Søren Johansen

A Statistical Analysis of Cointegration for I(2) Variables

Preprint
March
1991
2

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Søren Johansen

A STATISTICAL ANALYSIS OF COINTEGRATION FOR I(2) VARIABLES

Preprint 1991 No. 2

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

March 1991

University of Copenhagen

Institute of

Mathematical Statistics

February 1991

Søren Johansen

A STATISTICAL ANALYSIS OF COINTEGRATION FOR I(2) VARIABLES[1]

Abstract. This paper discusses inference for I(2) variables in a VAR model. The procedure suggested consists of two reduced rank regressions, and inference on the cointegration ranks can be conducted using the tables already prepared for the analysis of cointegration of I(1) variables. The paper contains a multivariate test for the existence of I(2) variables. The asymptotic distribution of the proposed estimators of the cointegrating coefficients is mixed Gaussian which implies that asymptotic inference can be conducted using the χ^2 distribution. The procedure is illustrated using a data set consisting of UK prices and exchange rates.

[1]This paper was written while the author was visiting Department of Statistics University of Helsinki. The visit was supported by the Danish Social Sciences Research Council acc. nr. 14–5793.

1. The models

Consider the vector autoregressive model with Gaussian errors in p dimensions

$$X_t = \Sigma_1^k \Pi_i X_{t-i} + \epsilon_t, \qquad t = 1,...,T,$$

where the initial values $X_{-k+1},...,X_0$ are fixed, where ϵ_t are independent Gaussian variables $N_p(0,\Omega)$, and the parameter space is given by

$$(1.1) \qquad (\Pi_1,...,\Pi_k,\Omega) \text{ unrestricted.}$$

It is convenient to rewrite the model in an error correction form that anticipates the analysis of the I(2) model given below

$$(1.2) \qquad \Delta^2 X_t = \Gamma \Delta X_{t-1} + \Pi X_{t-2} + \Sigma_1^{k-2} \Gamma_i \Delta^2 X_{t-i} + \epsilon_t.$$

The relation between the parameters $(\Gamma,\Pi,\Gamma_1,...,\Gamma_{k-2})$ and $(\Pi_1,...,\Pi_k)$ is found by identifying coefficients of the lagged levels in the two different expressions.

This model is the unrestricted VAR model and will be called the I(0) model in the following, since in some sense, see Theorem 1, most of the probability measures of interest in this model are those that make the process X_t stationary, and this is the model that is usually applied to describe the variation of stationary processes. There are, however, many probability measures, that is, choices of parameter values in the set described in (1.1), that make the process non stationary, and in the following we define and analyse two classes of such models. The first class consists of the first order cointegration models, or the I(1) models, given by the parameter restrictions

$$(1.3) \qquad H_r: \qquad \Pi = \alpha\beta',$$

where α and β are matrices of dimension p$\times$r. It follows from Granger's theorem, see Theorem 2, that in the I(1) model most processes are I(1). The likelihood analysis of this model is given in Johansen (1988a), see also Ahn and Reinsel (1988). We have used the notation H_r for the model defined by (1.3) where the dimension of α and β is r. We let H_r^0 denote the model where also the rank of α and β is r, that is, α and β have full rank. Then $H_r = \bigcup_{i=0}^{r} H_i^0$, and $H_0 \subset ... \subset H_p$, where H_p is just the I(0) model or the unrestricted VAR model.

We consider here the further sub models, the I(2) models, given by the parameter restrictions

(1.4) $\quad H_{r,s}$: $\quad \Pi = \alpha\beta'$

(1.5) $\quad H_{r,s}$: $\quad \alpha_\perp'\Gamma\beta_\perp = \varphi\eta'$

where α and β are $p\times r$ matrices of rank r and φ and η are $(p-r)\times s$ matrices.

Here and in the following we use for any $p\times r$ matrix γ of full rank the notation $\gamma_\perp$ to denote a $p\times(p-r)$ matrix of full rank such that $\gamma'\gamma_\perp = 0$. It is also convenient to have the notation $\bar{\gamma} = \gamma(\gamma'\gamma)^{-1}$, such that $\gamma'\bar{\gamma} = I$ and the projection onto the column space spanned by γ is $\gamma(\gamma'\gamma)^{-1}\gamma' = \bar{\gamma}\gamma' = \gamma\bar{\gamma}'$.

We also define $H_{r,s}^0$ as the submodel of $H_{r,s}$ defined by φ and η having full rank s. Then $H_{r,s} = \bigcup_{i=0}^{s} H_{r,i}^0$, and $H_{r,0} \subset H_{r,1} \subset \ldots \subset H_{r,p-r} = H_r^0 \subset H_r$.

Table 1

The relation between the various hypotheses in the I(2) model

p–r	r					
p	0	$H_{0,0}$	$\subset H_{0,1} \subset \ldots \subset$	$H_{0,p-1}$	$\subset H_{0,p} = H_0^0$	$\subset H_0$
						$\cap$
p–1	1		$H_{1,0} \subset \ldots \subset$	$H_{1,p-2}$	$\subset H_{1,p-1} = H_1^0$	$\subset H_1$
						$\cap$
						$\vdots$
1	p–1			$H_{p-1,0}$	$\subset H_{p-1,1} = H_{p-1}^0$	$\subset H_p$

Note that the model $H_{r,p-r}^0$ is the model that allows for r cointegrating relations and no I(2) variables. A multivariate test of this model is given in Corollary 9 .

The mathematical analysis of these models is given in Johansen (1990c) and can be briefly summarized as follows. Let

$$A(z) = I - \Sigma_1^k \Pi_i z^i = (1-z)^2 I - (1-z)\Gamma - \Pi z^2 - \Sigma_1^{k-2}\Gamma_i(1-z)^2 z^i$$

denote the characteristic polynomial of the autoregressive process (1.2). It is well known that the properties of this polynomial determines the properties of the process X_t defined by (1.2). In the probability analysis below we assume throughout that the roots of the characteristic polynomial satisfy the relation

(1.6) $\det A(z) = 0$ implies that $|z| > 1$ or that $z = 1$.

Thus we allow only unit roots, and exclude thereby both seasonal roots (on the unit circle) and explosive roots (inside the unit circle). We then formulate three theorems that describe the properties of the process X_t in the various models.

The first result is the well known condition for stationarity of the process and describes most of the processes in the I(0) model.

THEOREM 1. *If X_t satisfies (1.2) for all t and if Π has full rank then the process X_t is stationary.*

Thus condition (1.6) excludes unit roots and shows that most of the VAR′s, which satisfy condition (1.6), represent stationary processes, since the autoregressive representation (1.2) is invertible. In the next theorem which is a version of Granger′s representation theorem, see Engle and Granger (1987), we allow for unit roots, but in order to make sure that we only allow for non stationarity that can be removed by differencing we need another condition.

THEOREM 2. *If X_t satisfies (1.2) for all t, and if*

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

has reduced rank r and

(1.7) $$\alpha_\perp'\Gamma\beta_\perp$$

has full rank, then the process X_t is non stationary, ΔX_t is stationary and finally $\beta' X_t$ is stationary.

This result shows that in the I(1) model the processes which satisfy (1.7) are in fact I(1). This formulation of Granger′s result and the proof of Theorem 2 is given in Johansen (1990c). In view of Theorem 2 is looks natural to see what happens if the matrix in (1.7) has reduced rank. The next theorem shows that if this is the case and some further condition (1.11) is satisfied then we have I(2) processes.

In the following we apply the notation $A_{BC} = (B'B)^{-1}B'AC(C'C)^{-1}$ for matrices A,B and C of matching dimensions. The idea is of course that if in particular $B' = (I,0)$ and $C' = (I,0)$ and A is decomposed in block matrices $A = \{A_{ij}, i,j = 1,2\}$ then $A_{BC} = A_{11}$. We need the matrix

(1.8) $$\Theta = \Gamma\bar{\beta}\bar{\alpha}'\Gamma + I - \Gamma - \Sigma_1^{k-2}\Gamma_i$$

for the formulation of the next result.

