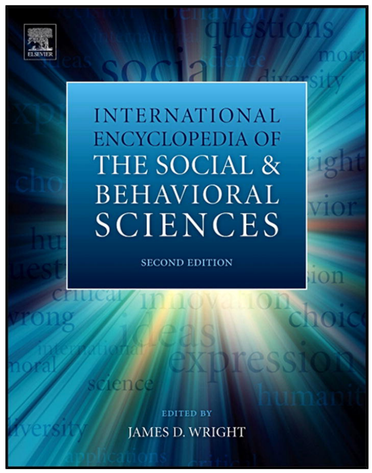

From Johansen, S., 2015. Time Series: Cointegration. In: James D. Wright (editor-in-chief), International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Vol 24. Oxford: Elsevier. pp. 322–330.
ISBN: 9780080970868

Time Series: Cointegration

Søren Johansen, University of Copenhagen, Copenhagen, Denmark; and Aarhus University, Aarhus, Denmark

This article is a revision of the previous edition article by N.H. Chan, volume 23, pp. 15709–15714, © 2001, Elsevier Ltd.

Abstract

An overview of results for the cointegrated vector autoregressive model for nonstationary $I(1)$ variables is given. The emphasis is on the analysis of the model and the tools for asymptotic inference. These include: formulation of criteria on the parameters, for the process to be nonstationary and $I(1)$, formulation of hypotheses of interest on the rank, the cointegrating relations and the adjustment coefficients. A discussion of the asymptotic distribution results that are used for inference. The results are illustrated by a few examples. A number of extensions of the theory are pointed out.

Introduction

The term cointegration was defined by Granger (1983) 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 (1987).

During the last 30 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; see Johansen (1996). 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 before conducting inference.

We start with some examples of cointegration.

Example 1: As a simple economic example of the main idea in cointegration, consider 229 observations of US monthly interest rates in the period 1987:1 to 2006:1, which defines the period when Greenspan was the chairperson of the Federal Reserve System. The data are taken from IMF's financial database and consist of the 6 months treasury bill rate and the 3 years bond rate, denoted by i_{6m} and i_{3y}, respectively. In Figure 1 we plot the two interests rates and their spread $i_{6m} - i_{3y}$. The expectations hypothesis implies that these interest rates should be equal up to a constant, $i_{3y} = i_{6m} + c$, and such a relation is not found in data. We can formulate it instead as their spread being stationary around a constant, possibly zero. This is an example of the formulation of an economic regularity as a cointegrating relation and we want below to analyze a statistical model, which allows such a formulation.

As simple examples of models for processes of this nature, we first consider a model for a random walk and a stationary process. Throughout we consider the sequence of p-dimensional errors ε_t, $t = 1,\ldots, T$, which are independent and identically distributed with mean zero and variance matrix, Ω.

Example 2: (Autoregressive processes) (Figure 2) Let $x_t = (y_t, z_t)'$ be given by the equations for $t = 1,\ldots,T$

$$y_t = y_{t-1} + \varepsilon_{yt} \qquad [1]$$

$$z_t = \rho z_{t-1} + \varepsilon_{zt} \qquad [2]$$

here $-1 < \rho < 1$. It is seen that $y_t = y_0 + \varepsilon_{y1} + \ldots + \varepsilon_{yt}$ and that $E(y_t|y_0) = y_0$ and $Var(y_t|y_0) = t\Omega$, so the variance is increasing and the process is nonstationary. We also find $z_t = \rho^t z_0 + \sum_{i=0}^{t-1}\rho^i\varepsilon_{z,t-i}$ which implies that $E(z_t|z_0) = \rho^t z_0$ and $Var(z_t|z_0) = \sigma_z^2\sum_{i=0}^{t-1}\rho^{2i} \to \sigma_z^2\sum_{i=0}^{\infty}\rho^{2i} = \sigma_z^2/(1-\rho^2)$. We can make z_t stationary by choosing $z_0 = \sum_{i=0}^{\infty}\rho^i\varepsilon_{z,-i}$ and then $z_t = \sum_{i=0}^{\infty}\rho^i\varepsilon_{z,t-i}$. We call y_t an $I(1)$ process and z_t an $I(0)$ process, see section Integration and Cointegration.

