon_t + \Phi d_t) + \Delta(z_t + \mu_t),$$

which cumulates to

$$x_t = C\sum_{i=1}^{t}(\varepsilon_i + \Phi d_i) + z_t + \mu_t + A,$$

where $A = x_0 - z_0 - \mu_0$. The distribution of x_0 is chosen so that $\beta' x_0 = \beta'(z_0 + \mu_0)$, and hence $\beta' A = 0$. Then x_t is $I(1)$ and $\beta' x_t = \beta' z_t + \beta' \mu_t$ is stationary around its mean $E(\beta' x_t) = \beta' \mu_t$. Finally, Δx_t is stationary around its mean $E(\Delta x_t) = C\Phi d_t + \Delta \mu_t$.

One of the useful applications of the representation (8) is to investigate the role of the deterministic terms. Note that d_t cumulates in the process with a coefficient $C\Phi$, but that D_t does not, because $C\alpha\Upsilon = 0$. A leading special case is the model with $D_t = t$, and $d_t = 1$, which ensures that any linear combination of the components of x_t is allowed to have a linear trend. Note that if $D_t = t$ is not allowed in the model, that is $\Upsilon = 0$, then x_t has a trend given by $C\Phi t$, but the cointegration relation $\beta' x_t$ has no trend because $\beta' C\Phi = 0$.

2.3 Interpretation of cointegrating coefficients

Consider first a usual regression

$$x_{1t} = \gamma_2 x_{2t} + \gamma_3 x_{3t} + \varepsilon_t, \tag{9}$$

with i.i.d. errors ε_t which are independent of the processes x_{2t} and x_{3t}. The coefficient γ_2 is interpreted via a counterfactual experiment, that is, the coefficient γ_2 is the effect on x_{1t} of a change in x_{2t}, keeping x_{3t} constant.

The cointegration relations are long-run relations. This means that they have been there all the time, and they influence the movement of the process x_t via the adjustment coefficients α. The more the process $\beta' x_t$ deviates from $E\beta' x_t$, the more the adjustment coefficients pull the process back towards its mean. Another interpretation is that they are relations that would hold in the limit, provided all shocks in the model are set to zero after a time t.

It is therefore natural that interpretation of cointegration coefficients involves the notion of a long-run value. From the Granger Representation Theorem 8 applied to the model with no deterministic terms, it can be proved, see [25], that

$$x_{\infty|t} = \lim_{h\to\infty} E(x_{t+h}|x_t, \ldots, x_{t-k+1}) = C(x_t - \sum_{i=1}^{k-1} \Gamma_i x_{t-i}) = C\sum_{i=1}^{t} \varepsilon_i + x_{\infty|0}.$$

This limiting conditional expectation is a long-run value of the process. Because $\beta' x_{\infty|t} = 0$, the point $x_{\infty|t}$ is in the attractor set $\{x : \beta' x = 0\} = sp\{\beta_\perp\}$, see Figure 1.

Thus if the current value, x_t, is shifted to $x_t + h$, then the long-run value is shifted from $x_{\infty|t}$ to $x_{\infty|t} + Ch$, which is still a point in the attractor set because $\beta' x_{\infty|t} + \beta' Ch = 0$. If a given long-run change $k = C\xi$ in $x_{\infty|t}$ is needed, Γk is added to the current value x_t. This gives the long-run value

$$x_{\infty|t} + C\Gamma k = x_{\infty|t} + C\Gamma C\xi = x_{\infty|t} + C\xi = x_{\infty|t} + k,$$

where the identity $C\Gamma C = C$ is applied; see (7). This idea is now used to give an interpretation of a cointegration coefficient in the simple case of $r = 1$, $p = 3$, and where the relation is normalized on x_1

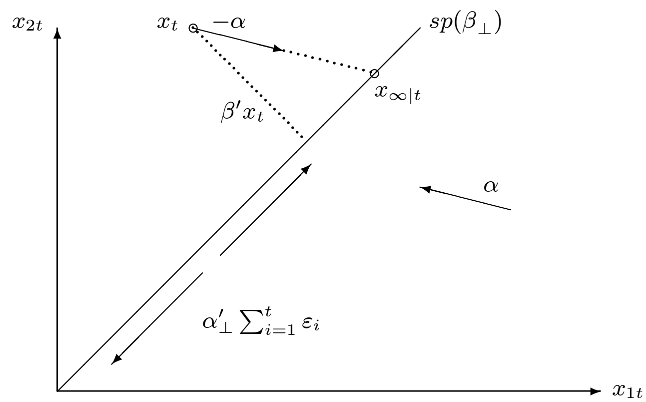

Fig. 1. In the model $\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t$, the point $x_t = (x_{1t}, x_{2t})$ is moved towards the long-run value $x_{\infty|t}$ on the attractor set $\{x|\beta' x = 0\} = sp(\beta_\perp\}$ by the forces $-\alpha$ or $+\alpha$, and pushed along the attractor set by the common trends $\alpha'_\perp \sum_{i=1}^t \varepsilon_i$.

$$x_1 = \gamma_2 x_2 + \gamma_3 x_3, \tag{10}$$

so that $\beta' = (1, -\gamma_2, -\gamma_3)$. In order to give the usual interpretation as a regression coefficient (or elasticity if the measurements are in logs), a long-run change with the properties that x_2 changes by one, x_1 changes by γ_2, and x_3 is kept fixed, is needed. Thus the long-run change is $k = (\gamma_2, 1, 0)$, which satisfies $\beta' k = 0$, so that $k = C\xi$ for some ξ, and this can be achieved by moving the current value from x_t to $x_t + C\Gamma k$. In this sense, a coefficient in an identified cointegration relation can be interpreted as the effect of a long-run change to one variable on another, keeping all others fixed in the long run. More details can be found in [25] and Proietti [43].