THEOREM 3. *If X_t satisfies (1.2) for all t, and if*

(1.9) $$\Pi = \alpha\beta'$$

has rank r and if

(1.10) $$\bar{\alpha}_{\perp}'\Gamma\bar{\beta}_{\perp} = \varphi\eta'$$

has rank s, where φ and η are $(p-r)\times s$, and if further

(1.11) $$(\alpha,\alpha_{\perp}\varphi)_{\perp}'\Theta(\beta,\beta_{\perp}\eta)_{\perp}$$

has full rank, then the process X_t is non stationary and Δ^2X_t is stationary. If we define $\beta_1 = \beta_{\perp}\eta$, $\alpha_1 = \alpha_{\perp}\varphi$, $\beta_2 = (\beta,\beta_1)_{\perp}$, and $\alpha_2 = (\alpha,\alpha_1)_{\perp}$ then it holds that $\beta_1'\Delta X_t$ and $\beta'X_t + \bar{\alpha}'\Gamma\bar{\beta}_2\beta_2'\Delta X_t$ are stationary.

Further X_t has the representation

(1.12) $$X_t = C_2\Sigma_1^t\Sigma_1^s\epsilon_i + C_1\Sigma_1^t\epsilon_i + Y_t + X_0 - Y_0,$$

for some stationary process Y_t, where

$$C_1 = -\bar{\beta}_1\Gamma^{-1}_{\alpha_1\beta_1}\bar{\alpha}_1' + \bar{\beta}_1\Gamma^{-1}_{\alpha_1\beta_1}\Theta_{\alpha_1\beta_2}\Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2' - \bar{\beta}\Gamma_{\alpha\beta_2}\Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2',$$
$$C_2 = \bar{\beta}_2\Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2'.$$

A simple consequence is that

(1.13) $$\beta'C_1 = -\Gamma_{\alpha\beta_2}\Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2',$$

which will be used below. It is a consequence of (1.12) that for instance $\beta'X_t$ is I(1) since $\beta'C_2 = 0$. Similarly we find from (1.13) and (1.12) that apart from a stationary process

$$\beta'X_t = \beta'C_1\Sigma_1^t\epsilon_i = -\Gamma_{\alpha\beta_2}\Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2'\Sigma_1^t\epsilon_i,$$

and since $\beta_2'C_2 = \beta_2'\bar{\beta}_2\Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2' = \Theta^{-1}_{\alpha_2\beta_2}\bar{\alpha}_2'$ it follows that

$$\overline{\alpha}'\Gamma\overline{\beta}_2\beta_2'\Delta X_t = \Gamma_{\alpha\beta_2}\Theta^{-1}_{\alpha_2\beta_2}\overline{\alpha}_2'\Sigma_1^t\epsilon_i,$$

which gives the result that $\beta' X_t + \overline{\alpha}'\Gamma\overline{\beta}_2\beta_2'\Delta X_t$ is stationary. Thus under conditions (1.9,1.10,1.11) the process itself is I(2), and the restrictions (1.9) and (1.10) describe the I(2) model $H^0_{r,s}$. The linear combinations β and β_1 reduce the order from 2 to 1. The vectors in β have the further property that the I(1) variable $\beta' X_t$ cointegrate with $\beta_2{}'\Delta X_t$. Thus the vectors in β and β_2 capture the notion of multi cointegration or polynomial cointegration see Engle and Yoo (1989), Granger and Lee (1988), and Johansen (1988b).

The illustrative example in section 4 needs a more complicated model involving seasonal dummies, constant term and strictly exogenous variables. We here treat in detail the model without these complications and later mention how the results have to be modified.

Before proceeding to the statistical analysis of the I(2) models we need one more result that is crucial for the asymptotic analysis and also for the understanding of the role of the coefficients β. It follows from the above that the vectors β do not necessarily reduce the order from 2 to 0, but there is another process derived from X_t for which β has this property. In order to see this define

$$R_{tT} = X_t - [\Sigma_1^T X_t\Delta X_t'][\Sigma_1^T \Delta X_t\Delta X_t']^{-1}\Delta X_t,$$

that is the levels corrected for the differences over the interval 1,...,T.

THEOREM 4: *Under the assumptions of Theorem 3 it holds that $\beta' X_t$ regressed on ΔX_t, $t = 1,\dots,T$, converges weakly to a stationary process, i.e.*

$$\beta' R_{tT} \overset{w}{\to} \beta' X_t + \overline{\alpha}'\Gamma\overline{\beta}_2\beta_2'\Delta X_t, \quad T \to \infty$$

which is stationary by Theorem 3.

The proof of this result is given in section 3.

2. The statistical analysis

The likelihood based analysis of the I(0) model reduces to ordinary least squares regression and will not be discussed here. The analysis of the I(1) model H_r is performed by a combination of regression and reduced rank regression, see Anderson 81951) or Johansen (1988a). In order to introduce the notation for the subsequent analysis of $H_{r,s}$, we give some details here.

In model H_r the matrix Π is restricted as $\Pi = \alpha\beta'$ but the parameters $(\alpha,\beta,\Gamma,\Gamma_1,...,\Gamma_{k-2})$ vary independently. Hence the parameters $\Gamma_1,...,\Gamma_{k-2}$ can be eliminated by regressing Δ^2X_t, ΔX_{t-1} and X_{t-2} on $\Delta^2X_{t-1},...,\Delta^2X_{t-k+2}$. This gives residuals R_{0t}, R_{1t} and R_{2t}, and residual product moment matrices

(2.1) $$M_{ij} = T^{-1}\Sigma_1^T R_{it}R_{jt}', \qquad i,j = 0,1,2.$$

The remaining analysis of both the I(1) and the I(2) models can be performed from the equations

(2.2) $$R_{0t} = \Gamma R_{1t} + \alpha\beta' R_{2t} + \epsilon_t.$$

The analysis of the I(1) model is given by first eliminating the unrestricted parameters Γ by regression and forming the residual product moment matrices

$$M_{ij.1} = M_{ij} - M_{i1}M_{11}^{-1}M_{1j}, \qquad i,j = 0,2,$$

and residuals

$$R_{i.1t} = R_{it} - M_{i1}M_{11}^{-1}R_{1t}, \qquad i = 0,2.$$

Next solve the eigenvalue problem

(2.3) $$|\lambda M_{22.1} - M_{20.1}M_{00.1}^{-1}M_{02.1}| = 0,$$

for eigenvalues $\hat{\lambda}_1 > ... > \hat{\lambda}_p$ and eigenvectors $V = (v_1,...,v_p)$. The maximum likelihood estimators are then given by

(2.4) $$\hat{\beta} = (v_1,...,v_r),\ \hat{\alpha} = M_{02.1}\hat{\beta},\ \hat{\Omega} = M_{00.1} - \hat{\alpha}\hat{\alpha}'.$$

Finally the maximized likelihood function is found from

$$L_{max}^{-2/T} = |\hat{\Omega}| = |M_{00.1}|\Pi_1^r(1-\hat{\lambda}_i).$$

From this it follows that if one wants to test the I(1) model H_r with rank $\leq r$ in the I(1) model with rank $\leq p$, H_p, i.e. in the unrestricted VAR model, the likelihood ratio test becomes

(2.5) $$T_r = -2\ln Q(H_r|H_p) = -T\Sigma_{r+1}^p \ln(1-\hat{\lambda}_i), \quad (r = 0,...,p-1).$$

The asymptotic properties of the estimators and test statistics were given in Johansen (1988a), under the assumptions of Theorem 2, i.e. when the process is I(1). There it was also shown that the limit distribution of $T(\hat{\beta}-\beta)$ is mixed Gaussian, and hence asymptotic inference on hypotheses on β can be conducted using the χ^2 distribution. The asymptotic distribution of the test statistic (2.5) is given under the assumptions of Theorem 2 as a functional of Brownian motion that can be expressed as

$$\mathrm{tr}\{\textstyle\int_0^1(dB)B'[\int_0^1 BB'du]^{-1}\int_0^1 B(dB)'\}. \tag{2.6}$$

Here B is a Brownian motion of dimension p–r on the unit interval. The dimension p–r is called the degrees of freedom for the test statistic.

Some applications of these methods are given in Johansen and Juselius (1990,1991).