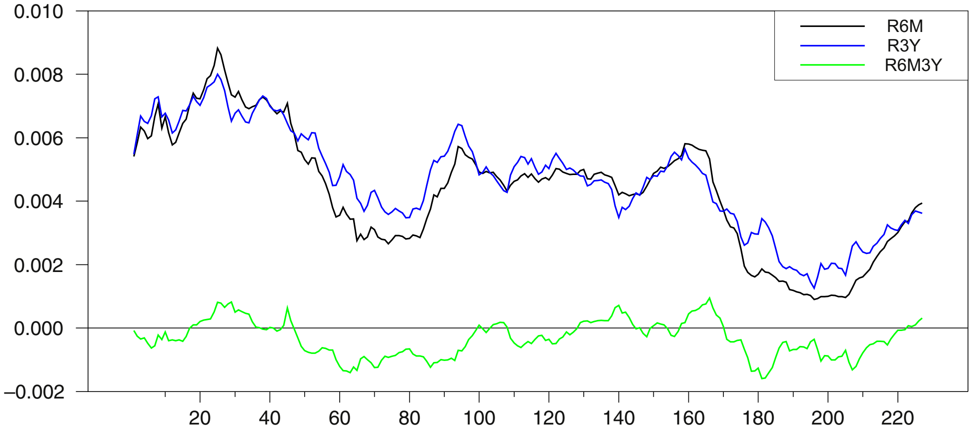

Figure 1 A plot of 229 monthly observations of the 6 months treasury bill rate and the 3 years bond rate and their spread $i_{6m} - i_{3y}$. Note the nonstationary behavior of the interest rates and the much more stationary spread, see Example 1.

International Encyclopedia of the Social & Behavioral Sciences, Second Edition, 2015, 322–330

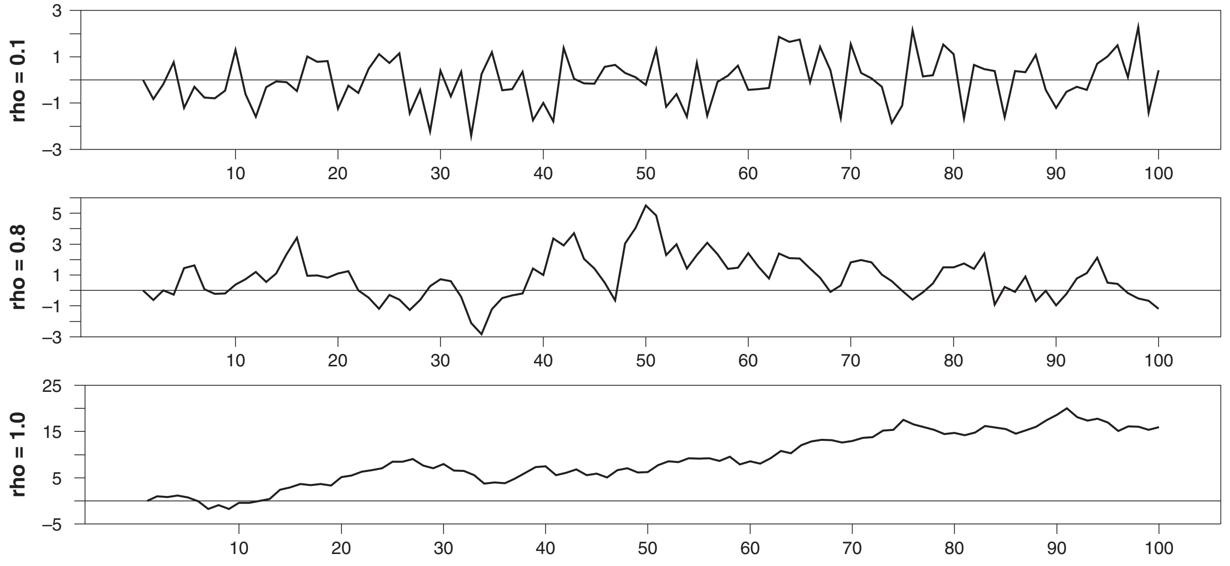

Figure 2 Three examples of *AR*(1) processes. Each plot has 100 observations. The first two are stationary with $\rho = 0.1$ and $\rho = 0.5$, and the last is a random walk with $\rho = 1$, see Example 2.

Next we give a model for nonstationary variables that are cointegrated, using the notation, $\Delta x_t = x_t - x_{t-1}$.

Example 3: (Cointegrated processes) A bivariate process is given for $t = 1,\ldots,T$ by the equations

$$\Delta x_{1t} = \alpha_1(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t} \quad [3]$$
$$\Delta x_{2t} = \alpha_2(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t}$$

Subtracting the equations, we find that the process $y_t = x_{1t} - x_{2t}$ is autoregressive, see [2] and stationary if $\rho = 1 + \alpha_1 - \alpha_2$ satisfies $|\rho| < 1$, 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, see [1], so that the process $x_t = (x_{1t}, x_{2t})'$ is given by

$$x_{1t} = \frac{s_t - \alpha_1 y_t}{\alpha_2 - \alpha_1} \quad \text{and} \quad x_{2t} = \frac{s_t - \alpha_2 y_t}{\alpha_2 - \alpha_1}$$

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 $\sum_{i=1}^{t}(\alpha_2\varepsilon_{1i} - \alpha_1\varepsilon_{2i})$, 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 (Figure 3).

Three Approaches to Cointegration

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

Regression Formulation

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

$$x_{1t} = \gamma' x_{2t} + \varepsilon_t$$
$$\Delta x_{2t} = \varepsilon_{2t}$$

This model implies that x_{2t} is a nonstationary random walk, and $x_{1t} - \gamma' x_{2t}$ gives p_1 stationary linear combinations. Hence, in this case, the cointegration rank of x_t is p_1, see section Integration and Cointegration. The first estimation method used in this model is least squares regression, Engle and Granger (1987), which is shown to give a superconsistent estimator by Stock (1987). This estimation method gives rise to residual based tests for cointegration. It was shown by Phillips and Hansen (1990) that, for a more general error term, a modification of the regression estimator gives useful methods for inference on coefficients of cointegration relations; see also Phillips (1991).