3 Interpretation of the $I(1)$ model for cointegration

In this section model $H(r)$ defined by (2) is discussed. The parameters in $H(r)$ are

$$(\alpha,\, \beta,\, \Gamma_1,\, \ldots,\, \Gamma_{k-1},\, \Upsilon,\, \Phi,\, \Omega)\,.$$

All parameters vary freely and α and β are $p \times r$ matrices. The normalization and identification of α and β are discussed, and some examples of hypotheses on α and β, which are of economic interest are given.

3.1 The models $H(r)$

The models $H(r)$ are nested

$$H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(p).$$

Here $H(p)$ is the unrestricted vector autoregressive model, so that α and β are unrestricted $p \times p$ matrices. The model $H(0)$ corresponds to the restriction $\alpha = \beta = 0$, which is the vector autoregressive model for the process in differences. Note that in order to have nested models, we allow in $H(r)$ for all processes with rank less than or equal to r.

The formulation allows us to derive likelihood ratio tests for the hypothesis $H(r)$ in the unrestricted model $H(p)$. These tests can be applied to check if one's prior knowledge of the number of cointegration relations is consistent with the data, or alternatively to construct an estimator of the cointegration rank.

Note that when the cointegration rank is r, the number of common trends is $p-r$. Thus if one can interpret the presence of r cointegration relations one should also interpret the presence of $p-r$ independent stochastic trends or $p-r$ driving forces in the data.

3.2 Normalization of parameters of the $I(1)$ model

The parameters α and β in (2) are not uniquely identified, because given any choice of α and β and any nonsingular $r \times r$ matrix ξ, the choice $\alpha\xi$ and $\beta\xi'^{-1}$ gives the same matrix $\Pi = \alpha\beta' = \alpha\xi(\beta\xi'^{-1})'$.

If $x_t = (x'_{1t}, x'_{2t})'$ and $\beta = (\beta'_1, \beta'_2)'$, with $|\beta_1| \neq 0$, we can solve the cointegration relations as

$$x_{1t} = \gamma' x_{2t} + u_t,$$

where u_t is stationary and $\gamma' = -(\beta'_1)^{-1}\beta'_2$. This represents cointegration as a regression equation. A normalization of this type is sometimes convenient for estimation and calculation of 'standard errors' of the estimate, see section 5.2, but many hypotheses are invariant with respect to a normalization of β, and thus, in a discussion of a test of such a hypothesis, β does not require normalization. As seen in subsection 3.3, many stable economic relations are expressed in terms of identifying restrictions, for which the regression formulation is not convenient.

From the Granger Representation Theorem we see that the $p-r$ common trends are the nonstationary random walks in $C\sum_{i=1}^t \varepsilon_i$, that is, can be chosen as $\alpha'_\perp \sum_{i=1}^t \varepsilon_i$. For any full rank $(p-r) \times (p-r)$ matrix η, $\eta\alpha'_\perp \sum_{i=1}^t \varepsilon_i$ could also be used as common trends because

$$C\sum_{i=1}^{t} \varepsilon_i = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}(\alpha'_\perp \sum_{i=1}^{t} \varepsilon_i) = \beta_\perp(\eta\alpha'_\perp \Gamma \beta_\perp)^{-1}(\eta\alpha'_\perp \sum_{i=1}^{t} \varepsilon_i).$$

Thus identifying restrictions on the coefficients in $\alpha_\perp$ are needed to find their estimates and standard errors.

In the cointegration model there are therefore three different identification problems: one for the cointegration relations, one for the common trends, and finally one for the short run dynamics, if the model has simultaneous effects.

3.3 Hypotheses on long-run coefficients

The purpose of modeling economic data is to test hypotheses on the coefficients, thereby investigating whether the data supports an economic hypothesis or rejects it. In the example with the series $x_t = (p_t^{au},\, p_t^{us}, i_t^{au}, i_t^{us}, e_t)'$ the hypothesis of PPP is formulated as the hypothesis that $(1, -1, 1, 0, 0)$ is a cointegration relation. Similarly, the hypothesis of price homogeneity is formulated as

$$R'\beta = (1, 1, 0, 0, 0)\beta = 0,$$

or equivalently as $\beta = R_\perp \varphi = H\varphi$, for some vector φ and $H = R_\perp$. The hypothesis that the interest rates are stationary is formulated as the hypothesis that the two vectors $(0, 0, 0, 1, 0)$ and $(0, 0, 0, 0, 1)$ are cointegration vectors. A general formulation of restrictions on each of r cointegration vectors, including a normalization, is

