Cointegration: an overview

Søren Johansen
Department of Applied Mathematics and Statistics
University of Copenhagen

November 2004

Contents

1   Introduction and methodology

The phenomenon that non-stationary processes can have linear combinations that are stationary was called cointegration by Granger (1983), who used it for modelling long-run economic relations. The paper by Engle and Granger (1987), which showed the equivalence of the error correction formulation and the phenomenon of cointegration, started a rapid development of the statistical and probabilistic analysis of the ideas.

There are now three different ways of modelling the cointegration idea in a statistical framework. To illustrate the ideas we formulate these in the simplest possible case, leaving out deterministic terms.

1.1   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

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

where we assume that u_t is a linear invertible process defined by i.i.d. errors ε_t with mean zero and finite variance. The assumptions behind the model imply that x_{2t} is non-stationary and not cointegrating, and hence that the cointegrating rank, p_1, is known so that the models for different ranks are not nested. The first estimation method used in this model is least squares regression, see Engle and Granger (1987), which gives superconsistent estimators as shown by Stock (1987) and which gives rise to residual based tests for cointegration. It was shown by Phillips and Hansen (1990) and Park (1992) that a modification of the regression, involving a correction using the long-run variance of the process u_t, would give useful inference for the coefficients of the cointegrating relations, see also Phillips (1991).

1.2   The autoregressive formulation

In this case the process x_t is given by the equations

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

where ε_t are i.i.d. errors with mean zero and finite variance, and α and β are $p \times r$ matrices. The formulation allows modelling of both the long-run relations and the adjustment, or feedback, towards the attractor set $\{\beta' x = 0\}$ defined by the long-run relations, and this is the model we shall focus upon in this chapter. The models for different cointegrating ranks are nested and the rank can be analysed by likelihood ratio tests. The methods usually applied for the analysis are derived from the Gaussian likelihood function, (Johansen 1996), which we shall discuss here, see also Ahn and Reinsel (1990) and Reinsel and Ahn (1992).

1.3   The unobserved component formulation

Let x_t be given by

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

where u_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, the hypotheses of different ranks are nested. The parameter are linked to the autoregressive formulation by $\xi = \beta_{\perp}$ and $\eta = \alpha_{\perp}$, where for any $p \times r$ matrix a of rank $r \leq p$, we define $a_{\perp}$ as a $p \times (p-r)$ matrix of rank $p-r$, for which $a'a_{\perp} = 0$. Thus both adjustment and cointegration can be discussed in this formulation. Rather than testing for unit roots one tests for stationarity, which is sometimes a more natural formulation. The estimation is usually performed by the Kalman filter, and the asymptotic theory of the rank tests has been worked out by Nyblom and Harvey (2000).

1.4   The statistical methodology for the analysis of cointegration

In this chapter we analyse cointegration as modelled by the vector autoregressive model

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

where ε_t are i.i.d. with mean zero and variance Ω, and d_t are deterministic terms, like constant, trend, seasonals or intervention dummies. Under suitable conditions, see section 2, the process $(\beta' x_t, \Delta x_t)$ stationary around its mean, and subtracting the mean from (1) we find

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

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

By working throughout with a statistical model we ensure that we get a coherent framework for formulating and testing economic hypotheses. Thus our understanding of the dynamic behavior of the economic processes is expressed by the model.

The application of the likelihood methods gives a set of methods for conducting inference without having to derive properties of estimators and tests since they have been derived once and for all. Thus one can focus on applying the methods, using the now standard software available, provided the questions of interest can be formulated in the framework of the model.

The price paid for all this, is that one has to be reasonably sure that the framework one is working in, the vector error correction model, is in fact a good description of the data. This implies that one should always ask the fundamental question

Which statistical model describes the data?

That means that one should carefully check that the basic assumptions are satisfied, and, if they are not, one should be prepared to change the model or find out what are the implications of deviations in the underlying assumptions for the properties of the procedures employed.

The statistical methodology employed is to analyze the Gaussian likelihood function with the purpose of deriving estimators and test statistics. Once derived under the ideal Gaussian assumptions one derives the properties of the estimators and test statistics under more general assumptions.

The rest of this chapter deals with the following topics. In section 2 we give the definitions of integration and cointegration and the basic properties of the vector autoregressive $I(1)$ process as formulated in the Granger Representation Theorem, which is then used to discuss the role of the deterministic terms and the interpretation of cointegrating coefficients. In section 3 we show how various hypotheses can be formulated in the cointegrated model and discuss briefly the impulse response function. Then in section 4 we give the likelihood theory and discuss the calculation of maximum likelihood estimators under various restrictions on the parameters. Section 5 has a brief discussion of the asymptotics, including the different Dickey-Fuller distributions for the determination of rank and the mixed Gaussian distribution for the asymptotic distribution of $\hat{\beta}$, which leads to asymptotic χ^2 inference. We briefly mention the small sample improvements.

In section 6, we give some further topics that involve cointegration, like the implication of rational expectations for the cointegration model and models for seasonal cointegration, explosive roots, the model for $I(2)$ variables, non-linear cointegration and a few comments on models for panel data cointegration. These topics are still to be developed in detail and thus offer scope for a lot more research.

There are many surveys of the theory of cointegration, see for instance Watson (1994), and the topic has become part of most text book in econometrics, see among others Lütkepohl (1991), Banerjee, Dolado, Galbraith and Hendry (1993), Hamilton (1994), and Hendry (1995). It is not possible to mention all the papers that have contributed to the theory and we shall use as a general reference the monograph by Johansen (1996), where many earlier references can be found. The purpose of this survey is to explain some basic ideas, and show how they have been extended since 1996 for the analysis of new problems in the autoregressive model, and for the analysis of some new models. The theory of cointegration is an interesting econometric technique, but the main interest and usefulness of the methods lies in the applications in macroeconomic problems. We do not deal with the many applications of the cointegration techniques, but refer to the monograph by Juselius (2004) for a detailed treatment of the macroeconomic applications.

2   The vector autoregressive process and Granger's Representation Theorem

In this section we formulate first the well known conditions for stationarity of an autoregressive process and then show how these results generalize to integrated variables of order 1. The solution of the equations, the Granger Representation Theorem, is applied to discuss the role of the deterministic terms, the interpretation of the cointegrating coefficients, and in section 5, the asymptotic properties of the process.

2.1   The stationary vector autoregressive process and the definition of integration and cointegration

The vector autoregressive model for the $p-$dimensional process x_t

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t \tag{2}$$

is a dynamic stochastic model for all the variables x_t. The model is discussed in detail in Chapter 5. We assume that ε_t is i.i.d. with mean zero and variance Ω. 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 $d_1, \ldots, d_t$, and the parameters $(\Pi, \Gamma_1, \ldots, \Gamma_{k-1}, \Phi, \Omega)$. The deterministic terms are constants, linear terms, seasonals, or intervention dummies. The properties of x_t are studied through the characteristic polynomial

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

with determinant $\det(\Pi(z)) = |\Pi(z)|$ of degree at most kp. Let ρ_i^{-1} be the roots of $|\Pi(z)| = 0$. Then $\det(\Pi(z)) = \prod_{i=1}^{kp}(1 - z\rho_i)$ and the inverse matrix is given by

$$C(z) = \Pi^{-1}(z) = \frac{\text{adj}(\Pi(z))}{\det(\Pi(z))}, \quad z \neq \rho_i^{-1}$$

We mention the well known result, see Chapter 5,

Theorem 1 *If $|\rho_i| < 1$, the coefficients of $\Pi^{-1}(z) = C(1) = \sum_{i=0}^{\infty} C_i z^i$ are exponentially decreasing. Let $\mu_t = \sum_{i=0}^{\infty} C_i \Phi d_{t-i}$. Then the initial values of x_t can be given a distribution so that $x_t - \mu_t$ is stationary. The moving average representation of x_t is*

$$x_t = \sum_{i=0}^{\infty} C_i(\varepsilon_{t-i} + \Phi d_{t-i}). \tag{3}$$

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 will generate non-stationary processes of various types and the coefficients will not be exponentially decreasing. The process (3) is called a linear process and will form the basis for the definitions of integration and cointegration.

Definition 2 *We say that the process x_t is integrated of order 1, $I(1)$, if Δx_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, then x_t is cointegrated with cointegrating vector β. The number of linearly independent cointegrating vectors is the cointegrating rank.*

Example 3 *A bivariate cointegrated process given by the moving average representation*

$$x_{1t} = a \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{2t}$$
$$x_{2t} = b \sum_{i=1}^{t} \varepsilon_{1i} + \varepsilon_{3t}$$

is a cointegrated $I(1)$ process with $\beta = (b, -a)'$ because $\Delta x_{1t} = a\varepsilon_{1t} + \Delta\varepsilon_{2t}, \Delta x_{2t} = b\varepsilon_{1t} + \Delta\varepsilon_{3t}$ and $bx_{1t} - ax_{2t} = b\varepsilon_{2t} - a\varepsilon_{3t}$ are stationary.

Example 4 *A bivariate process given by the vector autoregressive model allowing for adjustment is*

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

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

$$x_{1t} = (S_t - \alpha_1 y_t)/(\alpha_2 - \alpha_1),$$
$$x_{2t} = (S_t - \alpha_2 y_t)/(\alpha_2 - \alpha_1).$$

This shows that if $|1 + \alpha_1 - \alpha_2| < 1$, 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 cointegrating vector $\beta' = (1, -1)$. We call S_t a common stochastic trend and α the adjustment coefficients. Note that the properties of the processes are derived from the equations and depend on the parameters of the model.

Example 2 presents a special case of the Granger Representation Theorem, which we give next.

2.2 The Granger Representation Theorem

If the characteristic polynomial $\Pi(z)$ has a unit root, then $\Pi(1) = -\Pi$ is singular, of rank r, say, and the process is not stationary. We let the $r \times p$ matrix β' denote the r linearly independent rows of Π, and let the $p \times r$ matrix α contain the coefficients

that express each row of $-\Pi$ as a combination of the vectors β', so that $\Pi = \alpha\beta'$. Equation (2) becomes

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t. \tag{4}$$

This is called the error or equilibrium correction model. We next formulate a condition, the $I(1)$ condition, which guarantees that the solution of (4) is a cointegrated $I(1)$ process. We define $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$.

Condition 5 *(The $I(1)$ condition). We assume that $\det(\Pi(z)) = 0$ implies that $|z| > 1$ or $z = 1$ and assume that*

$$\det(\alpha_\perp' \Gamma \beta_\perp) \neq 0. \tag{5}$$

This condition is needed to avoid solutions which have seasonal roots or explosive roots, and solutions which are integrated of order 2 or higher, see section 6. The condition is equivalent to the condition that the number of roots of $\det \Pi(z) = 0$ is $p - r$.

Theorem 6 *(The Granger Representation Theorem) If $\Pi(z)$ has unit roots and the $I(1)$ condition (5) is satisfied, then*

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

is convergent for $|z| \leq 1 + \delta$ for some $\delta > 0$ and

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

The process x_t has the moving average representation

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

where A depends on initial values, so that $\beta' A = 0$. It follows that x_t is a cointegrated $I(1)$ process with r cointegrating vectors β and $p - r$ common stochastic trends $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$.