Autoregressive Formulation

In the rest of this contribution we focus on the autoregressive formulation of the p-dimensional process, x_t, defined by the equations

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

where α and β are $p \times r$ matrices of rank r. Under the condition that Δx_t is stationary, the solution is

$$x_t = C\sum_{i=1}^{t}\varepsilon_i + \sum_{i=0}^{\infty} C_i\varepsilon_{t-i} + A \quad [4]$$

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 coefficient α, toward the attractor set $\{x : \beta' x = 0\}$ defined by the long-run relations. Models for different cointegration ranks are nested and the smallest, for $\alpha = \beta = 0$, corresponds to p random walks. The rank can be analyzed by likelihood ratio tests. Methods usually applied for this analysis are derived from the Gaussian likelihood function, which is discussed here; see also Johansen (1988, 1996), and Ahn and Reinsel (1990).

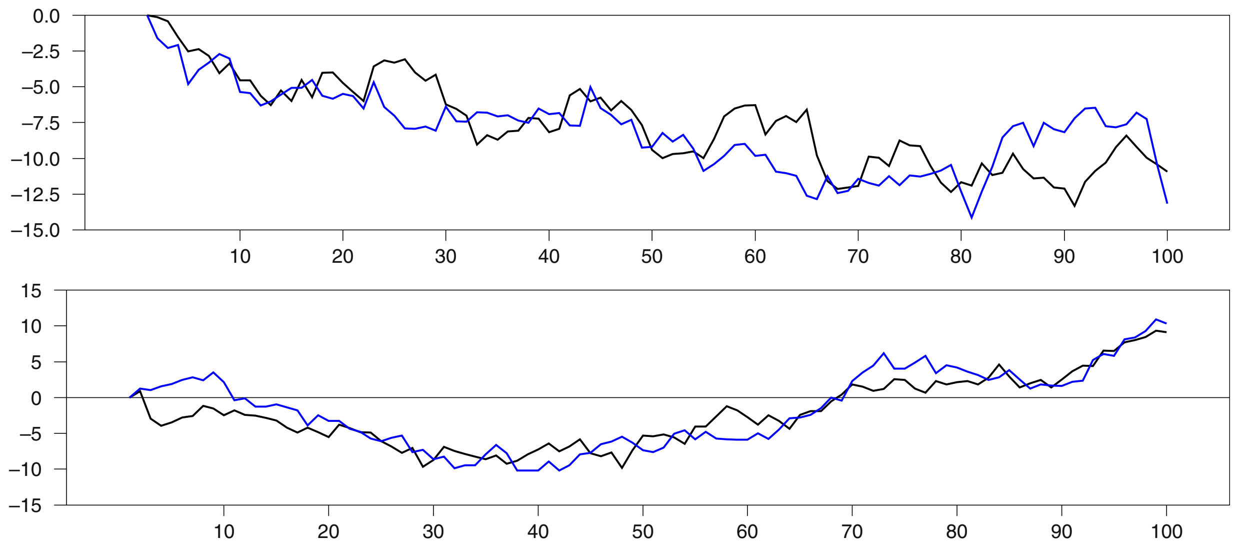

Figure 3 Two simulations of model [3] for cointegrated variables, see Example 3.

Unobserved Components Formulation

Let x_t be given by

$$x_t = \xi\eta' \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{2t}$$

where ε_{2t}, typically, is independent of ε_{1t}.

In this formulation too, hypotheses of different ranks are nested but in the opposite direction, and the smallest, for $\xi = \eta = 0$, corresponds to stationary processes. The parameters are linked to the autoregressive formulation by $\xi = \beta_\perp$ and $\eta = \alpha_\perp$, even though the linear process, $\sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$, in [4] depends on the random walk part, so the unobserved components model and the autoregressive model are not the same. However, 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 performed by the Kalman filter, and asymptotic theory of the rank tests has been worked out by Nyblom and Harvey (2000), see also Durbin and Koopman (2012).

The Model Analyzed in This Contribution

In the following we consider cointegration modeled by the cointegrated vector autoregressive (CVAR) model, $H(r)$, for the p-dimensional process x_t,

$$H(r) : \Delta x_t = \alpha(\beta' x_{t-1} + \gamma D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t \quad [5]$$

The terms 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 $0 \le r \le p$. In section Integration and Cointegration, conditions for the processes $\beta' x_t$ and Δx_t to be stationary around their means are given, and model [5] can then 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 (1994) and Stock (1994) or Johansen (2006, 2009), where the last two form the basis for the presentation here. The topic has become part of most textbooks in econometrics; see among others Banerjee et al. (1993), Hamilton (1994), Hendry (1995), and Lütkepohl (2006). For a general account of the methodology of the CVAR model with applications to the analysis of economic data, see Juselius (2006).