With the above notation we can now describe the statistical analysis of the I(2) model $H_{r,s}$. The model is defined by the reduced rank conditions (1.4) and (1.5) on Π and Γ, but since the last condition depends on the reduced rank of the first condition it is rather involved to estimate the parameters simultaneously, see Johansen (1990b) for an algorithm. Instead we give here a method that is very easy since it only involves regression and reduced rank regression.

The first step is to determine (r,α,β) from the I(1) model, that is estimate (1.2) or equivalently (2.2) with Γ unrestricted, giving the estimates (2.4). Next assume that these parameters are known and fixed and continue the analysis of (2.2) as follows. Multiply by $\bar{\alpha}'$ and $\bar{\alpha}_\perp'$ in (2.2) to obtain

$$\alpha' R_{0t} = \bar{\alpha}'\Gamma R_{1t} + \beta' R_{2t} + \bar{\alpha}'\epsilon_t, \tag{2.7}$$

$$\bar{\alpha}_\perp' R_{0t} = \bar{\alpha}_\perp'\Gamma R_{1t} + \bar{\alpha}_\perp'\epsilon_t. \tag{2.8}$$

Define $\omega = \bar{\alpha}'\Omega\bar{\alpha}_\perp(\bar{\alpha}_\perp'\Omega\bar{\alpha}_\perp)^{-1}$, and subtract ω times (2.8) from (2.7) to get

$$\bar{\alpha}' R_{0t} = \omega\bar{\alpha}_\perp' R_{0t} + (\bar{\alpha}'\Gamma - \omega\bar{\alpha}_\perp'\Gamma)R_{1t} + \beta' R_{2t} + (\bar{\alpha}' - \omega\bar{\alpha}_\perp')\epsilon_t. \tag{2.9}$$

Note that the errors in (2.8) and (2.9) are independent, and that the coefficients in (2.8) are variation independent of those in (2.9) even under the restriction (1.10), which only restricts the coefficients of (2.8). The equations can therefore be analysed independently, and the analysis of (2.9) is just a regression of $(\bar{\alpha}' R_{0t} - \beta' R_{2t})$ on $\bar{\alpha}_\perp' R_{0t}$ and R_{1t}.

We now focus on the analysis of (2.8), which we write using $I = \bar{\beta}\beta' + \bar{\beta}_\perp\beta_\perp'$ as

$$\bar{\alpha}_\perp' R_{0t} = \bar{\alpha}_\perp'\Gamma(\bar{\beta}\beta' + \bar{\beta}_\perp\beta_\perp')R_{1t} + \bar{\alpha}_\perp'\epsilon$$

or

$$\bar{\alpha}_\perp' R_{0t} = \bar{\alpha}_\perp'\Gamma\bar{\beta}(\beta' R_{1t}) + \varphi\eta'(\beta_\perp' R_{1t}) + \bar{\alpha}_\perp'\epsilon. \tag{2.10}$$

From this it is seen that the analysis consists in eliminating the parameters $\bar{\alpha}_\perp'\Gamma\bar{\beta}$ by regression of $\overline{\alpha}_\perp' R_{0t}$ and $\beta_\perp' R_{1t}$ on $\beta' R_{1t}$, and then solve a reduced rank problem

(2.11) $$|\rho M_{\beta_\perp\beta_\perp.\beta} - M_{\beta_\perp\alpha_\perp.\beta} M_{\alpha_\perp\alpha_\perp.\beta}^{-1} M_{\alpha_\perp\beta_\perp.\beta}| = 0.$$

We here use the rather compact notation

$$M_{\beta_\perp\beta_\perp.\beta} = \beta_\perp'(M_{11} - M_{11}\beta(\beta' M_{11}\beta)^{-1}\beta' M_{11})\beta_\perp,$$

$$M_{\beta_\perp\alpha_\perp.\beta} = \beta_\perp'(M_{10} - M_{11}\beta(\beta' M_{11}\beta)^{-1}\beta' M_{10})\alpha_\perp,$$

and

$$M_{\alpha_\perp\alpha_\perp.\beta} = \alpha_\perp'(M_{00} - M_{01}\beta(\beta' M_{11}\beta)^{-1}\beta' M_{10})\alpha_\perp.$$

Note that here $\beta_\perp$ and β transforms the differences, which are I(1) variables, and $\alpha_\perp$ the second differences which are stationary under the assumptions of Theorem 3. The notation indicates how the matrices are transformed and how they are conditioned. The solution of (2.11) gives eigenvalues $\hat{\rho}_1 > \ldots > \hat{\rho}_{p-r}$ and eigenvectors $W = (w_1,\ldots,w_{p-r})$, and the maximum likelihood estimators (for fixed known r, α and β) are then

$$\hat{\eta} = (w_1,\ldots,w_s),\ \hat{\varphi} = M_{\alpha_\perp\beta_\perp.\beta}\hat{\eta},\ \hat{\Omega}_{\alpha_\perp\alpha_\perp} = M_{\alpha_\perp\alpha_\perp.\beta} - \hat{\varphi}\hat{\varphi}'.$$

Finally the factor of the likelihood function that comes from equation (2.10) is given by

$$L_{\max}^{-2/T} = |\alpha_\perp'\alpha_\perp|^{-1} |M_{\alpha_\perp\alpha_\perp.\beta}| \ \Pi_1^s(1-\hat{\rho}_i).$$

This shows that, still for known values of r, α, and β, the likelihood ratio test of the model with reduced rank $\leq s$, $H_{r,s}$, in the model $H_{r,p-r} = H_r^0$ is given by

(2.12) $$T_{r,s} = -2\ln Q(H_{r,s}|H_r^0) = -T\Sigma_{s+1}^{p-r}\ln(1-\hat{\rho}_i),$$

since the factor that comes from equation (2.9) is the same with and without restriction (1.5).

In summary the procedure that we propose here is first to solve the estimation problem for the I(1) model for all values of r by solving (2.3). Next consider for each value of r, the estimates of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\alpha}_\perp$, and $\hat{\beta}_\perp$ and form the differences of the common trends $\alpha_\perp' X_t$, which satisfy the equation (2.10) which does not contain terms involving the levels. Since $\beta = \hat{\beta}$ is assumed known at this stage, the process $\beta' \Delta X_{t-1}$ is a known stationary process and since its coefficients are unrestricted by condition (1.5) they can be eliminated together with the coefficients of the lagged second differences $\Delta^2 X_{t-1}, \ldots \Delta^2 X_{t-k+2}$. What remains is a reduced rank regression problem for the variables in differences.

The cointegration ranks r and s are determined by testing the hypotheses H_0, H_1,... using the test statistics T_r, and let $\hat{r}$ be the index of the first non rejected hypothesis. Next we test for $r = \hat{r}$ the hypotheses $H_{r,0}$, $H_{r,1}$,... using the test statistics $T_{r,s}$, and let s be the the index of the first non rejected hypothesis. In particular if $H_{r,s}$ is rejected for all s, we accept the hypothesis that there are r cointegrating relations and no I(2) variables.

We give now some results which demonstrate the properties of the suggested procedure

THEOREM 5: *Under the conditions of Theorem 3, i.e. under model $H^0_{r,s}$ the asymptotic distribution of the likelihood ratio test statistic $T_{r,s}$ (2.12) of $H_{r,s}$ versus H^0_r derived for r,α and β known, is given by (2.6) with $p-r-s$ degrees of freedom. Further the asymptotic distribution of the maximum likelihood estimators of the coefficients η, see (2.10), is mixed Gaussian, such that usual inference can be conducted for testing hypotheses about η, if r, α and β are known.*

This result is of course not so interesting since in general r, α and β have to be estimated. We therefore analyse how the estimators $\hat{\alpha}$ and $\hat{\beta}$ from (2.4) behave, if infact there are I(2) variables in the model, that is, if model $H^0_{r,s}$ holds.

In order to describe the asymptotic distributions let W_t be a Brownian motion with covariance Ω defined by

$$T^{-\frac{1}{2}}\Sigma_0^{[Tt]}\epsilon_i \overset{w}{\to} W_t$$

and define

$$V_\alpha = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W.$$

We also need to define two more processes:

(2.13) $$F_{1t} = \beta_1'C_1(W_t - [\int_0^1 W_u W_u' du\alpha_2][\alpha_2'\int_0^1 W_u W_u' du\alpha_2]^{-1}\alpha_2' W_t),$$

and

(2.14) $$F_{2t} = \beta_2'C_2(\int_0^t W_u du - [\int_0^1(\int_0^u W_s ds)W_u' du\alpha_2][\alpha_2'\int_0^1 W_u W_u' du\alpha_2]^{-1}\alpha_2' W_t).$$

These processes are derived from the Brownian motion and the integrated Brownian motion respectively by correcting for the α_2 components of the Brownian motion $\alpha_2' W_t$. Finally we let $F' = (F_1', F_2')$. The dimension of F_1 is s and that of F_2 is p–r–s.