The result (6) rests on the observation that the singularity of $\Pi(z)$ for $z = 1$ implies that $\Pi(z)^{-1}$ has a pole at $z = 1$. Condition (5) is a condition for this pole to be of order one. We shall not prove this here, but show how this result can be applied to prove the representation result (8). We multiply $\Pi(L)x_t = \Phi d_t + \varepsilon_t$ by

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

and find

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

Now define the stationary process $z_t = C^*(L)\varepsilon_t$ and the deterministic function $\mu_t = C^*(L)\Phi d_t$. Then

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

which cumulates to

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

where $A = x_0 - z_0 - \mu_0$. We choose the distribution of x_0 so that $\beta' x_0 = \beta'(z_0 + \mu_0)$, and hence $\beta' A = 0$. It is seen that x_t is $I(1)$, that $\beta' x_t = \beta' z_t + \beta' \mu_t$, so that $\beta' x_t$ is stationary around its mean $E(\beta' x_t) = \beta' \mu_t$, and that Δx_t is stationary around its mean $E(\Delta x_t) = C\Phi d_t + \Delta \mu_t$.
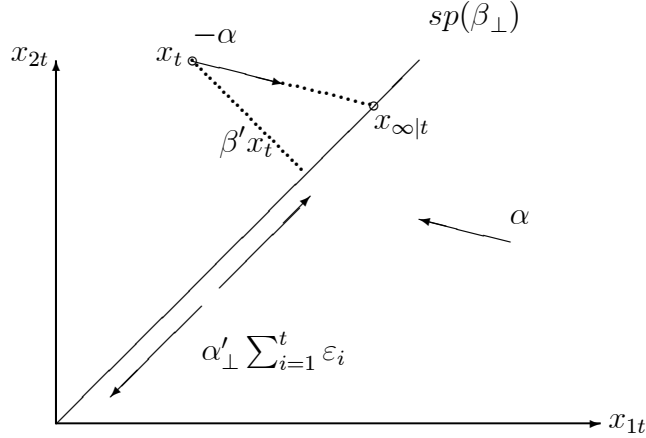
Figure 1: In the model $\Delta x_t = \alpha \beta' x_{t-1} + \varepsilon_t$, the point $x_t = (x_{1t}, x_{2t})$ is moved towards the long-run value $x_{\infty|t}$ on the attractor set $\{x|\beta' x = 0\} = sp(\beta_\perp\}$ by the force $-\alpha$ or $+\alpha$, and pushed along the attractor set by the common trends $\alpha'_\perp \sum_{i=1}^{t} \varepsilon_i$

It is easy to see that for a process with one lag we have $\Gamma = I_p$ and

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

so that the $I(1)$ condition is that the eigenvalues of $I_r + \beta'\alpha$ are bounded by one.

Engle and Granger (1987) show this result in the form that if $\Delta x_t = C(L)\varepsilon_t$ with $|C(1)| = 0$, then x_t satisfies an (infinite order) autoregressive model. We have chosen to start with the autoregressive formulation, which is the one estimated and which has the coefficients that have immediate interpretations, and derive the (infinite

order) moving average representation. Both formulations require the $I(1)$ condition (5). We next give three examples to illustrate the use of the Granger Representation Theorem.

Example 7 *First consider*

$$\Delta x_{1t} = -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t},$$

$$\Delta x_{2t} = \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t},$$

which gives an $I(1)$ process, with $\alpha' = \frac{1}{4}(-1, 1), \beta' = (1, -1)$, and $|1 + \alpha'\beta| = \frac{1}{2} < 1$.

Example 8 *Next consider*

$$\Delta x_{1t} = \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{1t},$$

$$\Delta x_{2t} = -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t},$$

for which $\alpha' = \frac{1}{4}(1, -1), \beta' = (1, -1)$, and $|1 + \alpha'\beta| = \frac{3}{2} > 1$. This describes an explosive process, and the $I(1)$ condition is not satisfied because there are roots inside the unit disk.

Example 9 *Finally consider a strange example:*

$$\Delta x_{1t} = \frac{1}{4}(x_{1t-1} - x_{2t-1}) + \frac{9}{4}\Delta x_{2t-1} + \varepsilon_{1t},$$
$$\Delta x_{2t} = -\frac{1}{4}(x_{1t-1} - x_{2t-1}) + \varepsilon_{2t}.$$

This process is a cointegrated $I(1)$ process with cointegrating relation $\beta' = (1, -1)$, because the $I(1)$ condition is satisfied despite the fact that the adjustment coefficients point in the wrong direction. The adjustment of the process towards equilibrium comes through the term Δx_{2t-1}.

2.3 The role of the deterministic terms

We apply the result (8) to discuss the role of a deterministic term in the equations, which could be called an 'innovation' term. It follows from (8) that d_t is cumulated into the trend $C\Phi \sum_{i=1}^{t} d_i$. We consider some special cases. Let first $\Phi d_t = \mu_0 + \mu_1 t$, so that

$$x_t = C \sum_{i=1}^{t}(\varepsilon_i + \mu_0 + \mu_1 i) + \sum_{i=0}^{\infty} C_i^*(\varepsilon_{t-i} + \mu_0 + \mu_1(t-i)) + A,$$

which shows, that in general, a linear term in the equation becomes a quadratic trend with coefficient $\frac{1}{2}Ct^2$ in the process. We decompose $\mu_i = \alpha\rho_i' + \alpha_\perp\gamma_i'$, where $\bar{\alpha}'\mu_i = \rho_i'$, and $\bar{\alpha}_\perp'\mu_i = \gamma_i'$. It is then seen that if $\gamma_1 = 0$, so that $\mu_1 = \alpha\rho_1'$, or

Model	linear term	restriction	trend in x_t	$E(\Delta x_t)$	$E(\beta' x_t)$
1	$\mu_0 + \mu_1 t$	none	quadratic	linear	linear
2	$\mu_0 + \alpha\rho_1' t$	$\xi_1 = 0$	linear	constant	linear
3	μ_0	$\mu_1 = 0$	linear	constant	constant
4	$\alpha\rho_0'$	$\mu_1 = 0, \xi_0 = 0$	constant	zero	constant
5	0	$\mu_1 = \mu_2 = 0$	zero	zero	zero

Table 1: The five models defined by restrictions on the deterministics terms in the equations. We decompose $\mu_i = \alpha\rho_i' + \alpha_\perp \gamma_i'$ ($\bar{\alpha}' \mu_i = \rho_i'$, $\bar{\alpha}_\perp' \mu_i = \gamma_i'$) and express the models by restrictions on ρ_i and γ_i.

$\alpha_\perp' \mu_1 = 0$, then the quadratic term has coefficient zero, $C\mu_1 = C\alpha\gamma_1' = 0$, so that only a linear trend is present. We, therefore, distinguish five cases as shown in Table 1.

We also consider a 'linear additive term' defined by

$$x_t = \tau_0 + \tau_1 t + z_t,$$

$$\Delta z_t = \alpha\beta' z_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta z_{t-i} + \varepsilon_t,$$

so that the deterministic part and the stochastic part are modelled independently. We eliminate z_t and find an error correction model (4) with

$$\mu_0 = -\alpha\beta'\tau_0 + \alpha\beta'\tau_1 + \Gamma\tau_1, \quad \mu_1 = -\alpha\beta'\tau_1,$$

thus corresponding to case 2 in Table 1. Similarly, if we take $\tau_1 = 0$, we get case 4, where the constant is restricted $\alpha_\perp' \mu_0 = 0$.

Finally we consider the case of an 'innovation dummy'. We define

$$d_t = 1_{\{t=t_0\}} = \begin{cases} 1, t = t_0 \\ 0, t \neq t_0 \end{cases}.$$

In this case the deterministic part of x_t is

$$C\Phi 1_{\{t \geq t_0\}} + \sum_{i=0}^{\infty} C_i^* \Phi d_{t-i} = (C\Phi + C_{t-t_0}^* \Phi) 1_{\{t \geq t_0\}}.$$

Because $C_{t-t_0}^* \to 0$, for $t \to \infty$, it is seen that the effect of an innovation dummy is that x_t changes from having 'level' zero up to time $t_0 - 1$ to having 'level' $C\Phi$ for large t.

On the other hand, if we model an 'additive dummy'

$$x_t = \Phi d_t + z_t,$$

$$\Delta z_t = \alpha\beta' z_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta z_{t-i} + \varepsilon_t,$$

we get the equation for x_t

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta z_{t-i} + \Phi\Delta d_t - \alpha\beta'\Phi d_{t-1} - \sum_{i=1}^{k-1} \Phi\Gamma_i \Delta d_{t-i} + \varepsilon_t,$$

which shows that, in the autoregressive formulation, we need the deterministics terms $\Delta d_t, d_{t-1}, \Delta d_{t-1}, \ldots, d_{t-k+1}$, with coefficient depending on $\Phi, \alpha, \beta, \Gamma_i$. The inclusion of deterministic terms in the equations thus requires careful consideration of which trending behavior or deterministic term is relevant for the data.

2.4 Interpretation of cointegrating coefficients

Usually regression coefficients in a regression like

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

are interpreted via a counterfactual experiment of the form: The coefficient γ_2 is the effect on x_{1t} of a change in x_{2t}, keeping x_{3t} constant. It is not relevant if in fact x_{3t} can in reality be kept constant when x_{2t} is changed, the experiment is purely a counterfactual or thought experiment.

The cointegrating relations are long-run relations. This is not taken to mean that these relations will eventually materialize if we wait long enough, but rather that these are relations, which have been there all the time and which influence the movement of the process x_t via the adjustments α, in the sense that 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.

It is therefore natural that the interpretation of cointegrating coefficients involves the notion of a long-run value. From the Granger Representation Theorem (8) applied to the model with no deterministic terms, one can find an expression for $E(x_{t+s}|x_t)$, which shows that

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

This limiting conditional expectation is the so-called long-run value of the process, which is a point in the attractor set, see Figure 1, because $\beta' x_{\infty|t} = 0$. The cointegrating relation can be formulated as a relation between long-run values: $\beta' x_{\infty|t} = 0$.

We see that if the current value is shifted from x_t 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 we want to achieve a given long-run change $k = C\xi$, say, we add Γk to the current value, as the long-run value becomes

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

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

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

so that $\beta' = (1, -\gamma_2, -\gamma_3)$. In order to give the usual interpretation as a regression coefficient (or elasticity if the variables are in logs), we would like to implement a long-run change so that x_2 changes by one, x_1 changes by γ_2, and x_3 is kept fixed. Thus the long-run change should be the vector $k = (\gamma_2, 1, 0)$, but this satisfies $\beta'k = 0$, and hence $k = C\xi$ for some ξ, so we can achieve the long-run change k by moving the current value to $x_t + C\Gamma k$.