Linear Stationary Processes

We consider p-dimensional linear stationary processes

$$z_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$$

which are well defined if the coefficient matrices satisfy the condition that $\sum_{i=0}^{\infty} tr^{1/2}(C_i' C_i) < \infty$, where the trace of a matrix, C, is $tr(C) = \sum_{i=1}^{p} C_{ii}$. If in [5] we consider $r = p$ we define the matrix $\Pi = \alpha\beta'$ and the matrix valued characteristic polynomial

$$\Psi(z) = (1-z)I_p - \Pi z - (1-z)\sum_{i=1}^{k-1} \Gamma_i z^i \quad [6]$$

with determinant $|\Psi(z)|$. The properties of the solution of [5] are determined by $\Psi(z)$. We define the roots $z_i, i = 1, \ldots, n$, as the solutions of $|\Psi(z)| = 0$, and get, because $|\Psi(0)| = 1$, that $|\Psi(z)| = \prod_{i=1}^{n}(1 - z/z_i)$. The inverse characteristic polynomial is given by

$$\Psi(z)^{-1} = \frac{\text{adj}(\Psi(z))}{\det\Psi(z)}, z \neq z_i$$

that is, the adjoint of $\Psi(z)$ divided by the determinant of $\Psi(z)$.

The function, $C(z) = \Psi(z)^{-1}$ has poles at the roots of the polynomial $|\Psi(z)|$, and the position of the poles determine the stochastic properties of the solution of [5]. We first mention a well-known result; see Anderson (1971).

Theorem 1: If the roots satisfy $|z_i| > 1$, then α and β have full rank $r = p$, 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 with moving average representation

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

Thus the exponentially decreasing coefficients are found by simply inverting the characteristic polynomial if the roots are outside the unit disk. The matrices, C_i contain the impulse response coefficient of the process in the sense that a shock at time zero to variable k will have the effect, $(C_t)_{ik}$ at time t to variable i.

Integration and Cointegration

The basic definitions of integration and cointegration are given together with a moving average representation of the solution of the error correction model [5]. This solution reveals the stochastic properties of the solution, see Example 3.

If the roots of $|\Psi(z)| = 0$ are not greater than 1, the equations generate nonstationary processes of various types, and the coefficients are not exponentially decreasing. Still, the coefficients of $C(z) = \Psi^{-1}(z)$ determine the stochastic properties of the solution of [5].

Definition 1: If $\sum_{i=0}^{\infty} tr^{1/2}(C_i' C_i) < \infty$, the linear process $x_t - E(x_t) = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$ is called $I(0)$ if $C(1) = \sum_{i=0}^{\infty} C_i \neq 0$ The process, x_t is called integrated of order 1, $I(1)$, if $\Delta x_t - E(\Delta x_t)$ is $I(0)$. 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.

We consider the process given by [5] and the characteristic polynomial, $\Psi(z)$ defined in [6]. This has a unit root, if $\Psi(1) = -\Pi$ is singular, and by Theorem 1, the process is not stationary. A singular matrix Π of rank r can be expressed as $\Pi = \alpha\beta'$, where α and β are $p \times r$. We next formulate a condition for the process to be $I(1)$. We define $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$.

Assumption 1: (The $I(1)$ condition) The $I(1)$ condition is satisfied if the roots, $|\Psi(z_i)| = 0$ satisfy $|z_i| > 1$ or $z_i = 1$ and it holds that

$$|\alpha_\perp' \Gamma \beta_\perp| \neq 0 \qquad [7]$$

Condition [7] is needed to avoid solutions that are integrated of order 2 or higher; see section Further Topics in Cointegration for references. For a process with one lag $\Gamma = I_p$, and [5] implies

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

In this case the $I(1)$ condition is equivalent to the condition that $\beta' x_t$ is stationary, that is, the absolute value of the eigenvalues of $I_r + \beta'\alpha$ are less than 1, and in Example 2 this condition reduces to $|1 + \alpha_1 - \alpha_2| < 1$.

Example 3 presents a special case of the Granger Representation Theorem, which gives the moving average representation of the solution of the error correction model.

Theorem 2: (The Granger Representation Theorem) If $\Psi(z)$ has unit roots and the $I(1)$ condition is satisfied, then

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

converges for $|z| \leq 1 + \delta$ for some $\delta > 0$. The matrix C is defined by

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

The solution x_t of [5] has the moving average representation

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

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

This result implies that Δx_t and $\beta' x_t$ are stationary around their mean, so that x_t is a cointegrated $I(1)$ process with r cointegration vectors β and $p - r$ common stochastic trends $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$.

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 $\beta' C\Phi = 0$.

Interpretation of the *I*(1) Model

In this section, model $H(r)$ defined by [5] 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 are given.

The Relation between the Models *H*(*r*)

The models $H(r)$ are nested

$$H(0) \subset \ldots \subset H(r) \subset \ldots \subset H(p)$$

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 of α and β 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.

Normalization of Parameters

The parameters α and β in [5] 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 $\alpha\xi(\beta\xi'^{-1})' = \alpha\xi\xi^{-1}\beta' = \alpha\beta'$.