We also define a Brownian motion $B' = (B_1', B_2')$ with covariance matrix I such that B_1 is s dimensional and B_2 is p–r–s dimensional, and processes $G' = (G_1', G_2')$ defined in analogy with F

(2.15) $$G_{1t} = B_{1t} - [\int_0^1 B_1 B_2' du][\int_0^1 B_2 B_2' du]^{-1} B_{2t},$$

(2.16) $$G_{2t} = \int_0^t B_{2u} du - [\int_0^1 \int_0^u B_{2s} ds B_2' du][\int_0^1 B_2 B_2' du]^{-1} B_{2t}.$$

THEOREM 6. *Under the assumptions of Theorem 3 the estimators $\hat{\alpha}$ and $\hat{\beta}$ given by (2.4) are consistent and the asymptotic distribution of $\hat{\beta}$ is given by*

$$(T\bar{\beta}'(\hat{\beta}-\beta), T^2\bar{\beta}_2'(\hat{\beta}-\beta)) \xrightarrow{w} (\int_0^1 FF' du)^{-1} \int_0^1 F(dV_\alpha)'$$

where F, given by (2.13) and (2.14), is independent of V_α.

The asymptotic distribution of $\hat{\alpha}$ is given by

$$T(\hat{\alpha} - \alpha) \xrightarrow{w} N_{p \times r}(0, \Omega \otimes (\beta' \Sigma_{22.1} \beta)^{-1}).$$

The asymptotic distribution of the likelihood ratio test statistic (2.5) is given by

$$-2lnQ(H_r | H_p) \xrightarrow{w} tr\{\int_0^1 (dB) dG' [\int_0^1 GG' du]^{-1} \int_0^1 G(dB)'\}$$

where G is given by (2.15) and (2.16).

It is an important consequence of Theorem 6 that the estimator $\hat{\beta}$, derived as if there are no I(2) variables, still has a mixed Gaussian distribution, since it means that the tests on restrictions derived as if there are no I(2) variables, remain valid under $H_{r,s}$. It is also seen that the limit distribution of T_r, if there are I(2) variables, is not given by (2.6). Simulations of the distribution given in Theorem 6 indicate that the tails are fatter than those derived from (2.6), indicating that the size of the tests used to determine the cointegration rank is too large. We shall show below how one can modify the test procedure to get tests with reasonable properties.

Combining the results of Theorem 5 with the asymptotic behaviour of the estimators in Theorem 6 we can then prove

THEOREM 7. *The results of Theorem 5 remain valid if α and β are replaced by the estimates given in (2.4).*

This result shows that in evaluating the distribution of $T_{r,s}$ under the hypothesis $H^0_{r,s}$, with α and β estimated, we need only apply the limit distributions already derived for I(1) variables.

The procedure of estimating r and s is investigated in the next theorem, and there the price for not calculating the maximum likelihood estimator in model $H_{r,s}$ and the likelihood ratio test against H_p is apparent in (2.17).

THEOREM 8. *Under the assumption of Theorem 3, that is if $H^0_{r,s}$ holds then*

$$\lim P\{(\hat{r},\hat{s}) = (i,j)\} = 0 \text{ if } i < r \text{ or } i = r \text{ and } j < s$$

$$\lim P\{(\hat{r},\hat{s}) = (r,s)\} = P\{T_r \leq c_r(\alpha), T_{r,s} \leq c_{r,s}(\alpha)\} \geq 1-2\alpha. \tag{2.17}$$

Finally we get the correct size for the test if we test for no I(2) variables:

COROLLARY 9. *The test that accepts r cointegrating relations and no I(2) variables if*

$$T_i \geq c_i(\alpha)\ , \ i = 0,\dots,r-1,\ T_r < c_r(\alpha)$$

and

$$T_{r,i} \geq c_{r,i}(\alpha),\ i = 0,\dots,p-r-1,$$

has asymptotic size α and asymptotic power 1. Here T_i, $T_{i,j}$ are given by (2.5) and (2.12) respectively, and the quantiles are calculated from the distribution (2.6) tabulated by simulation see Johansen (1988a).

The proof of these results will be given in the next section.

3. Asymptotic analysis

We first give some technical results about the behaviour of the process, the residuals and the product moment matrices. These will not be proved since they follow easily from the representation (1.12). Note that the normalization below ensures that the initial values will loose their importance in the limit.

LEMMA 1. *Under the assumptions of Theorem 3, and for $T \to \infty$, it holds that*

$$T^{-3/2}\beta_2' X_{[Tt]} \xrightarrow{w} \beta_2' C_2 \int_0^t W_u du,$$

$$T^{-\frac{1}{2}}\Delta X_{[Tt]} \xrightarrow{w} C_2 W_t,$$

$$T^{-\frac{1}{2}}(\beta,\beta_1)' X_{[Tt]} \xrightarrow{w} (\beta,\beta_1)' C_1 W_t,$$

$$T^{-2}\Sigma_1^T(\beta,\beta_1)' X_t \Delta X_t' \xrightarrow{w} (\beta,\beta_1)' C_1 \int W_t W_t' dt C_2',$$

$$T^{-2}\Sigma_1^T \Delta X_t \Delta X_t' \xrightarrow{w} C_2 \int W_t W_t' dt C_2',$$

such that

$$\beta' [\Sigma_1^T X_t \Delta X_t'][\Sigma_1^T \Delta X_t \Delta X_t']^{-1} C_2 \xrightarrow{P} \beta' C_1 = -\bar{\alpha}' \Gamma \bar{\beta}_2 \beta_2' C_2, \tag{3.1}$$

Here W_t is Brownian motion with variance Ω defined from the innovations ϵ_t. Notice the meaning of the different directions; in the direction β_2 the process is of the order 2, in the directions (β,β_1) the order is only 1, hence the different normalization. Note that (3.1) establishes the proof of Theorem 4.

LEMMA 2. *Under the assumptions of Theorem 3 and for $T \to \infty$ it holds that*

$$\beta' R_{2.1t} \xrightarrow{w} \text{stationary process}$$

$$T^{-\frac{1}{2}}\beta_1' R_{2.1t} \xrightarrow{w} F_{1t}$$

$$T^{-3/2}\beta_2' R_{2.1t} \xrightarrow{w} F_{2t}$$

such that

$$T^{-1}\Sigma_1^T \epsilon_t R_{2.1t}' \beta_1 \to \int_0^1 (dW) F_1'$$

$$T^{-2}\Sigma_1^T \epsilon_t R_{2.1t}' \beta_2 \to \int_0^1 (dW) F_2'$$

$$\beta' M_{22.1}\beta \xrightarrow{P} \beta' \Sigma_{22.1}\beta,\ \beta' M_{20.1} \xrightarrow{P} \beta' \Sigma_{20.1},\ M_{00.1} \xrightarrow{P} \Sigma_{00.1}$$

$$M_{11}(\beta,\beta_1),\ \beta' M_{22.1}\beta_1 \in O_P(1)$$

$$\beta_2' M_{11}\beta_2,\ \beta' M_{22.1}\beta_2 \in O_P(T)$$

$$(T^{-\frac{1}{2}}\beta_1, T^{-3/2}\beta_2)' M_{22.1} (T^{-\frac{1}{2}}\beta_1, T^{-3/2}\beta_2) \xrightarrow{w} \int FF' du.$$

This follows by considering the definition of the residuals and noting the different order of magnitude of the process in the directions $(\beta, \beta_1, \beta_2)$. Note that $\beta' R_{2.1t}$ is treated as a stationary process. We have here used the notation

$$\Sigma_{00.1} = \mathrm{Var}(\Delta^2 X_t \mid \Delta X_{t-1}, \Delta^2 X_{t-1}, \ldots, \Delta^2 X_{t-k+2}),$$
$$\beta' \Sigma_{20.1} = \mathrm{Cov}(\beta' X_{t-2}, \Delta^2 X_t \mid \Delta X_{t-1}, \Delta^2 X_{t-1}, \ldots, \Delta^2 X_{t-k+2}),$$
$$\beta' \Sigma_{22.1} \beta = \mathrm{Var}(\beta' X_{t-2} \mid \Delta X_{t-1}, \Delta^2 X_{t-1}, \ldots, \Delta^2 X_{t-k+2}).$$

Proof of Theorem 5.