In this sense, a coefficient in an identified cointegrating relation can be interpreted as the effect of a long-run change to one variable on another, keeping all others fixed. The difference with the usual interpretation of a regression coefficient is that because the relation is a long-run relation, that is, a relation between long-run values, the counterfactual experiment should involve a long-run change in the variables. More details can be found in Johansen (2004a), see also Proietti (1997).

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

We discuss here the model $H(r)$ defined by (1). The parameters are

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

All parameters vary freely, and α and β are $p \times r$ matrices. The models $H(r)$ form a nested sequence of hypotheses

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

where $H(p)$ is the unrestricted vector autoregressive model, or the $I(0)$ model, and $H(0)$ corresponds to the restriction $\Pi = 0$, which is the vector autoregressive model for the process in differences. The models in between, $H(1), \ldots, H(p-1)$, ensure cointegration and are the models of primary interest to us here. Note that in order to have nested models we allow in $H(r)$ 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 then be applied to check if ones prior knowledge of the number of cointegrating relations is consistent with the data, or alternatively to construct an estimator of the cointegrating rank.

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

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

The parameters α and β in (1) are not uniquely identified in the sense that given any choice of α and β and any non-singular matrix ξ $(r \times r)$ the choice $\alpha\xi$ and $\beta\xi'^{-1}$ will give the same matrix $\Pi = \alpha\beta' = \alpha\xi^{-1}(\beta\xi')'$ and hence determine the same probability distribution for the variables.

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

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

where u_t is stationary and $\theta' = -(\beta'_1)^{-1}\beta'_2$. 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, but many hypotheses are invariant to a normalization of β, and thus, in a discussion of a test of such a hypothesis, β does not require normalization. On the other hand, as seen in the next subsection, many economic hypotheses are expressed in terms of different restrictions, for which the regression formulation is not convenient.

Similarly, $\alpha_\perp$ and $\beta_\perp$ are not uniquely defined, so that the common trends are not unique. From the Granger Representation Theorem we see that common trends contribute with the non-stationary random walk term $C \sum_{i=1}^{t} \varepsilon_i$. For any full rank $(p - r) \times (p - r)$ matrix η, we have that $\eta \alpha'_\perp \sum_{i=1}^{t} \varepsilon_i$ could also be used because

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

This shows that we could have defined the common trends as $\eta \alpha'_\perp \sum_{i=1}^{t} \varepsilon_i$, so that identifying restrictions are needed in order to make sense of them.

3.2   Hypotheses on the long-run coefficients β

The main use of the concept of cointegration is as a precise definition of the economic concept of a long-run relation or equilibrium relation. In order to illustrate these ideas, consider the variables: m_t, log real money, y_t, log real income, $\pi_t = \Delta \log(p_t)$ the inflation rate, and two interest rates: a deposit rate i_t^d and a bond rate i_t^b. For simplicity, we first assume that we have only one cointegrating relation between the variables and formulate some natural hypotheses below.

The inverse money velocity is defined as $m_t - y_t$. We do not find that $m_t - y_t = c$ holds in the data, but instead interpret the statement that money velocity is constant as the statement that the process $m_t - y_t$ is stationary, or, in term of cointegration, that $\beta = (1, -1, 0, 0, 0)'$ is a cointegrating relation. Another hypothesis of interest is that velocity is a function of the interest rates only, which we formulate as $\beta = (1, -1, 0, \psi, \eta)'$, for some parameters ψ, η. We can formulate the question of whether the interest rate spread is stationary as the hypothesis that $\beta = (0, 0, 0, 1, -1)'$ and finally we can investigate the question of the stationarity of the inflation rate as the hypothesis $\beta = (0, 0, 1, 0, 0)'$.

Notice that stationarity of a variable is formulated as a question about the parameters of the model, that is, the model allows for both $I(0)$ variables and $I(1)$ variables.

If the cointegrating rank is two, we have more freedom in formulating hypotheses. In this situation we could ask if, in both relations, the coefficients to m and y add

to zero, and that the coefficient to the inflation rate is zero. This hypothesis is expressed as a restriction on the cointegrating relations

$$\beta = H\phi, \text{ or } R'\beta = 0,$$

where

$$R' = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, H = R_\perp.$$

The stationarity of velocity is formulated as the existence of a cointegrating relation of the form $b' = (1, -1, 0, 0, 0)$, that is, $\beta = (b, \phi)$, where b is known and the vector ϕ is the remaining unrestricted cointegrating vector.

A general formulation of linear restrictions on individual cointegrating relations is

$$\beta = (H_1\varphi_1, \ldots, H_r\varphi_r), \text{ or } R_i'\beta_i = 0, i = 1, \ldots, r, \quad (11)$$

where $H_i = R_{i\perp}$ is $p \times s_i$ and φ_i is $s_i \times 1$. In this way we impose $p - s_i$ restrictions on β_i. In order to identify the vector ϕ_i, we also have to normalize on one of its coefficients. An example of (11) for the case $r = 2$, is given by the hypotheses that m_t and y_t cointegrate and that i_t^d cointegrates with π_t. This hypothesis can be formulated as the existence of two cointegrating vectors of the form $(\varphi_{11}, \varphi_{12}, 0, 0, 0)$ and $(0, 0, \varphi_{21}, \varphi_{22}, 0)$ for some $\varphi_1 = (\phi_{11}, \phi_{12})'$, and $\varphi_2 = (\phi_{21}, \phi_{22})$. For this case we would take

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

The formulation (11) is the general formulation of linear restrictions on individual equations and identification is therefore possible, provided the identification condition is satisfied. In this particular case this means that, for instance, the first equation is identified by $R_1'\beta_1 = 0$, provided the rank condition is satisfied at the true value, that is, $\text{rank}(R_1'(\beta_2, \ldots, \beta_r)) \geq r - 1$.

Another set of conditions, which do not involve the true value, is given by

$$\text{rank}(R_1'(H_{i_1}, \ldots, H_{i_k})) \geq k,$$

for all $2 \leq i_1 \leq \ldots \leq i_k \leq r, k = 1, \ldots, r - 1$. These ensure that, for almost all values of the true parameter β, the rank condition is satisfied.

Finally one can, of course, impose general (cross equation) restrictions of the form $R'vec(\beta) = r_0$.

3.3 Hypotheses on the adjustment coefficients α and $\alpha_\perp$

There are two type of hypotheses on α that are of primary interest. The first is the hypothesis of weak exogeneity, see Engle, Hendry and Richard (1983), of some of the variables x_{2t}, say. We decompose x_t as $(x_{1t}', x_{2t}')'$ and the matrices similarly. The model equations without deterministics are then

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

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

$$\Delta x_{1t} = \omega \Delta x_{2t} + (\alpha_1 - \omega \alpha_2)\beta' x_{t-1} + \sum_{i=1}^{k-1}(\Gamma_{1i} - \omega \Gamma_{2i})\Delta x_{t-i} + \varepsilon_{1t} - \omega \varepsilon_{2t}, \qquad (12)$$

where $\omega = \Omega_{12}\Omega_{22}^{-1}$, if the errors are Gaussian. It is seen that, if $\alpha_2 = 0$, x_{2t} is weakly exogenous for α_1 and β, if there are no further restrictions on the parameters. This implies that efficient inference can be conducted on α_1 and β in the conditional model.

Another interpretation of the hypothesis of weak exogeneity is the following: if $\alpha_2 = 0$ then $\alpha_\perp = (0, I_{p-r})'$, so that the common trends are $\alpha_\perp' \sum_{i=1}^t \varepsilon_i = \sum_{i=1}^t \varepsilon_{2i}$. Thus the errors in the equations for x_{2t} cumulate in the system and give rise to the non-stationarity. This does not mean that the process x_{2t} cannot cointegrate, in fact it can be stationary for specific parameter values, as the next example shows

Example 10 *Consider the model*

$$\Delta x_t = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} \beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_{1t},$$

where evidently x_{2t} is weakly exogenous for the parameters α_1 and β, if all parameters are varying freely. The data generating process given by the equations

$$\begin{aligned} \Delta x_{1t} &= x_{2t-1} + \varepsilon_{1t}, \\ \Delta x_{2t} &= -\tfrac{1}{4}\Delta x_{1t-1} + \varepsilon_{2t}, \end{aligned}$$

is a special case with parameter values

$$\alpha' = (1, 0), \ \beta' = (0, 1), \ \Gamma_1 = -\frac{1}{4}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

which satisfy the $I(1)$ condition (5), and for which the weakly exogenous variable x_{2t} is stationary.

A general formulation of this type of hypothesis is

$$\alpha = A\psi,$$

which has the interpretation that $A_\perp' x_t$ is weakly exogenous for $A'\alpha_1$ and β.

Another hypothesis of interest is that there are some cointegrating relations that only appear in one equation. In this case one of the adjustment vectors is a unit vector, a say, or equivalently $\alpha_\perp$ has a zero row. The interpretation of this is that the shocks to the corresponding equation are not contributing to the common trends. This hypothesis can be formulated as $\alpha = (a, \psi)$, or equivalently that $\alpha_\perp = a_\perp \psi$. This is an example where a hypothesis on $\alpha_\perp$ is formulated as a hypothesis on α. Another such is of course the hypothesis $\alpha_\perp = (a, \psi)$ which is equivalent to $\alpha = a_\perp \phi$.

3.4 The structural error correction model

Multiplying equation (1) by a non-singular matrix A_0 gives the structural error (or equilibrium) correction model

$$A_0 \Delta x_t = \alpha^* \beta' x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i^* \Delta x_{t-i} + \Phi^* d_t + \varepsilon_t^*, \tag{13}$$

where the * indicates that the matrices have been multiplied by A_0. Note that the parameter β is the same as in (1), but that all the other coefficients have changed. In particular, A_0 is often chosen so that Ω^* is diagonal. Usually β is identified first, by suitable restrictions, and then the remaining parameters are identified by imposing restrictions on

$$\vartheta = \left(\alpha^*, \Gamma_0^*, \ldots, \Gamma_{k-1}^*, \Phi^*, \Omega^* \right). \tag{14}$$

Such restrictions are, of course, well known from econometric textbooks, see Fisher (1966), and the usual rank condition applies, as well as the formulations in connection with the identification of β.

The conclusion of this is that the presence of non-stationary variables allows two distinct identification problems to be formulated. First, the long-run relations must be identified uniquely in order that one can estimate and interpret them, and then the short-run parameters ϑ must be identified uniquely in the usual way.

The cointegration analysis allows us to formulate long-run *relations between variables*, but the structural error correction model formulates *equations for the variables* in the system. Thus if $r = 1$ in the example with money, income, inflation rate and interest rates, we can think of the cointegrating relation as a *money relation* if we solve it for money, but the equation for Δm_t in the structural model is a *money equation* and models the dynamic adjustment of money to the past and the other simultaneous variables in the system. The structural VAR is treated in Chapter 5, section 7.