If $x_t = (x'_{1t}, x'_{2t})'$ where x_{1t} is $r \times 1$ and x_{2t} is $(p - r) \times 1$, and $\beta = (\beta'_1, \beta'_2)'$, with β_1, $r \times r$, and $|\beta_1| \neq 0$, 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, see Section Regression Formulation. A normalization of this type is sometimes convenient for estimation and calculation of 'standard errors' of the estimator, see Section Asymptotic Distribution of Estimators, 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.

Similarly $\alpha_\perp$ and $\beta_\perp$ are not uniquely defined. 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, the processes, $\eta\alpha'_\perp \sum_{i=1}^{t}\varepsilon_i$ could also be used as common trends because

$$C = \beta_\perp\left(\alpha'_\perp \Gamma \beta_\perp\right)^{-1}\alpha'_\perp = \beta_\perp\left(\eta\alpha'_\perp \Gamma \beta_\perp\right)^{-1}\eta\alpha'_\perp$$

Thus identifying restrictions on the coefficients in $\alpha_\perp$ are needed to find their estimates and standard errors, and a similar result holds for $\beta_\perp$.

In the cointegration model there are therefore four separate identification problems: one for the cointegration relations, one for the common trends, one for $\beta_\perp$, and finally one for the short-run dynamics, if the model has simultaneous effects.

Hypotheses on Long-Run Coefficients

One purpose of modeling economic data is to test hypotheses on the coefficients, thereby investigating whether the data support an economic hypothesis or reject it. As an example consider the series $x_t = (e_t, p_t, p_t^*)'$, where p_t and p_t^* are the log price indices in two countries and e_t the exchange rate. The hypothesis of the law of one price, *PPP*, is that $e_t = p_t - p_t^*$. We formulate that as the hypothesis that (1, −1, 1) is a cointegration vector so that $e_t - p_t + p_t^*$ becomes stationary. Similarly, the hypothesis of price homogeneity is formulated as the restriction

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

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = H\varphi$$

where $H = R_\perp$. A general formulation of restrictions on each of r cointegration vectors, including a normalization, is

$$\beta = (h_1 + H_1\varphi_1, \ldots, h_r + H_r\varphi_r). \qquad [9]$$

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 the restrictions be $R_i = (h_i, H_i)_\perp$ then β_i satisfies $R'_i\beta_i = 0$, and the normalization $(h'_i h_i)^{-1} h'_i \beta_i = 1$. The usual rank condition for identification is that β_i is identified by $R'_i\beta_i = 0$, if

$$\operatorname{rank}\left(R'_i(\beta_1, \ldots, \beta_r)\right) = r - 1$$

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 et al. (1983). We decompose the process x_t as $x_t = (x'_{1t}, x'_{2t})'$ and the matrices are decomposed similarly so that the model equations (without deterministic terms and $k = 2$) become

$$\begin{aligned} \Delta x_{1t} &= \alpha_1\beta' x_{t-1} + \Gamma_{11}\Delta x_{t-1} + \varepsilon_{1t} \\ \Delta x_{2t} &= \alpha_2\beta' x_{t-1} + \Gamma_{21}\Delta x_{t-1} + \varepsilon_{2t} \end{aligned}$$

The conditional model for Δx_{1t} given Δx_{2t} and the past is

$$\Delta x_{1t} = \omega\Delta x_{2t} + \alpha_1\beta' x_{t-1} + (\Gamma_{11} - \omega\Gamma_{21})\Delta x_{t-1} + \varepsilon_{1t} - \omega\varepsilon_{2t} \qquad [10]$$

where $\omega = \Omega_{12}\Omega_{22}^{-1}$. 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 called weakly exogenous for α_1 and β. In this case the likelihood is a product of two factors depending on (Γ_{21}, Ω_{22}) and (α_1,β,Γ_{11},ω,$\Omega_{11.2}$), respectively. Because the parameters are unrestricted (variation independent), likelihood inference on β and α_1 can be conducted in the conditional model alone.

If the hypothesis of weak exogeneity is not satisfied, inference of the conditional model is complicated because limit distributions contain nuisance parameters, and asymptotic inference is not Gaussian.

If x_{2t} is weakly exogenous, $\alpha_\perp$ 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 Δx_{2t} cumulate in the system and give rise to nonstationarity.

Likelihood Analysis

This section contains first some comments on what aspects of the data are important for checking for model misspecification, and then describes the calculation of reduced rank regression, introduced by Anderson (1951). Then reduced rank regression and modifications thereof are applied to estimate the parameters of the $I(1)$ model [5] and various submodels defined by restrictions on β, see Johansen and Juselius (1990).

Checking for Specification

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 generated 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 independent identically distributed (iid) 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 (1998) and Hansen and Johansen (1999) contain some results on recursive tests in the cointegration model, and Doornik and Hendry (2013) contains a description of a general algorithm (*Autometrics*) for finding a model that describes the data.