$$\beta = (h_1 + H_1\varphi_1, \ldots, h_r + H_r\varphi_r)\,. \tag{11}$$

Here h_i is $p \times 1$ and orthogonal to H_i which is $p \times (s_i - 1)$ of rank $s_i - 1$, so that $p - s_i$ restrictions are imposed on the vector β_i. Let $R_i = (h_i, H_i)_\perp$ then β_i satisfies the restrictions $R_i'\beta_i = 0$, and the normalization $(h_i'h_i)^{-1}h_i'\beta_i = 1$. Wald's identification criterion is that β_i is identified if

$$R_i'(\beta_1, \ldots, \beta_r) = r - 1.$$

3.4 Hypotheses on adjustment coefficients

The coefficients in α measure how the process adjusts to disequilibrium errors. The hypothesis of weak exogeneity is the hypothesis that some rows of α are zero; see Engle, Hendry and Richard [11]. The process x_t is decomposed as $x_t = (x_{1t}',\, x_{2t}')'$ and the matrices are decomposed similarly so that the model equations without deterministic terms become

$$\begin{aligned}\Delta x_{1t} &= \alpha_1\beta' x_{t-1} + \textstyle\sum_{i=1}^{k-1} \Gamma_{1i}\Delta x_{t-i} + \varepsilon_{1t},\\ \Delta x_{2t} &= \alpha_2\beta' x_{t-1} + \textstyle\sum_{i=1}^{k-1} \Gamma_{2i}\Delta x_{t-i} + \varepsilon_{2t}.\end{aligned}$$

If $\alpha_2 = 0$, there is no levels feedback from $\beta' x_{t-1}$ to Δx_{2t}, and if the errors are Gaussian, x_{2t} is weakly exogenous for α_1, β. The conditional model for Δx_{1t} given Δx_{2t} and the past is

$$\Delta x_{1t} = \omega\Delta x_{2t} + \alpha_1\beta' x_{t-1} + \sum_{i=1}^{k-1}(\Gamma_{1i} - \omega\Gamma_{2i})\Delta x_{t-i} + \varepsilon_{1t} - \omega\varepsilon_{2t}, \tag{12}$$

where $\omega = \Omega_{12}\Omega_{22}^{-1}$. Thus full maximum likelihood inference on α_1 and β can be conducted in the conditional model (12).

An interpretation of the hypothesis of weak exogeneity is the following: if $\alpha_2 = 0$ then $\alpha_\perp$ contains the columns of $(0, I_{p-r})'$, so that $\sum_{i=1}^{t} \varepsilon_{2i}$ are common trends. Thus the errors in the equations for x_{2t} cumulate in the system and give rise to nonstationarity.

4 Likelihood analysis of the $I(1)$ model

This section contains first some comments on what aspects are important for checking for model misspecification, and then describes the calculation of reduced rank regression, introduced by Anderson [2]. Then reduced rank regression and modifications thereof are applied to estimate the parameters of the $I(1)$ model (2) and various submodels.

4.1 Checking the specifications of the model

In order to apply Gaussian maximum likelihood methods, the assumptions behind the model have to be checked carefully, so that one is convinced that the statistical model contains the density that describes the data. If this is not the case, the asymptotic results available from the Gaussian analysis need not hold. Methods for checking vector autoregressive models include choice of lag length, test for normality of residuals, tests for autocorrelation, and test for heteroscedasticity in errors. Asymptotic results for estimators and tests derived from the Gaussian likelihood turn out to be robust to some types of deviations from the above assumptions. Thus the limit results hold for i.i.d. errors with finite variance, and not just for Gaussian errors, but autocorrelated errors violate the asymptotic results, so autocorrelation has to be checked carefully.

Finally and perhaps most importantly, the assumption of constant parameters is crucial. In practice it is important to model outliers by suitable dummies, but it is also important to model breaks in the dynamics, breaks in the cointegration properties, breaks in the stationarity properties, etc. The papers by Seo [45] and Hansen and Johansen [16] contain some results on recursive tests in the cointegration model.

4.2 Reduced rank regression

Let u_t, w_t, and z_t be three multivariate time series of dimensions p_u, p_w, and p_z respectively. The algorithm of reduced rank regression, see Anderson [2], can be described in the regression model

$$u_t = \alpha\beta' w_t + \Gamma z_t + \varepsilon_t, \tag{13}$$

where ε_t are the errors with variance Ω. The product moments are

$$S_{uw} = T^{-1}\sum_{t=1}^{T} u_t w_t',$$

and the residuals, which we get by regressing u_t on w_t, are

$$(u_t|w_t) = u_t - S_{uw}S_{ww}^{-1}w_t,$$

so that the conditional product moments are

$$S_{uw.z} = S_{uw} - S_{uz}S_{zz}^{-1}S_{zw} = T^{-1}\sum_{t=1}^{T}(u_t|z_t)(w_t|zv)',$$

$$S_{uu.w,z} = T^{-1}\sum_{t=1}^{T}(u_t|w_t, z_t)(u_t|w_t, z_t)' = S_{uu.w} - S_{uz.w}S_{zz.w}^{-1}S_{zu.w}.$$