3.5 Shocks, changes and impulse responses

Model (1) shows that a change in ε_t ($\varepsilon_t \mapsto \varepsilon_t + c$) is equivalent to a change in x_t ($x_t \mapsto x_t + c$). We shall call ε_t a shock and c a change. The Granger Representation Theorem shows that the effect at time $t + h$ of a change c to ε_t (or x_t) is

$$\frac{\partial x_{t+h}}{\partial \varepsilon_t}(c) = (C + C_h)c \to Cc, h \to \infty,$$

so that the impulse response function converges to Cc, which we shall call the long-run (or permanent) impact of the change c.

Sometimes one can give an economic meaning to linear combinations of shocks $e_i = v_i' \varepsilon_t$ and therefore one may want to induce changes to one of these and keep the remaining ones fixed. We introduce the notation

$$B^{-1} = (w_1, \ldots, w_p), \ B' = (v_1, \ldots, v_p),$$

and call $e_i = v_i' \varepsilon_t$ the structural shock and w_i its loading. We find

$$(C + C_h)\varepsilon_t = \sum_{i=1}^{p}(C + C_h)w_i v_i' \varepsilon_t = \sum_{i=1}^{p}(C + C_h)w_i e_i.$$

Now a change of one unit, say, to the structural shock e_i, keeping the others fixed, gives the impulse response function

$$h \mapsto (C + C_h)w_i.$$

One such possibility is to choose the Cholesky decomposition, so that B is triangular and so that $B\Omega B' = I_p$.

Because the shocks $\alpha_{\perp}' \varepsilon_t$ cumulate to the common trends, we define them as *permanent* shocks. It is natural to define *transitory* shocks as independent of the permanent ones, that is as $\alpha' \Omega^{-1} \varepsilon_t$. The decomposition

$$\varepsilon_t = \underbrace{\alpha(\alpha'\Omega^{-1}\alpha)^{-1}}_{loading} \underbrace{\alpha'\Omega^{-1}\varepsilon_t}_{trans.\ shock} + \underbrace{\Omega\alpha_{\perp}(\alpha_{\perp}'\Omega\alpha_{\perp})^{-1}}_{loading} \underbrace{\alpha_{\perp}'\varepsilon_t}_{perm.\ shock}$$

is a decomposition of the shocks ε_t into the transitory shocks and the permanent shocks. Note that the transitory shock has a loading proportional to α, so that the long-run effect of a transitory shock is zero. Note also that the loadings of the permanent shocks are suitable combinations of columns of Ω. Such loadings are used in the so-called generalized impulse response analysis, see Koop, Pesaran and Potter (1996).

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 introduces a notation for the calculations of reduced rank regression, introduced by Anderson (1951). We then discuss how reduced rank regression and modifications thereof are used to estimate the parameters of the $I(1)$ model (1) and various submodels.

4.1   Checking the specifications of the model

In order to apply the Gaussian maximum likelihood methods one has to check the assumptions behind the model carefully, so that one is convinced that the statistical model one has chosen contains the density that describes the data. If this is not the case, the asymptotic results available from the Gaussian analysis need not hold. The methods for checking the VAR model are outlined in Chapter 5, including the choice of lag length, the test for normality and tests for autocorrelation and heteroscedasticity in the errors. The asymptotic results for estimators and tests derived from the Gaussian likelihood turns 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. It turns out that heteroscedasticity does not influence the limit distributions, see Rahbek, Hansen, and Dennis (2002), whereas autocorrelated error terms will influence limit results, so this 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 dummy variables, but it is also important to model breaks in the dynamics, breaks in the cointegrating properties, breaks in the stationarity properties etc. The papers by Seo (1998) and Hansen and Johansen (1999) contain a theory for recursive estimation in the cointegrating 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, p_z respectively. We define a notation, which can be used to describe the calculations performed in regression and reduced rank regression, see Anderson (1951). We consider a regression model

$$U_t = \Pi W_t + \Gamma Z_t + \varepsilon_t, \tag{15}$$

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 we get by regressing U_t on W_t are

$$(U|W)_t = U_t - S_{uw} S_{ww}^{-1} W_t,$$

so that the conditional product moments are

$$S_{uw.z} = S_{uw} - S_{uz} S_{zz}^{-1} S_{zw} = T^{-1} \sum_{t=1}^{T} (U|Z)_t (W|Z)_t'$$

$$S_{uu.w,z} = T^{-1} \sum_{t=1}^{T} (U|W,Z)_t (U|W,Z)_t' = S_{uu.w} - S_{uz.w} S_{zz.w}^{-1} S_{zu.w}.$$

The unrestricted regression estimates are $\hat{\Pi} = S_{uw.z} S_{ww.z}^{-1}$, $\hat{\Gamma} = S_{uz.w} S_{zz.w}^{-1}$ 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 (15), when we assume that $\Pi = \alpha\beta'$ and α is $p_u \times r$ and β is $p_w \times r$. We first solve the eigenvalue problem

$$|\lambda S_{ww.z} - S_{wu.z} S_{uu.z}^{-1} S_{uw.z}| = 0. \tag{16}$$

The eigenvalue are ordered $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$, and the corresponding eigenvectors are $\hat{v}_1, \ldots, \hat{v}_p$. The interpretation of $\hat{\lambda}_1$, say, is as the maximal squared canonical

correlation between U and W corrected for Z, that is,

$$\hat{\lambda}_1 = \max_{\xi,\eta} \frac{(\xi' S_{uw.z}\eta)^2}{\xi' S_{uu.z}\xi\eta' S_{ww.z}\eta}.$$

The reduced rank estimates of β, α, Γ and Ω are given by

$$\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r),$$
$$\hat{\alpha} = S_{uw.z}\hat{\beta}(\hat{\beta}' S_{ww.z}\hat{\beta})^{-1},$$
$$\hat{\Gamma} = S_{uz.\hat{\beta}'w} S^{-1}_{zz.\hat{\beta}'w},$$
$$\hat{\Omega} = S_{uu.z} - S_{uw.z}\hat{\beta}(\hat{\beta}' S_{ww.z}\hat{\beta})^{-1}\hat{\beta}' S_{wu.z},$$

and we find $|\hat{\Omega}| = |S_{uu.z}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i)$. Often the eigenvectors are normalized on $\hat{v}'_i S_{ww.z}\hat{v}_j = 0$, if $i \neq j$, and 1 if $i = j$. The calculations described here is called a reduced rank regression and will be denoted by $RRR(U, W|Z)$.

4.3 Reduced rank regression in the $I(1)$ model and derivation of the rank test

We saw in section (2.3) that the role of a deterministic term changes when its coefficient is proportional to α. We therefore consider a model where some deterministic terms have this property, that is, the model

$$\Delta x_t = \alpha(\beta' x_{t-1} + \Upsilon D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t, \qquad (17)$$

where D_t and d_t are deterministic terms. Note that the coefficients to D_t, $\alpha\Upsilon$, has been restricted to be proportional to α. We assume for the derivations of maximum likelihood estimators and likelihood ratio tests that ε_t is i.i.d. $N_p(0, \Omega)$. The Gaussian likelihood function shows that the maximum likelihood estimation can be solved by the reduced rank regression

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

With the notation

$$R_{0t} = (\Delta x_t|\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t),$$
$$R_{1t} = \left((x'_{t-1}, D'_t)'|\Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t\right),$$
$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt},$$

we find that $(\hat{\beta}', \hat{\Upsilon})'$ solves the eigenvalue problem

$$|\lambda S_{11} - S_{10}S^{-1}_{00}S_{01}| = 0,$$

and that the maximized likelihood is, apart from a constant, given by

$$L_{\max}^{-2/T} = |\hat{\Omega}| = |S_{00}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i). \tag{18}$$

Note that we have solved all the models $H(r)$, $r = 0, \ldots, p$, by the same eigenvalue calculation. The maximized likelihood is given for each r by (18) and by dividing the maximized likelihood function for r with the corresponding expression for $r = p$ we get the likelihood ratio test for cointegrating rank, the so-called rank test or trace test:

$$-2\log LR(H(r)|H(p)) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i). \tag{19}$$

The asymptotic distribution of this test statistic and the estimators will be discussed in section 5. Next we discuss how a number of hypotheses or submodels can be analysed by reduced rank regression.

4.4 Hypothesis testing for the long-run coefficients β

We first consider the hypothesis $H_0 : \beta = H\phi$. Under H_0 the equation becomes

$$\Delta x_t = \alpha \left(\begin{array}{c} \phi \\ \Upsilon' \end{array} \right)' \left(\begin{array}{c} H'x_{t-1} \\ D_t \end{array} \right) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t,$$

which is solved by

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

If $\hat{\lambda}_i^*$ denote the eigenvalues derived under H_0, we find

$$-2\log LR(H_0|H(r)) = T \sum_{i=1}^{r} \log\{(1 - \lambda_i^*)/(1 - \hat{\lambda}_i)\}. \tag{20}$$

Similarly the hypotheses $\beta = b$ and $\beta = (b, H\phi)$ can be solved by reduced rank regression, but the more general hypothesis

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

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

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

This is evidently a reduced rank problem, but with r reduced rank matrices 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 found to converge quite often. The algorithm has the property that the likelihood function is maximized in each step. The algorithm switches between the reduced rank regressions

$$RRR(\Delta x_t, (x'_{t-1}H_i, D'_t)'|(x'_{t-1}H_j\phi_j, D'_t\Upsilon_j)'_{j\neq i}, \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t).$$

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

Another useful algorithm, see Boswijk (1992), consists of noticing, that for fixed $\{\phi_j, \Upsilon_j\}_{j=1}^r$, the likelihood is easily maximized by regression of Δx_t on $\{\varphi'_j H'_j x_{t-1} + \Upsilon_j D_t\}_{j=1}^r$ and lagged differences and d_t. This gives estimates of $\{\alpha_j\}_{j=1}^r$, $\{\Gamma_i\}_{i=1}^{k-1}$, Φ, and Ω. For fixed values of these, however, the equations are linear in $\{\phi_j, \Upsilon_j\}_{j=1}^r$, which can therefore be estimated by generalized least squares. By switching between these steps until convergence one can calculate the maximum likelihood estimators. This algorithm has the further advantage that one can impose restrictions of the form $R'vec(\beta) = r_0$, and the second step is still feasible.