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 (1951), can be described in the regression model

$$u_t = \alpha\beta' w_t + \Xi z_t + \varepsilon_t$$

where ε_t are iid $(0,\Omega)$. 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

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

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

For fixed β the regression estimates are

$$\widehat{\alpha}(\beta) = S_{uw.z}\beta(\beta' S_{ww.z}\beta)^{-1}$$
$$\widehat{\Omega}(\beta) = S_{uu.z} - S_{uw.z}\beta(\beta' S_{ww.z}\beta)^{-1}\beta' S_{wu.z}$$

so that

$$|\widehat{\Omega}(\beta)| = |S_{uu.w}|\frac{|\beta' S_{ww.uz}\beta|}{|\beta' S_{ww.z}\beta|}$$

Minimizing over β gives the reduced rank estimators. This minimization problem is solved as follows. First we solve the eigenvalue problem

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

The eigenvalues are ordered $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_{p_w}$ with corresponding eigenvectors $\widehat{v}_1, \ldots \widehat{v}_{p_w}$. The reduced rank estimate of β is

$$\widehat{\beta} = (\widehat{v}_1, \ldots, \widehat{v}_r) \qquad [11]$$

and the other estimators are found by regression of u_t on $\widehat{\beta}' x_{t-1}$ and z_t. Finally we find

$$|\widehat{\Omega}| = |S_{uu.z}|\prod_{i=1}^{r}\left(1 - \widehat{\lambda}_i\right)$$

The eigenvectors are orthogonal with respect to $S_{ww.z}$, that is, $\widehat{v}_i' S_{ww.z}\widehat{v}_j = 0$ for $i \neq j$, and they are normalized by $\widehat{v}_i' S_{ww.z}\widehat{v}_i = 1$. The calculations described here are called a reduced rank regression and are denoted by $RRR(u_t, w_t \mid z_t)$.

Likelihood Analysis

It is assumed for the likelihood analysis that ε_t is iid $N_p(0, \Omega)$, 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 of Δx_t on $(x'_{t-1}, D'_t)'$ correcting for $\mathbb{X}_t = (\Delta x'_{t-1}, \ldots, \Delta x'_{t-k+1}, d'_t)'$

$$RRR\left(\Delta x_t, \begin{pmatrix} x_{t-1} \\ D_t \end{pmatrix} | \mathbb{X}_t\right)$$

The estimates are given by [11], and the maximized likelihood is, apart from a constant, given by

$$L_{\max}^{-2/T} = |\widehat{\Omega}| = |S_{00}|\prod_{i=1}^{r}\left(1 - \widehat{\lambda}_i\right) \qquad [12]$$

where $S_{00} = T^{-1}\sum_{t=1}^{T}(\Delta X_t|\mathbb{X}_t)(\Delta X_t|\mathbb{X}_t)'$.

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 [12] 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:

$$-2\log LR(H(r)|H(p)) = -T\sum_{i=r+1}^{p}\log\left(1 - \widehat{\lambda}_i\right)$$

The asymptotic distribution of this test statistic and the estimators are discussed in section Asymptotic Analysis.

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

$$RRR\left(\Delta x_t, \begin{pmatrix} H' X_{t-1} \\ D_t \end{pmatrix} | \mathbb{X}_t\right)$$

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

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

see [9], cannot be solved by reduced rank regression.

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 $\widehat{\beta}$ 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.

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 iid $(0, \Omega)$ and assume that $D_t = t$ and $d_t = 1$, in model [5]. Under the assumptions that the cointegration rank is r, the asymptotic distribution of the likelihood ratio test statistic $-2\log LR(H(r)|H(p))$, [13], is

$$tr\left\{\int_0^1 (dB)F'\left(\int_0^1 FF'du\right)^{-1}\int_0^1 F(dB)'\right\} \qquad [13]$$

where F is defined by

$$F(u) = \begin{pmatrix} B(u) \\ u \end{pmatrix}\Bigg| 1 \Bigg)$$

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

The limit distribution is tabulated by simulation. Note that it 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 Johansen (2002), and the bootstrap has been investigated by Rahbek et al. (2012).

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 asymptotic distributions given in Theorem 3 are called the Dickey–Fuller distributions with $p - r$ degrees of freedom; see Dickey and Fuller (1981).

Asymptotic Distribution of Estimators

The main result here is that the estimator of β, suitably normalized, converges to a mixed Gaussian distribution; see Johansen (1988). 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.

To illustrate how to conduct inference on a cointegrating coefficient, and why it becomes asymptotic χ^2 despite the asymptotic mixed Gaussian limit of $\widehat{\beta}$, we consider an example.