Let $\Pi = \alpha\beta'$. The unrestricted regression estimates are

$$\hat{\Pi} = S_{uw.z}S_{ww.z}^{-1},\ \hat{\Gamma} = S_{uz.w}S_{zz.w}^{-1},\ \text{and}\ \hat{\Omega} = S_{uu.w,z}.$$

Reduced rank regression of u_t on w_t corrected for z_t gives estimates of α, β and Ω in (13). First the eigenvalue problem

$$|\lambda S_{ww.z} - S_{wu.z}S_{uu.z}^{-1}S_{uw.z}| = 0 \tag{14}$$

is solved. The eigenvalues are ordered $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{p_w}$, with corresponding eigenvectors $\hat{v}_1, \ldots, \hat{v}_{p_w}$. The reduced rank estimates of β, α, Γ, and Ω are given by

$$\begin{aligned}
\hat{\beta} &= (\hat{v}_1, \ldots, \hat{v}_r),\\
\hat{\alpha} &= S_{uw.z}\hat{\beta},\\
\hat{\Gamma} &= S_{uz.\hat{\beta}'w}S_{zz.\hat{\beta}'w}^{-1},\\
\hat{\Omega} &= S_{uu.z} - S_{uw.z}\hat{\beta}\hat{\beta}'S_{wu.z},\\
|\hat{\Omega}| &= |S_{uu.z}|\textstyle\prod_{i=1}^{r}(1-\hat{\lambda}_i).
\end{aligned} \tag{15}$$

The eigenvectors are orthogonal because $\hat{v}_i'S_{ww.z}\hat{v}_j = 0$ for $i \neq j$, and are normalized by $\hat{v}_i'S_{ww.z}\hat{v}_i = 1$. The calculations described here are called a reduced rank regression and are denoted by $RRR(u_t, w_t|z_t)$.

4.3 Maximum likelihood estimation in the $I(1)$ model and derivation of the rank test

Consider the $I(1)$ model given by equation (2). Note that the multiplier $\alpha\Upsilon$ of D_t is restricted to be proportional to α so that, by the Granger Representation Theorem, D_t does not cumulate in the process. It is assumed for the derivations of maximum likelihood estimators and likelihood ratio tests that ε_t is i.i.d. $N_p(0, \Omega)$, but for asymptotic results the Gaussian assumption is not needed. The Gaussian likelihood function shows that the maximum likelihood estimator can be found by the reduced rank regression

$$RRR(\Delta x_t, (x_{t-1}', D_t')'|\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t).$$

It is convenient to introduce the notation for residuals

$$\begin{aligned}
R_{0t} &= (\Delta x_t|\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t)\\
R_{1t} &= ((x_{t-1}', D_t')'|\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t)
\end{aligned}$$

and product moments

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt}.$$

The estimates are given by (15), and the maximized likelihood is, apart from a constant, given by

$$L_{\max}^{-2/T} = |\hat{\Omega}| = |S_{00}|\prod_{i=1}^{r}(1-\hat{\lambda}_i). \tag{16}$$

Note that all the models $H(r)$, $r = 0, \ldots, p$, have been solved by the same eigenvalue calculation. The maximized likelihood is given for each r by (16) and by dividing the maximized likelihood function for r with the corresponding expression for $r = p$, the likelihood ratio test for cointegration rank is obtained:

$$-2\log LR(H(r)|H(p)) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i). \tag{17}$$

This statistic was considered by Bartlett [5] for testing canonical correlations. The asymptotic distribution of this test statistic and the estimators are discussed in section 5.

The model obtained under the hypothesis $\beta = H\varphi$, is analyzed by

$$RRR(\Delta x_t, (H'x'_{t-1}, D'_t)'|\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t),$$

and a number of hypotheses of this type for β and α can be solved in the same way, but the more general hypothesis

$$\beta = (h_1 + H_1\varphi_1, \ldots, h_r + H_r\varphi_r),$$

cannot be solved by reduced rank regression. With $\alpha = (\alpha_1, \ldots, \alpha_r)$ and $\Upsilon = (\Upsilon'_1, \ldots, \Upsilon'_r)'$, equation (2) becomes

$$\Delta x_t = \sum_{j=1}^{r}\alpha_j((h_j + H_j\varphi_j)'x_{t-1} + \Upsilon_j D_t) + \sum_{i=1}^{k-1}\Gamma_i\Delta x_{t-i} + \Phi d_t + \varepsilon_t.$$

This is reduced rank regression, but there are r reduced rank matrices $\alpha_j(1, \varphi'_j, \Upsilon_j)$ of rank one. The solution is not given by an eigenvalue problem, but there is a simple modification of the reduced rank algorithm, which is easy to implement and is quite often found to converge. The algorithm has the property that the likelihood function is maximized in each step. The algorithm switches between reduced rank regressions of Δx_t on $(x'_{t-1}(H_i, h_i), D'_t)'$ corrected for

$$(((h_j + H_j\varphi_j)'x_{t-1} + \Upsilon_j D_t)_{j\neq i}, \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t).$$

This result can immediately be applied to calculate likelihood ratio tests for many different restrictions on the coefficients of the cointegration relations. Thus, in particular, this can give a test of over-identifying restrictions.