4.5   Test on adjustment coefficients.

Under the hypothesis $H_0 : \alpha = A\psi$, in particular the hypothesis of weak exogeneity, we have the model

$$\Delta x_t = A\psi \left(\begin{array}{c} \beta \\ \Upsilon' \end{array} \right)' \left(\begin{array}{c} x_{t-1} \\ D_t \end{array} \right) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \Phi d_t + \varepsilon_t.$$

Multiplying by $\bar{A}' = (A'A)^{-1}A'$ and $A'_\perp$ and conditioning on $A'_\perp \Delta x_t$ we get the marginal and conditional models

$$\bar{A}'\Delta x_t = \omega A'_\perp \Delta x_t + \psi \left(\begin{array}{c} \beta \\ \Upsilon' \end{array} \right)' \left(\begin{array}{c} x_{t-1} \\ D_t \end{array} \right) + \sum_{i=1}^{k-1} \Gamma_{Ai} \Delta x_{t-i} + \Phi_A d_t + \varepsilon_{At} \qquad (21)$$

$$A'_\perp \Delta x_t = \sum_{i=1}^{k-1} \Gamma_{A_\perp i} \Delta x_{t-i} + \Phi_{A_\perp} d_t + \varepsilon_{A_\perp t}. \qquad (22)$$

where $\omega = \bar{A}'\Omega A_\perp (A'_\perp \Omega A_\perp)^{-1}$. The parameters in the marginal and the conditional model are variation independent, and β is estimated from the first equation by

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

4.6   Partial systems

The usual economic distinction between endogenous and exogenous variables is not present in the VAR formulation. If we decompose $x_t = (x'_{1t}, x'_{2t})'$ of dimension p_1

and p_2 and the matrices similarly, we get weak exogeneity when $\alpha_2 = 0$. Inference on β is efficiently conducted in the conditional model (21) with $A = (I_{p_1}, 0)$. Thus we can model the changes of the variables x_{1t} conditional on current changes of x_{2t} and lagged values of both variables, under the assumption of weak exogeneity. It is therefore tempting to use this conditional or partial model to make inference on both the cointegrating rank and the cointegrating relations.

The partial model is estimated by reduced rank regression

$$RRR(\Delta x_{1t}, (x'_{t-1}, D'_t)' | \Delta x_{2t}, \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, d_t).$$

and tests of rank and hypotheses on β and α_1 can be calculated as for the full model. However, the assumption of weak exogeneity, without which the analysis would not be efficient, has to be checked in the full model. If the full system is too large to analyse by cointegration, one can determine β from the conditional model and then test for the absence of $\hat{\beta}' x_{t-1}$ in a regression model for Δx_{2t}, see (22).

5 Asymptotic analysis

This section contains a brief discussion of the most important aspects of the asymptotic analysis of the cointegrating model without proofs and details. We give 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 tests for hypotheses on β are asymptotically distributed as χ^2, and the asymptotic distribution of $\hat{\beta}$ is mixed Gaussian. The asymptotic results are supplemented by a discussion of small sample corrections of the tests.

5.1 The asymptotic distribution of the rank test

We give the asymptotic distribution of the rank test in case the deterministic term is a polynomial of order d.

Theorem 11 *In model (17) with $D_t = t^d$ and $d_t = (1, t, \ldots, t^{d-1})$, the likelihood ratio test statistic $LR(H(r)|H(p))$ is given in (19). Under the assumption that the cointegrating rank is r, and ε_t i.i.d. $(0, \Omega)$ the asymptotic distribution is*

$$tr\{\int_0^1 (dB)F'(\int_0^1 FF'du)^{-1} \int_0^1 F(dB)'\}, \tag{23}$$

where F is defined by

$$F(u) = \left(\begin{array}{c} B(u) \\ t^d \end{array} \middle| 1, \ldots, t^{d-1} \right),$$

and where $B(u)$ is the $p - r$ dimensional Brownian motion. This 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 $\Gamma_1, \ldots, \Gamma_{k-1}$, Υ, Φ, Ω, but only on the dimension $p - r$, the number of common trends, and the order of the trend d. If $p = 1$ the limit distribution is the squared Dickey−Fuller distribution, see Dickey and Fuller (1981), and we therefore call the distribution (23) the Dickey–Fuller distribution with $p - r$ degrees of freedom, $DF_{p-r}(d)$.

If the deterministic terms are more complicated, they sometimes change the asymptotic distribution. 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, we normalize x_t 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 we let d_t be a centered seasonal dummy, or a 'innovation dummy' $d_t = 1_{\{t=t_0\}}$, they do not change the asymptotic distribution. If, on the other hand, we include the 'step dummy' $d_t = 1_{\{t \geq t_0\}}$, then the cumulation of this is a broken linear trend, and that will influence the limit distribution and requires special tables, see Johansen, Mosconi and Nielsen (2000).

For the partial models we also need special tables, see Harbo, Johansen, Nielsen and Rahbek (1998), or Pesaran, Shin and Smith (2000). The problem is that because $\alpha_2 = 0$, the cointegrating rank has to be less than the dimension, p_1, of the modelled variables as all the $\varepsilon's$ from the p_2 conditioning variables generate common trends. Thus the tables depends on the indices p_2 and $p_1 - r$. The general problem of including stationary regressors in the VAR has been treated by Mosconi and Rahbek (1999).

One can also test for rank when some of the cointegrating relations are known, see Horvath and Watson (1995) and Paruolo (2001). In that case we get new limit distributions which are convolutions of the Dickey-Fuller distributions for rank determination and χ^2 distributions used for inference for β.

5.2   Determination of cointegrating rank

Consider again model (17), with $D_t = t^d$ and $d_t = (1, \ldots, t^{d-1})$, where the limit distribution is given by (23). The tables are used as follows. If r represents prior knowledge that we want to test, we simply calculate the test statistic $Q_r = -2\log LR(H(r)|H(p))$ and compare it with the relevant quantile. Note that the tables give the asymptotic distribution only, and that the actual distribution depends not only on the finite value of T but also on the parameters $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1})$, but not on Φ, Υ, and Ω.

A common situation is that one has no, or very little, prior knowledge about r, and in this case it seems more reasonable to estimate r from the data. This is done as follows. First compare Q_0 with its quantile c_0, say. If $Q_0 < c_0$, we let $\hat{r} = 0$, if $Q_0 \geq c_0$ we calculate Q_1 and compare it with c_1. If now $Q_1 < c_1$ we define $\hat{r} = 1$, and if not we compare Q_2 with its quantile c_2, etc. This defines an estimator $\hat{r}$:

$$\{\hat{r} = r\} = \{Q_r < c_r, Q_{r-1} \geq c_{r-1}, \ldots, Q_0 \geq c_0\},$$

which takes on the values $0, 1, \ldots, p$, and which converges in probability to the

true value in the sense that, if 95% quantiles are used for the estimation, then

$$P_r(\hat{r} = r) \rightarrow 95\% \text{ and } P_r(\hat{r} < r) \rightarrow 0.$$

5.3   A small sample correction of the rank test

In Johansen (2002a, 2004c) a small sample correction for the rank test is developed under the assumption of Gaussian errors, which improves the usefulness of the asymptotic tables for the rank test. For finite samples, the distribution of the likelihood ratio test statistic depends on the unknown parameters under the null hypothesis. For $T \rightarrow \infty$ the dependence on the parameters disappears, but not uniformly in the parameter. Usually the distribution is shifted to higher values for finite T, and more so, the closer we are to the $I(2)$ boundary of the parameter space.

As an illustration of the results, consider the test for $\pi = \mu_1 = 0$ in the model for the univariate process x_t, with $k = 2s + 1$

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{2s} \gamma_i \Delta x_{t-i} + \mu_0 + \mu_1 t + \varepsilon_t.$$

Under the assumption that the process is $I(1)$, the limit distribution of the likelihood ratio test is the (squared) Dickey-Fuller test. We can then prove that if, instead of using $-2 \log LR(\pi = \mu_1 = 0)$, we divide by the quantity

$$(1 + 0.12T^{-1} + 4.05T^{-2})(1 + \frac{1.72}{T}[s + \frac{\sum_{i=1}^{2s} i\hat{\gamma}_i}{1 - \sum_{i=1}^{2s} \hat{\gamma}_i}]),$$

then the approximation to the limit distribution is improved. Note how the correction depends on the estimated parameters, in particular on values of $\sum_{i=1}^{k} \gamma_i$ close to one, where the correction tends to infinity. This corresponds to the process being almost $I(2)$. The numerical coefficients are determined by simulation of the various moments of a random walk for various value of T, and depend on the type of deterministics in the model.

Another example is given by the model with one lag, and p dimensions

$$\Delta x_t = \Pi x_{t-1} + \mu_1 t + \mu_0 + \varepsilon_t,$$

where we test $\Pi = \alpha\beta'$ and $\mu_1 = \alpha\beta_1'$, where α and β are $p \times 1$. Under the null we have

$$H_0 : \Delta x_t = \alpha(\beta' x_{t-1} + \beta_1' t) + \mu_0 + \varepsilon_t.$$

In this case the correction factor takes the form

$$(1 + T^{-1}a_1(p) + T^{-2}a_2(p))(1 + \frac{1}{T} \frac{k(\alpha, \beta, \Omega)}{\beta'\alpha})$$

where

$$k = -(2 + \beta'\alpha)m(p)\kappa + \{2(1 + \beta'\alpha)(p - 1) - 2\kappa(4 + 3\beta'\alpha)\}g(p)/(p - 1)^2$$

and where $\kappa = 1 - (\beta'\alpha)^2/\alpha'\Omega^{-1}\alpha\beta'\Omega\beta$, and the coefficients $a_1(p), a_2(p), m(p)$, and $g(p)$ are found by simulation. Notice that again the correction, and the test, give problems when $\alpha'\beta = 0$, which happens close to the $I(2)$ boundary.

5.4   The asymptotic distribution of β

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. The result is taken from Johansen (1996), but see also Anderson (2002). This result implies that likelihood ratio tests on β are asymptotically χ^2 distributed. We normalize $\hat{\beta}$ on $\bar{\beta}$, so that $\bar{\beta}'\hat{\beta} = I_r$, and find

Theorem 12 *In model (1) the asymptotic distribution of $\hat{\beta}$ is given by*

$$T\bar{\beta}'_\perp(\hat{\beta} - \beta) \xrightarrow{w} \left(\int_0^1 HH'du \right)^{-1} \int_0^1 H(dV)', \tag{24}$$

where

$$H = \beta'_\perp CW, \;\; and \; V = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W$$

are independent Brownian motions. An estimator of $\int_0^1 HH'du$ is $T^{-1}\hat{\beta}'_\perp S_{11}\hat{\beta}_\perp$.

Because H and V are independent, it follows that the limiting random variable Z, say, has a distribution given H that is Gaussian

$$Z|H \sim N_{(p-r)\times r}\left(0, (\alpha'\Omega^{-1}\alpha)^{-1} \otimes \left(\int_0^1 HH'du \right)^{-1} \right), \tag{25}$$

or equivalently

$$\left(\int_0^1 HH'du \right)^{1/2} Z(\alpha'\Omega^{-1}\alpha)^{1/2}|H \sim N_{(p-r)\times r}\left(0, I_r \otimes I_{p-r} \right), \tag{26}$$

so that

$$(\hat{\beta}'_\perp \sum_{t=1}^T R_{1t}R'_{1t}\hat{\beta}_\perp)^{1/2}\bar{\beta}'_\perp(\hat{\beta} - \beta)(\hat{\alpha}'\hat{\Omega}^{-1}\hat{\alpha})^{1/2} \xrightarrow{w} N_{(p-r)\times r}\left(0, I_r \otimes I_{p-r} \right).$$

This implies that Wald and likelihood ratio tests on β can be conducted using the asymptotic χ^2 distribution. Note that the asymptotic distribution is not Gaussian and that the scaling factor $\left(\int_0^1 HH'du \right)^{1/2}$ is not an inverse standard deviation, as we usually employ in inference for stationary processes. It is correct that the deviation $(\hat{\beta} - \beta)$ can be scaled to converge to the Gaussian distribution, but it is not correct that this scaling is done by an estimator of the asymptotic standard deviation.