We assume that r, α and β are known, and that α and β have rank r. The result follows then by the methods given in Johansen (1988a), since under the assumptions of Theorem 5 the process $Y_t = \Delta X_t$ is an I(1) process. The problem of determining the reduced rank of the coefficient matrix to $\beta_\perp' Y_{t-1}$ from the p–r equations given for $\alpha_\perp' \Delta^2 X_t = \alpha_\perp' \Delta Y_t$ in (2.10) is then exactly the problem solved in Johansen (1988a) except for the fact that on the right hand side we have the extra stationary variables $\beta' Y_{t-1} = \beta' \Delta X_{t-1}$ which are eliminated together with the short term dynamics as expressed by the coefficients Γ_i, i = 1,...,k–2, see (1.2). Thus (2.10) with known α and β corresponds to the situation considered in Johansen and Juselius (1991) where some cointegration vectors are assumed known. This establishes that the asymptotic distribution of the likelihood ratio test is as stated and also that the asymptotic distribution of the maximum likelihood estimators of the coefficients $\hat{\eta}$ is mixed Gaussian, whereas the remaining ones are asymptotically Gaussian. This completes the proof of Theorem 5.

Proof of Theorem 6.

The estimate of β is found by solving equation (2.3). We multiply the matrix in the equation by $(\beta, T^{-\frac{1}{2}}\beta_1, T^{-3/2}\beta_2)'$ and its transposed and then take the determinant. We can apply the results of Lemma 2 and find that for positive λ the limiting equation is

$$|\lambda\begin{bmatrix}\beta'\Sigma_{22.1}\beta & 0\\ 0 & \int FF'du\end{bmatrix} - \begin{bmatrix}\beta'\Sigma_{20.1}\\ 0\end{bmatrix}\Sigma_{00.1}^{-1}\begin{bmatrix}\beta'\Sigma_{20.1}\\ 0\end{bmatrix}'| = 0.$$

This shows that the r largest estimated eigenvalues converge to the solutions of

$$|\lambda\beta'\Sigma_{22.1}\beta - \beta'\Sigma_{20.1}\Sigma_{00.1}^{-1}\Sigma_{02.1}\beta| = 0,$$

and that the space spanned by the r first eigenvectors $sp(\hat{\beta})$ converges to the space spanned by β. In this sense $\hat{\beta}$ is consistent. From $\hat{\alpha} = M_{02.1}\hat{\beta}$ and the equation (2.2) we find the representation

(3.2) $T^{\frac{1}{2}}(\hat{\alpha} - \alpha) = T^{\frac{1}{2}}(M_{02.1}\hat{\beta}(\hat{\beta}'M_{22.1}\hat{\beta})^{-1} - M_{02.1}\beta(\beta'M_{22.1}\beta)^{-1}) + T^{-\frac{1}{2}}\Sigma_1^T\epsilon_t R_{2.1t}'\beta.$

The asymptotic stationarity of $\beta' R_{2.1t}$ and the consistency of $\hat{\beta}$ implies that the first term tend to zero, and the second tends to its expectation which is zero. This shows that $\hat{\alpha}$ is consistent.

We now find the asymptotic distributions. The estimate of α and β are found by first eliminating all parameters except α, β and the variance matrix Ω. In order not to overburden the notation we fix Ω and concentrate on α and β in the following. Results for Ω are easily derived in the same way. Then we let g denote minus the log of the partially maximized likelihood function $g(\alpha,\beta) = -\log L_{max}(\alpha,\beta)$ and find the derivatives in the directions h and k, which are p$\times$r matrices

$$g_\alpha(h) = tr\{\Omega^{-1}\,\Sigma_1^T\epsilon_t R_{2.1t}'\beta h'\}$$
$$g_\beta(h) = tr\{\Omega^{-1}\,\Sigma_1^T\epsilon_t R_{2.1t}' h\alpha'\} = tr\{\alpha'\Omega^{-1}\,\Sigma_1^T\epsilon_t R_{2.1t}' h\}$$

and second derivatives

$$g_{\alpha\alpha}(h,k) = tr\{\Omega^{-1}h\beta'\Sigma_1^T R_{2.1t}R_{2.1t}'\beta k'\} = Ttr\{\alpha'\Omega^{-1}h\beta' M_{22.1}\beta k'\}.$$
$$g_{\alpha\beta}(h,k) = tr\{\Omega^{-1}h\beta'\Sigma_1^T R_{2.1t}R_{2.1t}' k\alpha'\} = Ttr\{\alpha'\Omega^{-1}h\beta' M_{22.1}k\alpha'\}$$
$$g_{\beta\beta}(h,k) = tr\{\Omega^{-1}\alpha k'\Sigma_1^T R_{2.1t}R_{2.1t}' h\alpha'\} = Ttr\{\alpha'\Omega^{-1}\alpha k' M_{22.1}h'\}.$$

By decomposing $\hat{\beta}$ into the direction β and $\beta_\perp$ we see that

$$\hat{\beta} = \beta b + \beta_\perp c,$$

such that

(3.3) $$\hat{\beta}b^{-1} - \beta = \beta_\perp c b^{-1}$$

which shows that by suitably normalizing $\hat{\beta}$, which does not change $sp(\hat{\beta})$, we need only consider the deviations $\hat{\beta} - \beta$ in the directions $\beta_\perp = (\beta_1,\beta_2)$. Note that $\hat{\beta}b^{-1}$ gives a maximum of the likelihood function. In the following we shall see that the behaviour of the derivatives depend on which direction is considered. We therefore sometimes take $k = \beta_1 u_1$, where β_1 is $p{\times}s$ and u_1 is $s{\times}r$, or $k = \beta_2 u_2$, where β_2 is $p{\times}(p-r-s)$, and u_2 is $(p-r-s){\times}r$. Finally we also use the notation $\beta_0 = \beta u_0$, where u_0 is the identity matrix of dimensions $r{\times}r$. We then introduce these coordinates directly and find from Lemma 2 that the matrix of second derivatives satisfy that

$$g_{\beta\beta}(\beta,\beta),\ g_{\alpha\alpha}(h,k),\ g_{\alpha\beta}(h,\beta),\ g_{\alpha\beta}(h,\beta_1 u_1) \in O_P(T),$$
$$g_{\alpha\beta}(h,\beta_2 u_2) \in O_P(T^2),$$

and

$$g_{\beta\beta}(\beta_i u_i,\beta_j u_j) \in O_P(T^{i+j}) \quad i,j = 0,1,2,\ (i,j) \neq (0,0).$$

This implies that the proper normalization of the estimators is $T^{\frac{1}{2}}(\hat{\alpha}-\alpha)$, $T\beta_1'(\hat{\beta}-\beta)$, and $T^2\beta_2'(\hat{\beta}-\beta)$.

By expanding the function $(g_\alpha(h),g_\beta(h))$ around $\alpha = \hat{\alpha}$, and $\beta = \hat{\beta}$, we find that α satisfies

$$g_{\alpha\alpha}(\hat{\alpha} - \alpha,h) = g_\alpha(h), \text{ for all } h \ p{\times}r.$$

Similarly $\hat{\beta}$ satisfies

$$g_{\beta\beta}(\hat{\beta} - \beta,h) = g_\beta(h), \text{ for } h = \beta_1 u_1 \text{ and } h = \beta_2 u_2. \tag{3.4}$$

The equations separate since the off diagonal elements of the second derivative are small compared to the diagonal elements. This really has the consequence that in making asymptotic inference about α we can assume that β is fixed and vice versa.