5 Asymptotic analysis

A discussion of the most important aspects of the asymptotic analysis of the cointegration model is given. This includes the result that the rank test requires a family of Dickey-Fuller type distributions, depending on the specification of the deterministic terms of the model. The asymptotic distribution of $\hat{\beta}$ is mixed Gaussian and that of the remaining parameters is Gaussian, so that tests for hypotheses on the parameters are asymptotically distributed as χ^2. All results are taken from Johansen [21]

5.1 Asymptotic distribution of the rank test

The asymptotic distribution of the rank test is given in case the process has a linear trend.

Theorem 3. *Let ε_t be i.i.d. $(0, \Omega)$ and assume that $D_t = t$ and $d_t = 1$, in model* (2)*. Under the assumptions that the cointegration rank is r, the asymptotic distribution of the likelihood ratio test statistic* (17) *is*

$$-2logLR(H(r)|H(p)) \xrightarrow{d} tr\{\int_0^1 (dB)F'(\int_0^1 FF'du)^{-1} \int_0^1 F(dB)'\}, \qquad (18)$$

where F is defined by

$$F(u) = \left(\begin{matrix} B(u) \\ u \end{matrix} \middle| 1 \right),$$

and $B(u)$ is the $p - r$ dimensional standard Brownian motion.

The limit distribution is tabulated by simulating the distribution of the test of no cointegration in the model for a $p - r$ dimensional model with one lag and the same deterministic terms. Note that the limit distribution does not depend on the parameters $(\Gamma_1, \ldots, \Gamma_{k-1}, \Upsilon, \Phi, \Omega)$, but only on $p-r$, the number of common trends, and the presence of the linear trend. For finite samples, however, the dependence on the parameters can be quite pronounced. A small sample correction for the test has been given in [24], and the bootstrap has been investigated by Swensen [47].

In the model without deterministics the same result holds, but with $F(u) = B(u)$. A special case of this, for $p = 1$, is the Dickey-Fuller test and the distributions (18) are called the Dickey–Fuller distributions with $p-r$ degrees of freedom; see [9].

The asymptotic distribution of the test statistic for rank depends on the deterministic terms in the model. It follows from the Granger Representation Theorem that the deterministic term d_t is cumulated to $C\Phi \sum_{i=1}^{t} d_i$. In deriving the asymptotics, x_t is normalized by $T^{-1/2}$. If $\sum_{i=1}^{t} d_i$ is bounded, this normalization implies that the limit distribution does not depend on the precise form of $\sum_{i=1}^{t} d_i$. Thus, if d_t is a centered seasonal dummy, or an 'innovation dummy' $d_t = 1_{\{t=t_0\}}$, it does not change the asymptotic distribution.

If, on the other hand, a 'step dummy' $d_t = 1_{\{t \geq t_0\}}$ is included, then the cumulation of this is a broken linear trend, and that influences the limit distribution and requires special tables; see [29].

5.2 Asymptotic distribution of the estimators

The main result here is that the estimator of β, suitably normalized, converges to a mixed Gaussian distribution, even when estimated under continuously differentiable restrictions [20]. This result implies that likelihood ratio tests on β are asymptotically χ^2 distributed. Furthermore the estimators of the adjustment parameters α and the short-run parameters Γ_i are asymptotically Gaussian and asymptotically independent of the estimator for β.

In order to illustrate these results, the asymptotic distribution of $\hat{\beta}$ for $r = 2$ is given, when β is identified by

$$\beta = (h_1 + H_1\varphi_1, h_2 + H_2\varphi_2). \tag{19}$$

Theorem 4. *In model* (2) *without deterministic terms and* ε_t *i.i.d.* $(0, \Omega)$, *the asymptotic distribution of* $T vec(\hat{\beta} - \beta)$ *is given by*

$$\begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} \rho_{11} H_1' \mathcal{G} H_1 & \rho_{12} H_1' \mathcal{G} H_2 \\ \rho_{21} H_2' \mathcal{G} H_1 & \rho_{22} H_2' \mathcal{G} H_2 \end{pmatrix}^{-1} \begin{pmatrix} H_1' \int_0^1 G(dV_1) \\ H_2' \int_0^1 G(dV_2) \end{pmatrix}, \tag{20}$$

where

$$\begin{aligned} T^{-1/2} x_{[Tu]} &\xrightarrow{d} G = CW, \\ T^{-1} S_{11} &\xrightarrow{d} \mathcal{G} = C \int_0^1 WW' du C', \end{aligned}$$

and

$$\begin{aligned} V &= \alpha' \Omega^{-1} W = (V_1, V_2)', \\ \rho_{ij} &= \alpha_i' \Omega^{-1} \alpha_j. \end{aligned}$$

The estimators of the remaining parameters are asymptotically Gaussian and asymptotically independent of $\hat{\beta}$.