One could say that the proper scaling is an estimator of the asymptotic conditional variance, given the function H, see Johansen (1995). The available information in the data is measured by the matrix $\hat{\alpha}'\hat{\Omega}^{-1}\hat{\alpha} \otimes \hat{\beta}'_\perp \sum_{t=1}^T R_{1t}R'_{1t}\hat{\beta}_\perp$, and if this is very large, $\hat{\beta}$ has a small 'standard error', but occasionally the information is small, and then large deviations of $\hat{\beta} - \beta$ can occur. We end by giving, without proof, a result on the test for identifying restrictions on β. We denote by $\{A_{ij}\}$ the matrix with blocks A_{ij} and by $\{H_i\}$ a block diagonal matrix with H_i in the diagonal.

Theorem 13 *Let β be identified by the restrictions $\beta = \{h_i + H_i\varphi_i\}_{i=1}^r$ where H_i is $p \times (s_i - 1)$ and φ_i is $(s_i - 1) \times 1$. Then the asymptotic distribution of $T(\hat{\beta} - \beta)$ is mixed Gaussian with an estimate of the asymptotic conditional variance given by*

$$Tdiag(\{H_i\}_{i=1}^r)\{\hat{\rho}_{ij}H_i'S_{11}H_j\}^{-1}diag(\{H_i'\}_{i=1}^r),$$

with $\rho_{ij} = \alpha_i'\Omega^{-1}\alpha_j$. The asymptotic distribution of the likelihood ratio test statistic for these restrictions is χ^2 with degrees of freedom given by $\sum_{i=1}^r(p - r - s_i + 1)$.

A small sample correction for the test on β has been developed by Johansen (2000a, 2002b), see also Omtzigt and Fachin (2001) for a discussion of this result and a comparison with the bootstrap.

To illustrate how to conduct inference on a cointegrating coefficient, and why it becomes asymptotic χ^2 despite the asymptotic mixed Gaussian limit of $\hat{\beta}$, we may consider a very simple case. Let x_t be a bivariate process with one lag for which $\alpha' = (-1, 0)$ and $\beta = (1, \theta)'$. The equations become

$$x_{1t} = \theta x_{2t-1} + \varepsilon_{1t}, \tag{27}$$
$$\Delta x_{2t} = \varepsilon_{2t}.$$

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 satisfies

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

Let us first analyse the distribution of $\hat{\theta}$ conditional on the process $\{x_{2t}\}$. We find that

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

aIt follows that $\hat{\theta}$ is mixed Gaussian with mixing parameter $1/\sum_{t=1}^T x_{2t-1}^2$, and hence has mean θ and variance $\sigma_1^2 E(1/\sum_{t=1}^T x_{2t-1}^2)$. When constructing a test for $\theta = \theta_0$ we do not base our inference on the Wald test

$$\frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} = \frac{\hat{\theta} - \theta}{\sqrt{E(1/\sum_{t=1}^T x_{2t-1}^2)}},$$

but rather on the Wald test which comes from an expansion of the likelihood function and is based on the observed information:

$$\left(\sum_{t=1}^T x_{2t-1}^2\right)^{1/2}(\hat{\theta} - \theta), \tag{28}$$

which is distributed as $N(0, \sigma_1^2)$. Thus we normalize by the observed information not the expected information often used when analysing stationary processes. In order
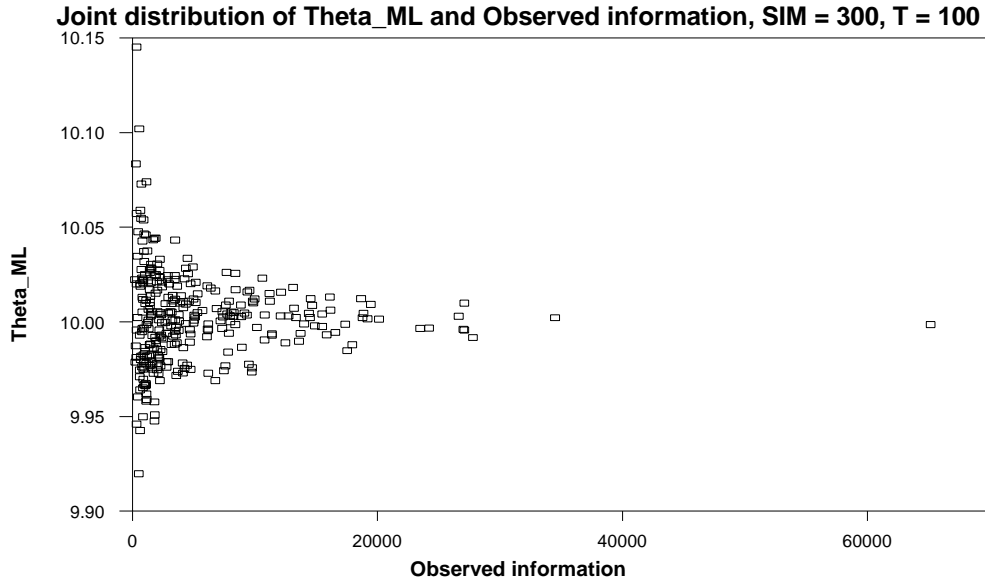
Joint distribution of Theta_ML and Observed information, SIM = 300, T = 100

Figure 2: The joint distribution of $\hat{\theta}$ and the observed information $\left(\sum_{i=1}^{T} x_{2t-1}^2/\hat{\sigma}^2\right)$ in the model (27). Note that the larger the information, the smaller is the uncertainty in the estimate $\hat{\theta}$.

to conduct inference we should therefore not consider the marginal distribution of the estimator, but the joint distribution of the estimator, $\hat{\theta}$, and the information in the data, $\sum_{t=1}^{T} x_{2t-1}^2/\sigma_1^2$, see Figure 2. The information should be exploited by conditioning, in order to achieve Gaussian inference, see Johansen (1995), and because the information and the t-ratio (28) are asymptotically independent, we can perform the conditioning (asymptotically) by simply considering the marginal distribution of the t-ratio instead.

6 Further topics in cointegration

The basic model for $I(1)$ variables can be applied to test economic hypotheses of different types and extended to allow for other types of non-stationarity. We mention here an application to test for rational expectations, and some extensions to models that allow for seasonal roots, explosive roots, and $I(2)$ variables, some non-linear error correction models and some models for panel data.

6.1 Rational expectations

Many economic models operate with the concept of rational or model based expectations, see Hansen and Sargent (1991). An example of such a formulation is the

uncovered interest parity,

$$\Delta^e e_{t+1} = i_t^1 - i_t^2, \tag{29}$$

which expresses a balance between the interest rates in two countries and the expected exchange rate changes. If we have fitted a vector autoregressive model to the data $x_t = (e_t, i_t^1, i_t^2)'$ of the form

$$\Delta x_t = \alpha\beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_t, \tag{30}$$

the assumption of model based expectations, (Muth 1961), means that $\Delta^e e_{t+1}$ can be replaced by $E_t \Delta e_{t+1}$ based upon the model (30). That is

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

The assumption (29) implies the identity

$$i_t^1 - i_t^2 = \alpha_1 \beta' x_t + \Gamma_{11} \Delta x_{t-1}.$$

Hence the cointegrating relation has to have the form

$$\beta' x_{t-1} = i_t^1 - i_t^2,$$

the adjustment is $\alpha_1 = 1$, and finally the first row of Γ_1 is zero: $\Gamma_{11} = 0$. Thus, the hypothesis (29) implies a number of testable restrictions on the vector autoregressive model. The implications of model based expectations and the cointegrated vector autoregressive model is explored in Johansen and Swensen (1999, 2004), where it is shown that, as in the example above, the rational expectation restrictions have information on the cointegrating relations and the short run adjustments. It is demonstrated how estimation under the rational expectation restrictions can be performed by regression and reduced rank regression in certain cases. See also Campbell and Shiller (1987) for an analysis of the expectation hypothesis in the cointegration model.

6.2 Seasonal cointegration

The cointegration theory so far described works under the assumption that the only unstable root of the process is $z = 1$. If roots are allowed at $z = -1, i, -i$ we get models that exhibit quarterly seasonal non-stationary variation.

This model has been analysed from the point of view of maximum likelihood by Lee (1992), Ahn and Reinsel (1994), and Johansen and Schaumburg (1998). In order to illustrate the concepts in a simple setting, we consider the model with roots only at $z = 1$ and $z = -1$, where we need $(1 - L)(1 + L)x_t = (1 - L^2)x_t$ to achieve stationarity under a condition corresponding to (5).

The error correction model, see Hylleberg, Engle, Granger, and Yoo (1990), in this case is found by expanding the autoregressive polynomial around $z = 1$ and $z = -1$ and a simplified version, without deterministics and lags, is

$$(1 - L)(1 + L)x_t = (1 + L)\alpha_1 \beta_1' x_{t-1} + (1 - L)\alpha_{-1} \beta_{-1}' x_{t-1} + \varepsilon_t. \tag{31}$$

The Granger Representation Theorem can be generalized to this case and gives the solution of the equations, or the moving average representation, under an $I(1)$ condition of the form (5)

$$x_t = C_1 \sum_{i=1}^{t} \varepsilon_i + C_{-1}(-1)^t \sum_{i=1}^{t} (-1)^i \varepsilon_i + A_1 + (-1)^t A_{-1} + y_t,$$

where y_t is stationary and A_1 and A_{-1} depend on initial values and satisfy $\beta_1' A_1 = 0$, and $\beta_{-1}' A_{-1} = 0$, and C_1 and C_{-1} have expressions like (7). This implies that the processes

$$(1-L)(1+L)x_t = C_1(1+L)\varepsilon_t + C_{-1}(1-L)\varepsilon_t + (1+L)(1-L)y_t,$$
$$(1+L)\beta_1' x_t = \beta_1' C_{-1}\varepsilon_t + (1+L)\beta_1' y_t,$$
$$(1-L)\beta_{-1}' x_t = \beta_{-1}' C_1 \varepsilon_t + (1-L)\beta_{-1}' y_t,$$

are stationary. The non-stationarity of x_t is due to the processes $S_t^{(1)} = \sum_{i=1}^{t} \varepsilon_i$ and $S_t^{(-1)} = (-1)^t \sum_{i=1}^{t} (-1)^i \varepsilon_i$, for which $(1-L)S_t^{(1)} = \varepsilon_t$, and $(1+L)S_t^{(-1)} = \varepsilon_t$. Maximum likelihood estimation of (31) involves two reduced rank regressions and can be performed by a switching algorithm, see also Cubadda (2001) for complex reduced rank regression in this model. Asymptotic inference can be conducted much along the same lines as for the usual $I(1)$ model, and we get the same basic results: inference on the rank requires new Dickey-Fuller distributions which can be expressed as stochastic integrals of complex Brownian motions, and inference on the remaining parameters is asymptotic χ^2.