Example 4: (Mixed Gaussian distribution) Let x_t be a bivariate process with one lag for which $\alpha = (-1, 0)'$ and $\beta = (1, \theta)'$. The equations become

$$\begin{aligned} x_{1t} &= \theta x_{2t-1} + \varepsilon_{1t} \\ \Delta x_{2t} &= \varepsilon_{2t} \end{aligned} \qquad [14]$$

This model as a special case of [5] with $\alpha' = (-1, 0), \beta' = (1, -\theta), p = 2, k = 1$. If we add the assumption that ε_t is Gaussian with mean zero and variance $\Omega = diag(\sigma_1^2, \sigma_2^2)$ the maximum likelihood estimator simplifies to a regression estimator, and becomes

$$\widehat{\theta} = \frac{\sum_{t=1}^T x_{1t}x_{2t-1}}{\sum_{t=1}^T x_{2t-1}^2} = \theta + \frac{\sum_{t=1}^T \varepsilon_{1t}x_{2t-1}}{\sum_{t=1}^T x_{2t-1}^2}$$

The distribution of $\widehat{\theta}$ conditional on the whole process $\{x_{2t}\}_{t=1}^T$ is clearly Gaussian:

$$\widehat{\theta}\Big|\{x_{2t}\}\text{is distributed as } N\left(\theta, \sigma_1\Big/\sum_{t=1}^T x_{2t-1}^2\right)$$

By integrating out the process x_{2t} we get a distribution which we call mixed Gaussian with mixing parameter $1/\sum_{t=1}^T x_{2t-1}^2$, and hence

$$E\left(\widehat{\theta}\right) = \theta \quad\text{and}\quad Var\left(\widehat{\theta}\right) = \sigma_1^2 E\left(1\Big/\sum_{t=1}^T x_{2t-1}^2\right)$$

When constructing a test for $\theta = \theta_0$ we do not use

$$\left(\widehat{\theta} - \theta\right)\Big/Var\left(\widehat{\theta}\right)^{1/2}$$

but instead expand the likelihood function and find the Wald test, which is based on the observed information:

$$\widehat{\sigma}_1^{-1}\left(\sum_{t=1}^T x_{2t-1}^2\right)^{1/2}\left(\widehat{\theta} - \theta\right)$$

This statistic is asymptotically distributed as $N(0, 1)$: Thus we normalize by the observed information, $\sum_{t=1}^T x_{2t-1}^2/\sigma_1^2$, not the expected information often used when analyzing stationary processes.

Figure 4 shows a scatter diagram of 1000 simulations of $(\widehat{\theta}, \sum_{t=1}^T x_{2t-1}^2/\sigma_1^2)$. That is, the estimator and the information about the parameter. We note that when the information is large, the variation of $\widehat{\theta}$ is small, and when the information is small, the variation of $\widehat{\theta}$ is much larger. Thus the variation of $\widehat{\theta}$ should be measured by its conditional variance, which is the reciprocal information in the data. This has the further advantage that if we only consider those estimates with a given information we see that $\widehat{\theta}$ is approximately Gaussian.

The main result is that tests on β are asymptotically distributed as χ^2, and we formulate that as.

Theorem 4: Let εt be iid $(0, \Omega)$. The asymptotic distribution of the likelihood ratio test statistic for the restrictions [9] in model [5] is distributed as χ^2 with degrees of freedom given by $\sum_{i=1}^r (p - r - s_i + 1)$. A small sample correction for some tests on β has been developed in Johansen (2000).

Further Topics in Cointegration

The basic model for $I(1)$ processes has been extended to other types of nonstationarity. In particular models for seasonal roots, Ahn and Reinsel (1994) and Johansen and Schaumburg (1998), explosive processes, Nielsen (2010), $I(2)$ processes, Johansen (1997), fractional processes, Johansen and Nielsen (2012), nonlinear processes, Lange and Rahbek (2009), panel data cointegration, Larsson et al. (2001) and Pesaran et al.

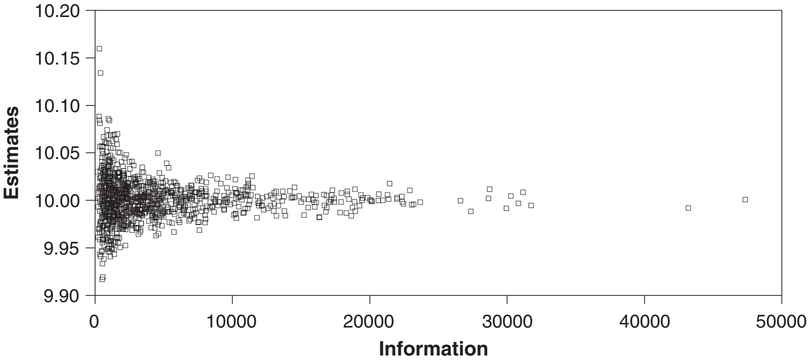

Figure 4 A scatter plot of $\widehat{\theta} = \sum_{i=1}^{T} x_{1t}x_{2t-1} / \sum_{i=1}^{T} x_{2,t-1}^2$ and the information. $\sum_{i=1}^{T} x_{2,t-1}^2/\sigma_1^2$ for model [14]. The number of observations is $T = 100$, and the number of simulations is 1000.

(2004), and finally applications to rational expectation models, Johansen and Swensen (2011).

Concluding Remarks

In summary one can say, that what has been developed for the CVAR model is a set of useful tools for the analysis of many types of economic time series. The theory is now 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, Gauss, GRETL, etc.

We have given a brief tour in the cointegration landscape showing some of the major sights without indicating, except by examples, how the formal analysis is conducted. We concluded with a list of extensions of the basic model, which shows that the ideas behind the CVAR extend to a large number of other models.