Note that G and V are independent Brownian motions so that the limit distribution is mixed Gaussian and the asymptotic conditional distribution given G is Gaussian with asymptotic conditional variance

$$\begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} \rho_{11} H_1' \mathcal{G} H_1 & \rho_{12} H_1' \mathcal{G} H_2 \\ \rho_{21} H_2' \mathcal{G} H_1 & \rho_{22} H_2' \mathcal{G} H_2 \end{pmatrix}^{-1} \begin{pmatrix} H_1' & 0 \\ 0 & H_2' \end{pmatrix}.$$

A consistent estimator for the asymptotic conditional variance is

$$T \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \begin{pmatrix} \hat{\rho}_{11} H_1' S_{11} H_1 & \hat{\rho}_{12} H_1' S_{11} H_2 \\ \hat{\rho}_{21} H_1' S_{11} H_2 & \hat{\rho}_{22} H_2' S_{11} H_2 \end{pmatrix}^{-1} \begin{pmatrix} H_1' & 0 \\ 0 & H_2' \end{pmatrix}. \tag{21}$$

In order to interpret these results, note that the observed information about β in the data (keeping other parameters fixed) is given by

$$\mathcal{J}_T = T \begin{pmatrix} \rho_{11} H_1' S_{11} H_1 & \rho_{12} H_1' S_{11} H_2 \\ \rho_{21} H_2' S_{11} H_1 & \rho_{22} H_2' S_{11} H_2 \end{pmatrix},$$

which normalized by T^2 converges to the stochastic limit

$$\mathcal{J} = \begin{pmatrix} \rho_{11} H_1' \mathcal{G} H_1 & \rho_{12} H_1' \mathcal{G} H_2 \\ \rho_{21} H_2' \mathcal{G} H_1 & \rho_{22} H_2' \mathcal{G} H_2 \end{pmatrix}.$$

Thus the result (20) states that, given the asymptotic information or equivalently the limit of the common trends, $\alpha_\perp' W$, the limit distribution of $T(\hat{\beta} - \beta)$ is Gaussian with a variance that is a function of the inverse limit information. Hence the asymptotic distribution of

$$\mathcal{J}_T^{1/2} \begin{pmatrix} \bar{H}_1'(\hat{\beta}_1 - \beta_1) \\ \bar{H}_2'(\hat{\beta}_2 - \beta_2) \end{pmatrix}$$

is a standard Gaussian distribution. Here $\bar{H}_i' = (H_i' H_i)^{-1} H_i'$. This implies that Wald and therefore likelihood ratio tests on β can be conducted using the asymptotic χ^2 distribution.

It is therefore possible to scale the deviations $\hat{\beta} - \beta$ in order to obtain an asymptotic Gaussian distribution. Note that the scaling matrix $\mathcal{J}_T^{1/2}$ is not an estimate of the asymptotic variance of $\hat{\beta}$, but an estimate of the asymptotic *conditional* variance given the information in the data. It is therefore not the asymptotic distribution of $\hat{\beta}$ that is used for inference, but the *conditional* distribution given the information; see Basawa and Scott [6] or [19] for a discussion. Finally the result on the likelihood ratio test for the restrictions given in (19) is formulated.

Theorem 5. *Let ε_t be i.i.d.* $(0, \Omega)$*. The asymptotic distribution of the likelihood ratio test statistic for the restrictions* (19) *in model* (2) *with no deterministic terms is χ^2 with degrees of freedom given by $\sum_{i=1}^r (p - r - s_i + 1)$.*

This result is taken from [21], and a small sample correction for some tests on β has been developed in [23].

6 Further topics in the area of cointegration

It is mentioned here how the $I(1)$ model can be applied to test hypotheses implied by rational expectations. The basic model for $I(1)$ processes can be extended to other models of nonstationarity. In particular models for seasonal roots, explosive roots, $I(2)$ processes, fractionally integrated processes and nonlinear cointegration. We discuss here models for $I(2)$ processes, and refer to the paper by Lange and Rahbek [33] for some models of nonlinear cointegration.

6.1 Rational expectations

Many economic models operate with the concept of rational or model based expectations; see Hansen and Sargent [15]. An example of such a formulation is uncovered interest parity,

$$\Delta^e e_{t+1} = i_t^1 - i_t^2, \tag{22}$$

which expresses a balance between interest rates in two countries and economic expectations of exchange rate changes. If a vector autoregressive model

$$\Delta x_t = \alpha\beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_t, \tag{23}$$

fits the data $x_t = (e_t, i_t^1, i_t^2)'$, the assumption of model based expectations, Muth [35], means that $\Delta^e e_{t+1}$ can be replaced by the conditional expectation $E_t \Delta e_{t+1}$ based upon model (23). That is,

$$\Delta^e e_{t+1} = E_t \Delta e_{t+1} = \alpha_1 \beta' x_t + \Gamma_{11} \Delta x_t.$$

Assumption (22) implies the identity

$$i_t^1 - i_t^2 = \alpha_1 \beta' x_t + \Gamma_{11} \Delta x_t.$$

Hence the cointegration relation is

$$\beta' x_t = i_t^1 - i_t^2,$$

and the other parameters are restricted by $\alpha_1 = 1$, and $\Gamma_{11} = 0$. Thus, the hypothesis (22) implies a number of testable restrictions on the vector autoregressive model. The implications of model based expectations for the cointegrated vector autoregressive model is explored in [30], where it is shown that, as in the example above, rational expectation restrictions assume testable information on cointegration relations and short-run adjustments. It is demonstrated how estimation under rational expectation restrictions can be performed by regression and reduced rank regression in certain cases.