6.3 Models for explosive roots

If the characteristic polynomial has one real explosive root $z = \lambda < 1$ and roots at $z = 1$, we get explosive processes for which $(1-L)(1-\lambda^{-1}L)x_t$ is stationary under an $I(1)$ condition corresponding to (5). We find, by expanding around $z = \lambda^{-1}$ and $z = 1$, the error correction model

$$(1-L)(1-\lambda^{-1}L)x_t = \alpha_1 \beta_1'(1-\lambda^{-1}L)x_{t-1} + \alpha_\lambda \beta_\lambda'(1-L)x_{t-1} + \varepsilon_t.$$

The solution under the $I(1)$ condition is

$$x_t = C_1 \sum_{i=1}^{t} \varepsilon_i + C_\lambda \lambda^{-t} \sum_{i=1}^{t} \lambda^i \varepsilon_i + \lambda^{-t} A_\lambda + A_1 + y_t,$$

where A_1 and A_λ depend on initial values and satisfy $\beta_1' A_1 = 0$ and $\beta_\lambda' A_\lambda = 0$, and y_t is stationary. The matrices C_1 and C_λ have expressions as given in (7). The non-stationarity is due to $S_t^{(1)} = \sum_{i=1}^{t} \varepsilon_i$, and $S_t^{(\lambda)} = \lambda^{-t} \sum_{i=1}^{t} \lambda^i \varepsilon_i$, which satisfies $(1-L)S_t^{(1)} = \varepsilon_t$ and $(1-\lambda^{-1}L)S_t^{(\lambda)} = \varepsilon_t$. Cointegration can remove one root, the other is removed by either $(1-L)$ or $(1-\lambda^{-1}L)$ so that the processes

$$(1-L)(1-\lambda^{-1}L)x_t = C_1(1-\lambda^{-1}L)\varepsilon_t + C_\lambda(1-L)\varepsilon_t + (1-L)(1-\lambda^{-1}L)y_t,$$
$$(1-\lambda^{-1}L)\beta_1' x_t = \beta_1' C_\lambda \varepsilon_t + (1-\lambda^{-1}L)\beta_1' y_t,$$
$$(1-L)\beta_\lambda' x_t = \beta_\lambda' C_1 \varepsilon_t + (1-L)\beta_\lambda' y_t,$$

are stationary. The sum $\sum_{i=0}^{t} \lambda^i \varepsilon_i$ converges for $t \to \infty$, and the explosiveness is due to the factor $\lambda^{-t} \to \infty$. This implies that the asymptotic theory is much more complicated as no central limit theorem can be invoked, but the limit distributions involve the random variable $\sum_{i=0}^{\infty} \lambda^i \varepsilon_i$, see Anderson (1959). The asymptotic theory is developed by Nielsen (2001a, 2001b, 2002). Maximum likelihood estimation of the model involves two reduced rank regressions and an estimation of λ. This can be performed by a suitable switching algorithm.

6.4   The I(2) model

If $z = 1$ is the only unstable root and $\alpha'_\perp \Gamma \beta_\perp$ has reduced rank, see (5), then we get integration of orders more than one. Under suitable conditions, similar to (5), we find that we need two differences to make the process stationary. We find an error correction model which can be parametrized as

$$\Delta^2 x_t = \alpha(\beta' x_{t-1} + \psi' \Delta x_{t-1}) + \Omega \alpha_\perp (\alpha'_\perp \Omega \alpha_\perp)^{-1} \kappa' \tau' \Delta x_{t-1} + \varepsilon_t, \ \ \beta = \tau \rho, \qquad (32)$$

where α and β are $p \times r$ and τ is $p \times (r + s)$, see Johansen (1997), or

$$\Delta^2 x_t = \alpha \left(\begin{array}{c} \beta \\ \delta' \end{array} \right)' \left(\begin{array}{c} x_{t-1} \\ \tau'_\perp \Delta x_{t-1} \end{array} \right) + \zeta \tau' \Delta x_{t-1} + \varepsilon_t, \qquad (33)$$

see Paruolo and Rahbek (1999). Here $\delta = \psi' \tau_\perp$ is of dimension $r \times (p - r - s)$. Under suitable conditions on the parameters, the equation has a solution of the form

$$x_t = C_2 \sum_{i=1}^{t} \sum_{j=1}^{i} \varepsilon_j + C_1 \sum_{i=1}^{t} \varepsilon_i + A_1 + t A_2 + y_t.$$

The coefficient matrices satisfy

$$\tau' C_2 = 0, \ \beta' C_1 + \psi' C_2 = 0, \ \tau'(A_1, A_2) = 0, \ \beta' A_1 + \psi' A_2 = 0,$$

so that the processes

$$\begin{aligned} \Delta^2 x_t &= C_2 \varepsilon_t + C_1 \Delta \varepsilon_t + \Delta^2 y_t, \\ \beta' x_t + \psi' \Delta x_t &= \beta' y_t + \psi' C_1 \varepsilon_t + \psi' \Delta y_t \\ \tau' \Delta x_t &= \tau' C_1 \varepsilon_t + \tau' \Delta y_t, \end{aligned}$$

are stationary. Thus the solution is an $I(2)$ process, and the cointegrating relations are given by $\tau' x_t$ (and hence $\beta' x_t = \rho' \tau' x_t$) is $I(1)$, but the model also allows for multicointegration, see Engle and Yoo (1991), that is, cointegration between the levels and the differences: $\beta' x_t + \psi' \Delta x_t$ is stationary. Equivalently one can say, since $\tau' \Delta x_t$ is stationary, that $\beta' x_t + \delta \tau'_\perp \Delta x_t$ is stationary, where δ is the so-called multicointegration parameter. Maximum likelihood estimation can be performed by a switching algorithm using the two parametrizations given in (32) and (33). The same techniques can be used for a number of hypotheses on the cointegrating parameters β and τ.

The asymptotic theory of likelihood ratio tests and maximum likelihood estimators is developed by Johansen (1997), Rahbek, Kongsted, and Jørgensen (1999), and Paruolo (1996, 2000). 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. Many different hypotheses on the parameters can be tested using asymptotic χ^2 tests, see Boswijk (2000) and Johansen (2004b).

6.5   Non-linear cointegration

There are obviously many different ways in which a linear model can be generalized to a non-linear model. We focus here on the non-linear error correction model, without deterministic terms, formulated as

$$\Delta x_t = f(\beta' x_{t-1}) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t, \qquad (34)$$

see Bec and Rahbek (2004) for a survey and recent results. For linear $f(\beta' x_{t-1}) = \alpha \beta' x_{t-1}$ we get model (1), and for the choice

$$f(\beta' x_{t-1}) = \left\{ \begin{array}{l} \alpha_1 \beta' x_{t-1}, \text{ if } |\beta' x_{t-1}| > \lambda \\ \alpha_2 \beta' x_{t-1}, \text{ if } |\beta' x_{t-1}| \leq \lambda \end{array} \right. ,$$

we get the Threshold Autoregressive (TAR) model, where the adjustment coefficients switch between α_1 and α_2, depending on the regime defined by the size of the disequilibrium error $|\beta' x_{t-1}|$. This kind of model has been used for testing for no cointegration, see Enders and Siklos (2001) and for testing for linear cointegration, see Balke and Fomby (1997) and Hansen and Seo (2002). No general results exist for inference on β, which is difficult to even calculate because of the discontinuous function f.

If we take

$$f(\beta' x_{t-1}) = \alpha_1 \exp(-|\beta' x_{t-1}|) + \alpha_2(1 - \exp(-|\beta' x_{t-1}|))$$

we get a smooth transition model, see Granger and Teräsvirta (1993).

In both cases one should think of the function f as modelling that the reaction to a disequilibrium error is different depending on the regime, but the switching is endogenous and does not depend on any outside influence.

Another type of model is where f is allowed to depend on an outside shock. Consider, as an example, a zero-one variable s_t, and a model of the form

$$\Delta x_t = f(\beta' x_{t-1}, s_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t, \qquad (35)$$

with f given by

$$f(\beta' x_{t-1}, s_t) = (s_t \alpha_1 + (1 - s_t)\alpha_2)\beta' x_{t-1}.$$

The distribution of s_t given the past and ε_t is given by for instance

$$P\{s_t = 1 | x_{t-1}, \dots x_{t-k+1}, \varepsilon_t\} = (1 - \exp(-|\beta' x_{t-1}|))/(1 + \exp(-|\beta' x_{t-1}|)).$$

In this case the adjustment coefficient switches between two states, where the probability of the two states is a smooth function of the process. Such a process is considered by Rahbek and Shephard (2002). The Markov switching models, where s_t is an independent Markov chain, were introduced in econometrics for stationary autoregressive processes by Hamilton (1989), and have gained widespread use. Only recently, see Douc, Moulines, and Rydén (2003), have properties of maximum likelihood estimators been established, but extensions of the results to the cointegrated model are still to be developed.

In general, this kind of model is difficult to analyse because of the non-linear reaction function. Instead of finding a linear representation of the process in terms of the errors, one has instead to prove the properties of the process directly. One can replace the usual notion of $I(0)$ by the notion of 'geometric ergodicity', (Bec and Rahbek 2004) or 'near epoch dependence', (Escribano and Mira 2002), and attempt to define $I(1)$ by the requirement that x_t converges weakly to a Brownian motion, but the final concepts have not been developed yet. What replaces the Granger Representation Theorem are results about $\beta' x_t$ and Δx_t being $I(0)$, whereas, under regularity conditions, x_t is not, so that the process is cointegrated. Furthermore, one can discuss existence of moments of the process x_t, which is useful for developing an asymptotic theory for the process and eventually for the estimators. Estimation of this model is relatively straightforward if β is known, but, as mentioned, the theory has yet to be developed for β unknown.

6.6   Panel data cointegration.

If we follow a panel of N units (countries) each with p variables, the data comes in the form $x_{it}, \; i = 1, \dots, N, \; t = 1, \dots, T$. We stack the vectors into the Np dimensional vector x_t and want to build a statistical model for x_t that reflects the panel structure. For illustration we assume that x_t satisfies the simple $I(1)$ model

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

The panel structure could then be formulated by the conditions that α, β and Ω are block diagonal corresponding to no feedback from one unit to another, no cointegration between units, and independent units. This model has been investigated by Larsson, Lyhagen and Løthgren (2001) and Groen and Kleibergen (2003) for general Ω.

For macro data, this is not a useful set of assumptions, and the problem is to model a large dimensional vector so that there is the possibility of 1) cointegration within a unit, 2) some cointegration between units, 3) the possibility of feedback

from disequilibrium in other units, and finally 4) some possibility of correlation between the shocks to different units.