Solving (3.4) for β we find the expression

$$\hat{\beta} - \beta = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}T^{-1}\Sigma_1^T \epsilon_t R_{2.1t}'M_{22.1}^{-1}.$$

or

$$(\hat{\beta} - \beta)'(T\bar{\beta}_1,T^2\bar{\beta}_2) = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\times$$
$$T^{-1}\Sigma_1^T \epsilon_t(R_{2.1t}'\beta_1,T^{-1}R_{2.1t}'\beta_1)\begin{bmatrix} T^{-1}\beta_1'M_{22.1}\beta_1 & T^{-2}\beta_1'M_{22.1}\beta_2 \\ T^{-2}\beta_2'M_{22.1}\beta_1 & T^{-3}\beta_2'M_{22.1}\beta_2 \end{bmatrix}^{-1}$$

We now apply the results of Lemma 2 to find that

$$(T(\hat{\beta} - \beta)'\bar{\beta}_1,T^2(\hat{\beta} - \beta)'\bar{\beta}_2)$$
$$\xrightarrow{w} (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}\int(dW)F'(\int FF'du)^{-1}.$$

Thus the estimator β is still super consistent, that is, T consistent in the directions β_1, where the process X_t is I(1), and T^2 consistent in the directions β_2, where the process X_t is I(2). This result holds for β normalized as discussed in (3.3). Results for other normalizations can then be derived, see Johansen (1990a). The main conclusion is, however, that the limit distribution is mixed Gaussian and this clearly holds for the other normalizations as well.

The asymptotic distribution of $\hat{\alpha}$ is found from (3.2) applying the super consistency of $\hat{\beta}$. We find

$$T^{\frac{1}{2}}(\hat{\alpha} - \alpha) = T^{-\frac{1}{2}}\Sigma_1^T \epsilon_t R_{2.1t}' \beta(\beta' M_{22.1}\beta)^{-1} + o_P(1),$$

which is asymptotically Gaussian with mean zero and variance $\Omega \otimes (\beta'\Sigma_{22.1}\beta)^{-1}$.

Finally we find the asymptotic distribution of the likelihood ratio test statistics given in (2.5). The result follow by the same method as in Johansen (1988a). We return again to the equation determining the eigenvalues (2.3), and find, letting $\lambda \to 0$ such that $\lambda T = \psi$ is constant, that for $T \to \infty$, we get by multiplying the matrix in the expression by $(\beta,\beta_1,T^{-1}\beta_2)$

$$\left| \begin{bmatrix} 0 & 0 \\ 0 & \psi\int FF'du \end{bmatrix} - \begin{bmatrix} \beta'\Sigma_{20.1} \\ K \end{bmatrix} \Sigma_{00.1}^{-1} \begin{bmatrix} \beta'\Sigma_{20.1} \\ K \end{bmatrix}' \right| = 0.$$

where K is the weak limit of $(\beta_1,T^{-1}\beta_2)'M_{20.1}$. Thus in the limit ψ will satisfy the equation

$$|\rho\int FF'du - KNK'| = 0,$$

where the matrix N is defined by

$$N = \Sigma_{00.1}^{-1} - \Sigma_{00.1}^{-1}\Sigma_{20}\beta(\beta'\Sigma_{20.1}\Sigma_{00.1}^{-1}\Sigma_{02.1}\beta)^{-1}\beta'\Sigma_{20.1}\Sigma_{00.1}^{-1}$$
$$= \alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp',$$

see Johansen (1990a) Lemma A.1. From (2.3) we then find that

$$(\beta_1,T^{-1}\beta_2)'M_{20.1}\alpha_\perp = T^{-1}\Sigma_1^T(\beta_1,T^{-1}\beta_2)'R_{2.1t}\epsilon_t'\alpha_\perp.$$

This converges by Lemma 2 towards

$$\int_0^1(F_1',F_2')(dW)'\alpha_\perp.$$

Collecting terms we find that the asymptotic distribution of the test statistic (2.5) is given by

$$tr\{[\int_0^1 FF'du]^{-1}\int_0^1 F(dW)'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\int_0^1(dW)F'\}.$$

The processes F and $\alpha_\perp'W$ only depend on W through the p–r linear combinations $B_2 = \Omega_{22}^{-\frac{1}{2}}\bar{\alpha}_2'W$ and $B_1 = \Omega_{11.2}^{-\frac{1}{2}}(\bar{\alpha}_1'W - \Omega_{12}\Omega_{22}^{-1}\bar{\alpha}_2'W)$, which have been normalized to have variance matrix I and to be independent. Here $\Omega_{22} = \bar{\alpha}_2'\Omega\bar{\alpha}_2$ and $\Omega_{11.2} = \bar{\alpha}_1'\Omega\bar{\alpha}_1 -$

$\bar{\alpha}_1'\Omega\alpha_2(\alpha_2'\Omega\alpha_2)^{-1}\alpha_2'\Omega\bar{\alpha}_1$. Note that the expression for F_2, see (2.14), is just

$$F_{2t} = \Theta^{-1}_{\alpha_2\beta_2}\Omega^{\frac{1}{2}}_{22}G_{2t},$$

and from (2.13)

$$\begin{aligned} F_{1t} &= \beta_1'C_1(W_t - [\textstyle\int_0^1 WW'du\alpha_2][\alpha_2'\int WW'du\alpha_2]^{-1}\alpha_2'W_t) \\ &= \Gamma^{-1}_{\alpha_1\beta_1}\bar{\alpha}_1(W_t - [\textstyle\int_0^1 WW'du\alpha_2][\alpha_2'\int WW'du\alpha_2]^{-1}\alpha_2'W_t) \\ &= \Gamma^{-1}_{\alpha_1\beta_1}\Omega^{\frac{1}{2}}_{11.2}G_{1t}, \end{aligned}$$

since $C_1\alpha = 0$. As the test statistics is invariant under linear transformation of F it is seen that the distribution does not depend on any parameters, and is given by the expression in Theorem 6. This completes the proof of Theorem 6.

Proof of Theorem 7.

The main idea of this proof is the well known result that if a sequence of random variables converge in distribution $X_n \overset{w}{\to} X$, and if $Y_n \overset{P}{\to} c$ then $Y_nX_n \overset{w}{\to} cX$. The asymptotic distribution of $\hat{\eta}$ is found from the equation (2.11) with β replaced by $\hat{\beta}$. Thus one can apply the above mentioned principle, except that the order of magnitude of the matrices is different in different directions. The eigenvalues are determined from the equation

$$|\rho M_{\hat{\beta}_\perp\hat{\beta}_\perp.\hat{\beta}} - M_{\hat{\beta}_\perp\hat{\alpha}_\perp.\hat{\beta}}M^{-1}_{\hat{\alpha}_\perp\hat{\alpha}_\perp.\hat{\beta}}M_{\hat{\alpha}_\perp\hat{\beta}_\perp.\hat{\beta}}| = 0. \tag{3.5}$$

We want to show that one can replace $\hat{\beta}$ and $\hat{\alpha}$ by β and α without changing the limit distribution. Hence one has to go through the various cases.

First $\hat{\alpha}_\perp \overset{P}{\to} \alpha_\perp$ and $M_{00} \overset{P}{\to} \Sigma_{00}$ implies that $\hat{\alpha}_\perp'M_{00}\hat{\alpha}_\perp \overset{P}{\to} \alpha_\perp'M_{00}\alpha_\perp$, and $\hat{\alpha}_\perp'M_{01} - \alpha_\perp'M_{01} \overset{P}{\to} 0$, which means that we can replace one by the other in the limit argument. Next consider the difference

$$\begin{aligned} &\hat{\beta}'M_{11}\hat{\beta} - \beta'M_{11}\beta \\ &= \beta'M_{11}(\hat{\beta}-\beta) + (\hat{\beta}-\beta)'M_{11}\beta + (\hat{\beta}-\beta)'M_{11}(\hat{\beta}-\beta). \end{aligned} \tag{3.6}$$

From Lemma 2 it follows that $(\beta,\beta_1)'M_{11}$ is $O_P(1)$, whereas $\beta_2'M_{11}\beta_2$ is $O_P(T)$. The difference $\hat{\beta}-\beta$ is concentrated on $sp(\beta_\perp) = sp(\beta_1,\beta_2)$, and $\hat{\beta}-\beta$ tends to zero in probability. This takes care of the first two terms in (3.6). The third term is expanded as

$$(\hat{\beta}-\beta)'M_{11}(\hat{\beta}-\beta)$$

$$= (\hat{\beta}-\beta)'(\bar{\beta}\beta' + \bar{\beta}_1\beta_1' + \bar{\beta}_2\beta_2')M_{11}(\beta\bar{\beta}' + \beta_1\bar{\beta}_1' + \beta_2\bar{\beta}_2')(\hat{\beta}-\beta).$$

Here all terms tend to zero in probability, even terms of the form

$$(\hat{\beta}-\beta)'\bar{\beta}_2\beta_2'M_{11}\beta_2\bar{\beta}_2'(\hat{\beta}-\beta), \tag{3.7}$$

since $(\hat{\beta}-\beta)'\beta_2 \in O_P(T^{-2})$.