6.2 The $I(2)$ model

It is sometimes found that inflation rates are best described by $I(1)$ processes and then log prices are $I(2)$. In such a case $\alpha'_\perp \Gamma \beta_\perp$ has reduced rank; see (5). Under this condition model (2) can be parametrized as

$$\Delta^2 x_t = \alpha(\beta' x_{t-1} + \psi' \Delta x_{t-1}) + \Omega\alpha_\perp(\alpha'_\perp \Omega \alpha_\perp)^{-1} \kappa' \tau' \Delta x_{t-1} + \varepsilon_t, \tag{24}$$

where α and β are $p \times r$ and τ is $p \times (r+s)$, or equivalently as

$$\Delta^2 x_t = \alpha \begin{pmatrix} \beta \\ \delta' \end{pmatrix}' \begin{pmatrix} x_{t-1} \\ \bar{\tau}'_\perp \Delta x_{t-1} \end{pmatrix} + \zeta \tau' \Delta x_{t-1} + \varepsilon_t, \tag{25}$$

where

$$\delta = \psi'\tau_\perp,\ \zeta = \alpha\psi'\bar{\tau} + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\kappa';$$

see [22] and Paruolo and Rahbek [39]. Under suitable conditions on the parameters, the solution of equations (24) or (25) has the form

$$x_t = C_2\sum_{i=1}^{t}\sum_{j=1}^{i}\varepsilon_j + C_1\sum_{i=1}^{t}\varepsilon_i + A_1 + tA_2 + y_t,$$

where y_t is stationary and C_1 and C_2 are functions of the model parameters. One can prove that the processes $\Delta^2 x_t$, $\beta' x_t + \psi'\Delta x_t$, and $\tau'\Delta x_t$ are stationary. Thus $\tau' x_t$ are cointegration relations from $I(2)$ to $I(1)$. The model also allows for multicointegration, that is, cointegration between levels and differences because $\beta' x_t + \psi'\Delta x_t$ is stationary; see Engle and Yoo [12]. Maximum likelihood estimation can be performed by a switching algorithm using the two parametrizations given in (24) and (25). The same techniques can be used for a number of hypotheses on the cointegration parameters β and τ.

The asymptotic theory of likelihood ratio tests and maximum likelihood estimators is developed by Johansen [22, 28], Rahbek, Kongsted, and Jørgensen [44], Paruolo [38, 40], Boswijk [7] and Nielsen and Rahbek [36]. It is shown that the likelihood ratio test for rank involves not only Brownian motion, but also integrated Brownian motion and hence some new Dickey-Fuller type distributions that have to be simulated. The asymptotic distribution of the maximum likelihood estimator is quite involved, as it is not mixed Gaussian, but many hypotheses still allow asymptotic χ^2 inference; see [28].

7 Concluding remarks

What has been developed for the cointegrated vector autoregressive model is a set of useful tools for the analysis of macroeconomic and financial time series. The theory is part of many textbooks, and software for the analysis of data has been implemented in several packages, e.g. in CATS in RATS, Givewin, Eviews, Microfit, Shazam, R, etc.

Many theoretical problems remain unsolved, however. We mention here three important problems for future development.

1. The analysis of models for time series strongly relies on asymptotic methods, and it is often a problem to obtain sufficiently long series in economics which actually measure the same variables for the whole period. Therefore periods which can be modelled by constant parameters are often rather short, and it is therefore extremely important to develop methods for small sample correction of the asymptotic results. Such methods can be analytic or simulation based. When these will become part of the software packages, and are routinely applied, they will ensure more reliable inference.

2. A very interesting and promising development lies in the analysis of cointegration in nonlinear time series, where the statistical theory is still in its

beginning. Many different types of nonlinearities are possible, and the theory has to be developed in close contact with applications in order to ensure that useful models and concepts are developed; see the overview [33].

3. Most importantly, however, is the development of an economic theory which takes into account the findings of empirical analyses of nonstationary economic data. For a long time, regression analysis and correlations have been standard tools for quantitative analysis of relations between variables in economics. Economic theory has incorporated these techniques in order to learn from data. In the same way economic theory should be developed to incorporate nonstationarity of data and develop theories consistent with the findings of empirical cointegration analyses.