An interesting solution has been proposed by Pesaran, Schuermann, Weiner (2004), who suggest to construct for each unit a "rest of the world" index $x_{it}^* = \sum_{j \neq i} w_{ij} x_{jt}$, and model the ith unit as

$$\Delta x_{it} = \alpha_i(\beta_i' x_{it-1} + \beta_i^{*\prime} x_{it-1}^*) + \varepsilon_{it}.$$

By stacking the observations into x_t and solving the models for x_t they obtain a model, that takes into account all the four requirements above.

All the models are of course submodels of the basic $I(1)$ model but the asymptotic theory is different, because we can let $N \to \infty$ or $T \to \infty$ or both, see Phillips and Moon (1999).

7 Conclusion

Granger (1983) coined the term cointegration, and it was his investigations of the relation between cointegration and error correction Engle and Granger (1987), that brought the modelling of vector autoregressions with unit roots into the center of attention in macro econometrics.

During the last 20 years, many have contributed to the development of the theory and the applications of cointegration. The account given here focusses on theory, more precisely on likelihood based theory for the vector autoregressive model and its extensions. The reason for focussing on model based inference is that, although we hope to derive methods with wide applicability, all methods have a limited applicability. By building a statistical model as a framework for hypothesis testing, one has to make explicit assumptions about the model used.

Therefore it becomes a natural part of the methodology to check assumptions, because if the assumptions are not satisfied, the same may hold for the results derived. Applying a rank test to some given data, without checking that the underlying vector autoregressive model has errors with no residual autocorrelation, and that the parameters of the model are constant, is as wrong as applying the continuos mapping theorem in asymptotic analysis, without checking that the function in question is in fact continuos.

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 text books, and the $I(1)$ procedures have been implemented in many different software packages, CATS in RATS, Givewin, Eviews, Microfit, Shazam etc. The $I(2)$ model is less developed but a version will appear in CATS.

Many theoretical problems remain unsolved, however. Time series rely heavily on asymptotic methods and it is often a problem to obtain 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. When these become part of the software packages, they will be routinely applied and ensure more reliable inference.

A very interesting and promising development lies with non-linear time series analysis, where the statistical theory is still in its beginning. There are many different types of non-linearities possible, and the theory has to be developed in close contact with the applications in order to ensure that useful models and concepts are developed.

Apart form this there is going to be a development and extension of cointegration in the area of panel data cointegration, seasonal cointegration, and the models for explosive roots. This development should also include software for the various models, in order that the theory can be easily applied and extended in interaction with applications.

Most importantly, however, is a totally different development which is needed, and that is a development of economic theory, which takes into account the findings of the empirical analyses of non-stationary economic data.

For a long time regression analysis and correlations have been standard ways of analysing relations between variables and cause and effect in economics. Economic theory has incorporated regression analysis as a useful tool for checking or falsifying economic predictions.

Similarly, empirical cointegration analysis of economic data reveals new ways of understanding economic data, and there is a need for building an economic theory that supports and explains these understanding.

8   References

Ahn, S. K., Reinsel, G. C., 1990. Estimation for partially non-stationary 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., 1959. On asymptotic distributions of estimates of parameters of stochastic difference equations. Annals of Mathematical Statistics 30, 676-687.

Anderson, T. W., 2002. Reduced rank regression in cointegrated models. Journal of Econometrics 106, 203-216.

Balke, N. S., Fomby, T. B., 1997. Threshold contegration. International Economic Review 38, 627-645.

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

Bec, F., Rahbek, A. C., 2002. Vector equilibrium correction models with non-linear discontinuous adjustments. Forthcoming Econometrics Journal.

Boswijk, P., 1992. Cointegration, Identification and Exogeneity: Inference in Structural Error Correction Models. Thesis Publisher, Tinbergen Institute, Amsterdam.

Boswijk, P., 2000. Mixed normality and ancillarity in $I(2)$ systems. Econometric Theory 16, 878-904.

Campbell, J., Shiller, R. J., 1987. Cointegration and tests of present value models. Journal of Political Economy 95, 1062-1088.

Cubadda, G., 2001. Complex reduced rank models for seasonally cointegrated time series. Oxford Bulletin of Economics and Statistics 63, 497-511.

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

Douc, R., Moulines, E., Rydén, T., 2003. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. Forthcoming Annals of Statistics.

Enders, W., Siklos, P. L., 2001. Cointegration and threshold adjustment. Journal of Business and Economic Statistics 19, 166-176.

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

Engle, R. F., Yoo B. S., 1991. Cointegrated economic time series: A survey with new results, in: C. W. J. Granger and Engle, R. F. (eds.,). Long-run Economic Relations. Readings in Cointegration. Oxford, Oxford University Press.

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

Escribano, A., Mira, S., 2002. Nonlinear error correction models. Journal of Time Series Analysis 23, 509-522.

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

Granger C. W. J., Teräsvirta, T., 1993. Modelling Non-linear Economic Relationships. Oxford, Oxford University Press.

Groen, J., Kleibergen, F. R., 2003. Likelihood based cointegration analysis in panels of vector error correction models. Journal of Business and Economic Statistics 21, 295-318.

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

Hansen, L. P., Sargent, T. J., 1991. Exact linear rational expectations models: Specification and estimation, in: L. P. Hansen, Sargent, T. J., (eds.,). Rational Expectations Econometrics, Westview Press, Boulder.

Hansen, B. E., Seo, B., 2002. Testing for two-regime threshold cointegration in vector error-correction models. Journal of Econometrics 110, 293-318.

Harbo, I., Johansen, S., Nielsen, B. G., Rahbek, A. C., 1998. Asymptotic inference on cointegrating rank in partial systems. Journal of Business Economics and Statistics 16, 388-399.

Harvey, A., Nyblom, J., 2000. Tests of common stochastic trends. Econometric Theory 16, 151-175

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

Horvath, M. T. K., Watson. M. 1995. Testing for cointegration when some of the cointegrating vectors are prespecified. Econometric Theory 11, 952-984.

Hylleberg, S., Engle, R. F., Granger, C. W. J., Yoo, S. B., 1990. Seasonal integration and cointegration. Journal of Econometrics 44, 215-238.

Johansen, S., 1995. The role of ancillarity in inference for non-stationary variables. Economic Journal 13, 302–320.

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., 2000a. A Bartlett correction factor for tests on the cointegrating relations. Econometric Theory 16, 740-778.

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

Johansen, S., 2002b. A small sample correction for tests of hypotheses on the cointegrating vectors. Journal of Econometrics 111, 195-221.

Johansen, S., 2004a. The interpretation of cointegrating coefficients in the cointegrated vector autoregressive model. Forthcoming Oxford Bulletin of Economics and Statistics.

Johansen, S., 2004b. The statistical analysis of hypotheses on the cointegrating relations in the $I(2)$ model. Forthcoming Journal of Econometrics.

Johansen, S., 2004c. A small sample correction for the Dickey-Fuller test, in: A. Welfe (ed.,). New Directions in Macromodelling, Chapter 3, Elsevier.

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

Johansen, S., Swensen, A. R., 1999. Testing rational expectations in vector autoregressive models. Journal of Econometrics 93, 73-91.

Johansen, S., Swensen, A. R., 2004. More on testing exact rational expectations in cointegrated vector autoregressive models: Restricted drift terms. The Econometric Journal 7, 389-397.

Johansen, S., Mosconi, R., Nielsen, B., 2000. Cointegration analysis in the presence of structural breaks in the determinstic trend. The Econometric Journal 3, 1-34.

Juselius. K., 2004. The Cointegrated VAR model: Econometric Methodology and Macroeconomic Applications. Monograph. Forthcoming, Oxford University Press, Oxford.

Koop, G., Pesaran, H. M., Potter, S. M., 1996. Impulse response analysis in nonlinear multivariate models. Journal of Econometrics 74, 119-147.

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

Lee, H.S., 1992. Maximum likelihood inference on cointegration and seasonal cointegration. Journal of Econometrics 54, 1-47.

Lütkepohl, H., 1991. Introduction to Multiple Times Series Analysis. Springer Verlag, Berlin.

Mosconi R., Rahbek, A.C., 1999. Cointegration rank inference with stationary regressors in VAR models. Econometrics Journal 2, 76-91.

Muth, J. F., 1961. Rational expectations and the theory of price movements. Econometrica 29, 315-335.

Nielsen, B., 2001a. The asymptotic distribution of likelihood ratio test statistics for cointegration in unstable vector autoregressive processes. Discussion paper, Nuffield, Oxford.

Nielsen, B., 2001b. Asymptotic properties of least squares statistics in general vector autoregressive models. Forthcoming Econometric Theory.

Nielsen, B., 2002. Cointegration analysis of explosive processes. Discussion paper, Nuffield, Oxford.

Omtzigt, P., Fachin, S. 2001. Bootstrapping and Bartett corrections in the cointegrated VAR model. Discussion paper, University "La Sapienza."

Park, J. Y., 1992. Canonical cointegrating regressions. Econometrica 60, 119-143.

Paruolo P., 1996. On the determination of integration indices in $I(2)$ systems. Journal of Econometrics 72, 313-356.

Paruolo P., 2000. Asymptotic efficiency of the two stage estimator in $I(2)$ systems. Econometric Theory 16, 4, 524-550.

Paruolo P., 2001. LR tests for cointegration when some cointegrating relations are known. Statistical Methods and Applications. Journal of the Italian Statistical Society 10, 123-137.

Paruolo. P., Rahbek, A. C., 1999. Weak Exogeneity in $I(2)$ VAR Systems. Journal of Econometrics 93, 281-308,

Pesaran, M. H., Shin, Y., Smith, R. J., 2000. Structural analysis of vector error correction models with exogenous $I(1)$ variables. Journal of Econometrics 97, 293-343.

Pesaran, M. H., Schuermann, T., Weiner, S., 2004. Modelling regional interdependencies using a global error-correction macroeconometric model. Journal of Business Economics and Statistics 22, 129-162.

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.

Phillips, P. C. B., Moon, H. R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67, 1057–1111.

Proietti, T., 1997. Short-run dynamics in cointegrated systems. Oxford Bulletin of Economics and Statistics 59, 405-422.

Rahbek, A. C., Kongsted, H. C., Jørgensen, C., 1999. Trend-stationarity in the $I(2)$ cointegration model. Journal of Econometrics 90, 265-289.

Rahbek, A. C., Hansen E., Dennis, J. G., 2002. ARCH innovations and their impact on cointegration rank testing. Preprint 12, 1998, University of Copenhagen.

Rahbek, A. C., Shephard, N., 2002. Autoregressive conditional root model: Inference and geometric ergodicity. Working paper, Nuffield College, Oxford.

Reinsel, G. C., Ahn, S. K., 1992. Vector autoregressive models with unit roots and reduced rank structure, estimation, likelihood ratio test, and forecasting. Journal of Time Series Analysis 13, 353-375.

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

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

Watson, M., 1994. Vector autoregressions and cointegration, in: R. F. Engle and McFadden, D. (eds.,). Handbook of Econometrics Vol. 4, North Holland.