In order to discuss matrices of the form $\hat{\beta}_\perp' M_{11}\hat{\beta}_\perp$ we need to choose $\hat{\beta}_\perp$ in a continuous fashion. The following representation is convenient $\hat{\beta}_\perp = \beta_\perp - \beta(\hat{\beta}'\beta)^{-1}\hat{\beta}'\beta_\perp$. This shows that $\hat{\beta}_\perp - \beta_\perp$ is in the space spanned by β. Hence from

$$\hat{\beta}_\perp' M_{11}\hat{\beta}_\perp - \beta_\perp' M_{11}\beta_\perp$$
$$= \beta_\perp' M_{11}(\hat{\beta}_\perp-\beta_\perp) + (\hat{\beta}_\perp-\beta_\perp)'M_{11}\beta_\perp + (\hat{\beta}_\perp-\beta_\perp)'M_{11}(\hat{\beta}_\perp-\beta_\perp).$$

one finds from the results in Lemma 2 that $\hat{\beta}_\perp' M_{11}\hat{\beta}_\perp - \beta_\perp' M_{11}\beta_\perp \overset{P}{\to} 0$.

Combining the results we find that the equation (3.5) has the same limit as the equation (2.11). This shows that the eigenvalues $\hat{\rho}_i$ from (3.5) have the same limit as those of (2.11) thus proving consistency of $\hat{\eta}$. Since the asymptotic distribution of $T_{r,s}$ is based upon the normalized eigenvalues $T\hat{\rho}_i$, $i = s+1,\dots,p-r$ one finds that even with estimated α and β the distribution of $T_{r,s}$ is given by (2.6) with p–r–s degrees of freedom.

The likelihood equation of η gives the following representation for the estimate

$$T(\hat{\eta}-\eta) = (T^{-1}M_{\hat{\beta}_\perp\hat{\beta}_\perp.\hat{\beta}})^{-1}(M_{\hat{\beta}_\perp\hat{\alpha}_\perp.\hat{\beta}} - M_{\hat{\beta}_\perp\hat{\beta}_\perp.\hat{\beta}}\eta\hat{\varphi}')\hat{\Omega}^{-1}\hat{\varphi}(\hat{\varphi}'\hat{\Omega}\hat{\varphi})^{-1}$$

The above arguments show that we can replace in this expression the estimated quantities by their limits without changing the limit distribution. Hence the limit law of $T(\hat{\eta}-\eta)$ is the same as if the coefficients α and β were known, i.e mixed Gaussian.

Note at this point that it was not enough that $\hat{\beta}$ was consistent. In order for the terms of the form (3.7) to be small we need that $T^{\frac{1}{2}}(\hat{\beta}-\beta) \overset{P}{\to} 0$. The full maximum likelihood estimator in the model $H_{r,s}$, which we do not treat here, has this property. Hence the analysis of the cointegrating rank and hypotheses on η seem to be equivalent to that based on the full likelihood methods. Inference for α and β, however, is expected to be less efficient in the approach suggested here.

Proof of Theorem 8

We define the critical sets $C_r = \{T_r \geq c_r(\alpha)\}$ and $C_{r,s} = \{T_{r,s} \geq c_{r,s}(\alpha)\}$. Then the definition of $\hat{r},\hat{s}$ shows that

$$P\{(\hat{r},\hat{s}) = (i,j)\} = P\{\cap_1^{i-1} C_k \cap C_i^c \cap_1^{j-1} C_{i,k} \cap C_{i,j}^c\},$$

The probability is calculated in the model $H^0_{r,s}$. For $i < r$, T_i is the sum of p–i smallest eigenvalues, and since even in the limit one of these is positive then the statistic T_i is forced to ∞. This shows that the probability tends to zero if $i < r$. Now let $i = r$. The same argument holds for $T_{r,j}$ if $j < s$, which again shows that the probability tends to zero. Finally let $i = r$ and $j = s$, then by taking limits we obtain

$$P\{(\hat{r},\hat{s}) = (i,j)\} \to P\{T_r \geq c_r(\alpha), T_{r,s} \geq c_{r,s}(\alpha)\}.$$

The last set has probability $1-\alpha$, and the first set has a probability derived from the limit distribution in Theorem 6. Simulations show that the first set has a probability $\geq 1-\alpha$ if in fact there are I(2) variables present, hence we can evaluate the probability down by $1-2\alpha$.

Proof of Corollary 9.

The proof follows by noting that if there are no I(2) variables then all eigenvalues ρ_i from (2.11) will be positive in the limit, which shows that all statistics $T_{r,s}$ will tend to ∞, hence the probability of accepting the hypothesis $H^0_{r,p-r}$

$$P\{C_0 \cap ... \cap C_{r-1} \cap C_r^c \cap C_{r,0} \cap ... \cap C_{r,p-r-1}\}$$

tends to

$$P\{C_r^c\} = P\{T_r < c_r(\alpha)\}$$

which is just $1-\alpha$ when calculated under the hypothesis that there are no I(2) variables. If we evaluate the probability under model $H^0_{r+1,j}$ then since $T_r \to \infty$ the acceptance probability tends to zero and the power tends to 1.

4. An example on the determination of the cointegration ranks.

In the previous section various results about the limit behaviour of the test statistics were derived. In this section we indicate how they are applied to make inference about the cointegration ranks r and s. As an example we consider the UK data analysed in Johansen and Juselius (1991). The data consists of quarterly observations from 1972.1 to 1987.2 of the five variables p_1 (a UK wholesale price index), p_2 (a trade weighted foreign wholesale price index), e_{12} (the UK effective exchange rate), i_1 (the three months treasury bill rate in UK), and i_2 (the three months Eurodollar interest rate). The detailed analysis in Johansen and Juselius (1991) involved fitting an autoregressive model with two lags, seasonal dummies and constant term since the data clearly indicated linear trends. In addition current and lagged values of the changes in the world oil price were included as strictly exogenous variables. It was further assumed that the processes were not I(2). This was justified by inspection of the graphs of the differences.

Thus the model analysed was not (1.2) but rather

$$\Delta^2 X_t = \Gamma\Delta X_{t-1} + \Pi X_{t-2} + \Sigma_1^{k-2}\Gamma_i\Delta^2 X_{t-i} + \Psi D_t + \mu + \gamma_0\Delta O_t + \gamma_1\Delta O_{t-1} + \epsilon_t,$$

where D_t are the seasonal dummies and O_t the oil price. From the representation (1.12) one finds by inserting $\Psi D_t + \mu + \gamma_0\Delta O_t + \gamma_1\Delta O_{t-1} + \epsilon_t$ instead of ϵ_t that X_t has a quadratic trend with coefficients $\tau_2 \in sp(\beta_2)$, if we assume for simplicity that O_t is described by a linear trend plus an I(1) process. In the space $sp(\beta_2)$ the process is dominated by the I(2) component except in the direction τ_2 where the quadratic trend is dominating. Similarly in $sp(\beta_1)$ there is a linear trend in a direction τ_1. Thus the process is dominated by the I(1) component in $sp(\beta_1)$ except in the direction τ_1, where the linear trend takes over. In order to formulate the limit result we therefore choose γ_i in $sp(\beta_i)$ orthogonal to τ_i, such that (τ_i,γ_i) span $sp(\beta_i)$. The limit result for $\hat\beta$ then has to be formulated as a result for the vector

$$T^{5/2}\tau_2'(\hat\beta-\beta),\ T^2\gamma_2'(\hat\beta-\beta),\ T^{3/2}\tau_1'(\hat\beta-\beta),\ T\gamma_1'(\hat\beta-\beta).$$

Again the limit distribution is mixed Gaussian. The limit distribution of the likelihood ratio test statistic is more involved and will not be discussed here, since it is not used. The main conclusion remains the same namely

1) That $\hat\alpha,\hat\beta$ are consistent and $\hat\beta$ super consistent, even if I(2) variables are present.