References

1. Ahn SK, Reinsel GC (1990) Estimation for partially nonstationary multivariate autoregressive models. Journal of the American Statistical Association 85:813–823
2. Anderson TW (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22:327–351
3. Anderson TW (1971) The statistical analysis of time series. Wiley, New York
4. Banerjee A, Dolado JJ, Galbraith JW, Hendry DF (1993) Co-integration error-correction and the econometric analysis of nonstationary data. Oxford University Press, Oxford
5. Bartlett M (1948) A note on the statistical estimation of the demand and supply relations from time series. Econometrica 16:323–329
6. Basawa IV, Scott DJ (1983) Asymptotic optimal inference for non-ergodic models. Springer, New York
7. Boswijk P (2000) Mixed normality and ancillarity in $I(2)$ systems. Econometric Theory 16:878–904
8. Campbell J, Shiller RJ (1987) Cointegration and tests of present value models. Journal of Political Economy 95:1062–1088
9. Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49:1057–1072
10. Engle RF, Granger CWJ (1987) Co-integration and error correction: Representation, estimation and testing. Econometrica 55:251–276
11. Engle RF, Hendry DF, Richard J-F (1983) Exogeneity. Econometrica 51:277–304
12. Engle RF, Yoo BS (1991) Cointegrated economic time series: A survey with new results. In: Granger CWJ, Engle RF (eds) Long-run economic relations. Readings in cointegration. Oxford University Press, Oxford
13. Granger CWJ (1983) Cointegrated variables and error correction models. UCSD Discussion paper 83-13a
14. Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton New Jersey
15. Hansen LP, Sargent TJ (1991) Exact linear rational expectations models: Specification and estimation. In: Hansen LP, Sargent TJ (eds) Rational expectations econometrics. Westview Press, Boulder

16. Hansen H, Johansen S (1999) Some tests for parameter constancy in the cointegrated VAR. The Econometrics Journal 2:306–333
17. Hendry DF (1995) Dynamic econometrics. Oxford University Press, Oxford
18. Johansen S (1988) Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12:231–254
19. Johansen S (1995) The role of ancillarity in inference for nonstationary variables. Economic Journal 13:302–320
20. Johansen S (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59:1551–1580
21. Johansen S (1996) Likelihood-based inference in cointegrated vector autoregressive models. Oxford University Press, Oxford
22. Johansen S (1997) Likelihood analysis of the $I(2)$ model. Scandinavian Journal of Statistics 24:433–462
23. Johansen S (2000) A Bartlett correction factor for tests on the cointegration relations. Econometric Theory 16:740–778
24. Johansen S (2002) A small sample correction of the test for cointegration rank in the vector autoregressive model. Econometrica 70:1929–1961
25. Johansen S (2005) The interpretation of cointegration coefficients in the cointegrated vector autoregressive model. Oxford Bulletin of Economics and Statistics 67:93–104
26. Johansen S (2006a) Cointegration: a survey. In Mills TC and Patterson K (eds) Palgrave handbook of econometrics: Volume 1, Econometric theory. Palgrave Macmillan, Basingstoke
27. Johansen S (2006b) Representation of cointegrated autoregressive processes with application to fractional processes. Forthcoming in Econometric Reviews
28. Johansen S (2006c) Statistical analysis of hypotheses on the cointegration relations in the $I(2)$ model. Journal of Econometrics 132:81–115
29. Johansen S, Mosconi R, Nielsen B (2000) Cointegration analysis in the presence of structural breaks in the deterministic trend. The Econometrics Journal 3:1–34
30. Johansen S, Swensen AR (2004) More on testing exact rational expectations in vector autoregressive models: Restricted drift term. The Econometrics Journal 7:389–397
31. Juselius K, MacDonald R (2004) The international parities between USA and Japan. Japan and the World Economy 16:17–34
32. Juselius K (2006) The cointegrated VAR model: Econometric methodology and macroeconomic applications. Oxford University Press, Oxford
33. Lange T, Rahbek A (2006) An introduction to regime switching time series models. This Volume
34. Lütkepohl H (2006) Introduction to multiple times series analysis. Springer, New York
35. Muth JF (1961) Rational expectations and the theory of price movements. Econometrica 29:315–335
36. Nielsen HB, Rahbek A (2004) Likelihood ratio testing for cointegration ranks in $I(2)$ Models. Forthcoming Econometric Theory
37. Nyblom J, Harvey A (2000) Tests of common stochastic trends. Econometric Theory 16:176–199
38. Paruolo P (1996) On the determination of integration indices in $I(2)$ systems. Journal of Econometrics 72:313–356
39. Paruolo P, Rahbek A (1999) Weak exogeneity in $I(2)$ VAR systems. Journal of Econometrics 93:281–308

40. Paruolo P (2000) Asymptotic efficiency of the two stage estimator in $I(2)$ systems. Econometric Theory 16:524–550
41. Phillips PCB (1991) Optimal inference in cointegrated systems. Econometrica 59:283–306
42. Phillips PCB, Hansen BE (1990) Statistical inference on instrumental variables regression with $I(1)$ processes. Review of Economic Studies 57:99–124
43. Proietti T (1997) Short-run dynamics in cointegrated systems. Oxford Bulletin of Economics and Statistics 59:405–422
44. Rahbek A, Kongsted HC, Jørgensen C (1999) Trend-stationarity in the $I(2)$ cointegration model. Journal of Econometrics 90:265–289
45. Seo B (1998) Tests for structural change in cointegrated systems. Econometric Theory 14:222–259
46. Stock JH (1987) Asymptotic properties of least squares estimates of cointegration vectors. Econometrica 55:1035–1056
47. Swensen AR (2006) Bootstrap algorithms for testing and determining the cointegrating rank in VAR models. Forthcoming Econometric Theory
48. Watson M (1994) Vector autoregressions and cointegration. In: Engle RF, McFadden D (eds) Handbook of econometrics Vol. 4. North Holland Publishing Company, Netherlands