Acknowledgment

The author acknowledges the support of the Center for Research in Econometric Analysis of Time Series (CREATES – DNRF78, funded by the Danish National Research Foundation).

See also: Probability Theory: Formal; Time Series: General; Time Series: Unit-Roots and Nonstandard Limiting Distributions.

Bibliography

Ahn, S.K., Reinsel, G.C., 1990. Estimation for partially nonstationary multivariate autoregressive models. Journal of the American Statistical Association 85, 813–823.

Ahn, S.K., Reinsel, G.C., 1994. Estimation of partially non-stationary vector autoregressive models with seasonal behavior. Journal of Econometrics 62, 317–350.

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327–351.

Anderson, T.W., 1971. The Statistical Analysis of Time Series. Wiley, New York.

Banerjee, A., Dolado, J.J., Galbraith, J.W., Hendry, D.F., 1993. Co-integration Error-correction and the Econometric Analysis of Nonstationary Data. Oxford University Press, Oxford.

Dickey, D.A., Fuller, W.A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49, 1057–1072.

Doornik, J., Hendry, D.F., 2013. Empirical Model Discovery and Theory Evaluation. MIT Press, Cambridge.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55, 251–276.

Engle, R.F., Hendry, D.F., Richard, J.-F., 1983. Exogeneity. Econometrica 51, 277–304.

Granger, C.W.J., 1983. Cointegrated variables and error correction models. UCSD Discussion Paper, 83–13a.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, New Jersey, NJ.

Hansen, H., Johansen, S., 1999. Some tests for parameter constancy in the cointegrated VAR. The Econometrics Journal 2, 306–333.

Hendry, D.F., 1995. Dynamic Econometrics. Oxford University Press, Oxford.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1996. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.

Johansen, S., 1997. Likelihood analysis of the $I(2)$ model. Scandinavian Journal of Statistics 24, 433–462.

Johansen, S., 2000. A Bartlett correction factor for tests on the cointegration relations. Econometric Theory 16, 740–778.

Johansen, S., 2002. A small sample correction of the test for cointegration rank in the vector autoregressive model. Econometrica 70, 1929–1961.

Johansen, S., 2006. Cointegration, a survey. In: Mills, T.C., Patterson, K. (Eds.), Palgrave Handbook of Econometrics, vol. 1. Econometric Theory. Palgrave Macmillan, Basingstoke, pp. 671–693.

Johansen, S., 2009. Cointegration. Overview and development. In: Andersen, T.G., Davies, R., Kress, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer, New York, pp. 671–694.

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration — with applications to the demand for money. Oxford Bulletin of Economics and Statistics 52, 169–210.

Johansen, S., Nielsen, M.Ø., 2012. Likelihood inference for a fractionally cointegrated vector autoregressive model. Econometrica 80, 2667–2732.

Johansen, S., Schaumburg, E., 1998. Likelihood analysis of seasonal cointegration. Journal of Econometrics 88, 301–339.

Johansen, S., Swensen, A.R., 2011. On a numerical and graphical technique for evaluating some models involving rational expectations. Journal of Time Series Econometrics 3, Article 9. http://dx.doi.org/10.2202/1941-1928.1089.

Juselius, K., 2006. The Cointegrated VAR Model,Econometric Methodology and Macroeconomic Applications. Oxford University Press, Oxford.

Lange, T., Rahbek, A., 2009. An Introduction to regime switching time series models. In: Andersen, T.G., Davies, R., Kress, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer, New York, pp. 871–888.

Larsson, R.L., Lyhagen, J., Löthgren, M., 2001. Likelihood based cointegration tests in heterogenous panels. The Econometric Journal 4, 109–142.

Lütkepohl, H., 2006. New Introduction to Multiple Time Series Analysis. Springer, New York.

Nielsen, B., 2010. Analysis of explosive processes. Econometric Theory 26, 882–915.

Nyblom, J., Harvey, A., 2000. Tests of common stochastic trends. Econometric Theory 16, 176–199.

Pesaran, M.H., Schuermann, T., Weiner, S.M., 2004. Modeling regional interdependencies using a global error-correcting macroeconometric model. Journal of Business and Economic Statistics 22, 169–172.

Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. Econometrica 59, 283–306.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference on instrumental variables regression with $I(1)$ processes. Review of Economic Studies 57, 99–124.

Rahbek, A., Cavaliere, G., Taylor, A.M.R., 2012. Bootstrap determination of the cointegration rank in VAR models. Econometrica 80, 1721–1740.

Seo, B., 1998. Tests for structural change in cointegrated systems. Econometric Theory 14, 222–259.

Stock, J.H., 1987. Asymptotic properties of least squares estimates of cointegration vectors. Econometrica 55, 1035–1056.

Stock, J.H., 1994. Unit roots, structural breaks and trends. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV. Elsevier, Netherlands, pp. 2739–2841.

Watson, M., 1994. Vector autoregressions and cointegration. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. IV. Elsevier, Netherlands, pp. 871–888.