2) For fixed r,α, and β the analysis of (2.10) remains the same except that now there is a constant term in the equation.

3) For given r we can replace α and β by their estimates and analyse the equation (2.10), including a constant term, for cointegration and hence determine the cointegration rank s.

The influence of a constant term in (2.10) on the limit distribution of (2.12) is discussed in detail in Johansen (1990) and illustrated in Johansen and Juselius (1990).

We find under the assumption of the presence of a linear trend a limit distribution of the form

$$\int_0^1(dB)F'[\int_0^1 FF'du]^{-1}\int_0^1 F(dB)' \tag{4.1}$$

where F is a p–r–s dimensional process, with components defined by $F_i(t) = B_i(t) - \int_0^1 B_i(u)du$, $i = 1,\ldots,p-r-s-1$ and the last component defined by $F_{p-r-s}(t) = t - 1/2$. This is tabulated in Johansen and Juselius (1990) as Table A1.

As an illustration of the above technique we determine below the cointegration ranks r and s for the UK data, and show that there are no I(2) variables in the data.

The test statistic for $H_{r,s}$ in H_r^0 is given by (2.12) and has the limit distribution (4.1) if in fact the cointegrating rank is s, that is if $H_{r,s}^0$ is the true model and if condition (1.11), which rules out I(3) variables, holds. If $H_{r,s-1}^0$ is true, and this is part of the null hypothesis $H_{r,s}$ being tested the limit distribution of (2.12) is different, and this should be taken into account when evaluating the size of the test. It is not difficult to show that under $H_{r,s-1}^0$, say, the limit distribution of (2.12) is expressed as the p–r–s smallest eigenvalues of a matrix of the form (4.1), where B is a Brownian motion in p–r–s+1 dimensions.

Similarly if the linear trend is absent, but the estimation has taken it into account, the distribution of the statistic is different and has broader tails, see Johansen and Juselius (1990). This particular point is not so relevant for this example since we have explicitly included the trending oil prices in the model.

The approach taken here, and formulated in Theorem 8 and Corollary 9 is inspired by Pantula (1989) who made an important contribution towards formulating and solving this type of problem. The problem is also known in the statistical literature see Berger and Sinclair (1984). Briefly the idea can be described as follows: Consider a situation where a certain null hypothesis is being tested by the likelihood ratio test T, say. Sometimes it turns out that even asymptotically the distribution of the test statistic depends on the parameter in the null hypothesis. Let us assume that the null can be decomposed into a finite set of hypotheses Θ_i, $i = 1,\dots,n$, such that for $\vartheta \in \Theta_i$ the (limit) distribution of the test statistic is the same P_i, say. The idea of applying the critical level c, such that $\max_i P_i\{T \geq c\} = \alpha$, in general gives a test with a rather wide acceptance region, with poor power properties. Instead one can calculate the likelihood ratio test T_i for each sub hypothesis Θ_i, and apply the critical region $C = \cap_1^n \{T_i \geq c_i\}$, where c_i is determined from the (limit) distribution of T_i under the hypothesis $\vartheta \in \Theta_i$. See Johansen (1991) for detailed formulation of this idea and an application to the determination of the cointegrating rank for I(1) variables.

We apply this sets of ideas to an analysis of the UK data, and determine the cointegrating ranks r and s, where we assume that condition (1.11) holds to exclude I(3) variables.

Below we give the results of the calculation for the UK data. First the eigenvalue problem (2.3) is solved to determine for each r the test statistics T_r given by (2.5) and the estimate of α and β and hence $\alpha_\perp$ and $\beta_\perp$. For each value of r, and the appropriate estimated parameters we then eliminate $\beta' R_{1t}$ from equation (2.10) and solve (2.11) and calculate the test statistic $T_{r,s}$ given in (2.12). This gives the results in Table 2

Table 2.

The test statistics T_r and $T_{r,s}$ for the hypotheses H_r and $H_{r,s}$ in the I(2) model for the UK data.

p–r	r			$T_{r,s}$			T_r	$c_r(5\%)$
5	0	165.13	105.38	55.46	26.28	8.20	80.75	68.91
4	1		108.42	39.18	28.89	7.70	49.42	47.18
3	2			60.39	28.78	6.96	29.26	29.51
4	3				35.27	9.72	11.67	15.20
5	4					2.46	5.19	3.96
$c_r(5\%)$		68.91	47.18	29.51	15.20	3.96		
p–r–s		5	4	3	2	1	0	

The table is then read from top to bottom in the column for T_r. If we first assume that we do not have I(2) variables, then the last two columns give the test statistics, and the quantiles. The procedure then suggests starting at the top and reject $p-r = 5$ (or $r = 0$), similarly $p-r = 4$ is rejected, but, at the chosen level, $p-r = 3$ is not rejected. This then defines $\hat{r} = 2$.

Now suppose that we want to check for I(2) variables in the data. Then we continue with $r = 2$, and investigate the test statistics $T_{0,s}$, $s = 0,1,2$. They are compared with the corresponding quantiles given below the table. These quantiles are the same as are being applied for the test statistics T_r, by the result in Theorem 7. It is seen that, at the chosen level, $T_{2,0}$, $T_{2,1}$, and $T_{2,2}$ are all rejected, hence $\hat{s} = 3 = p-\hat{r}$ is accepted, corresponding to no I(2) components in the model.

6. Acknowledgement.

It is a great pleasure to thank Pentti Saikkonen for some interesting and technical discussions on the topic and the proofs of the paper. The motivation for the present work was as usual provided by Katarina Juselius.

5. *References.*

Ahn, S.K. and G.C. Reinsel, 1988, Nested reduced–rank autoregressive models for multiple time series, *Journal of the American Statistical Association*, 83, 849–856.

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

Berger, R.L and D.F. Sinclair, 1984, Testing hypotheses concerning unions of linear sub spaces, *Journal of the Americam Statistical Association,* 79, 158–163.

Engle, R.F. and C.W.J. Granger, 1987, Co–integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251–276.

Engle, R.F. and B. S. Yoo, 1989, Cointegrated economic time series: A survey with new results. UCSD Discussion paper.

Granger, C.W.J. and T–H. Lee, 1988, Investigations of Production, Sales and Inventory Relationships using multi cointegration and non symmetric error correction models. UCSD discussion paper # 38.

Johansen, S., 1988a, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, 12, 231–254.

Johansen, S., 1988b, The mathematical structure of error correction models, *Contemporary Mathematics*, 80 359–386.

Johansen, S. and K. Juselius, 1990, Maximum Likelihood estimation and Inference on Cointegration – with Applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

Johansen, S. and K. Juselius, 1991, Structural tests in a multivariate cointegration analysis of the PPP and UIP for UK, to appear in *Journal of Econometrics.*

Johansen, S., 1990a, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, to appear in *Econometrica.*

Johansen, S., 1990b, An algorithm for estimating the cointegration relations in vector autoregressive processes allowing for I(2) variables. University of Copenhagen

Johansen, S., 1990c, A representation of vector autoregressive processes integrated of order 2, to appear in *Econometric Theory.*

Johansen, S., 1991, Determination of cointegration rank in the presence of a linear trend, Preprint, University of Copenhagen.

Pantula, S.G., 1989, Testing for unit roots in time series data, *Econometric Theory*, 5, 256–271.

PREPRINTS 1990

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

No. 1 Johansen, Søren and Juselius, Katarina: Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK.

No. 2 Tjur, Tue: Analysis of Variance and Design of Experiments.

No. 3 Johansen, Søren: A Representation of Vector Autoregressive Processes Integrated of Order 2.

No. 4 Johansen, Søren: Cointegration in Partial Systems and the Efficiency of Single Equation Analysis.

PREPRINTS 1991

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

No. 1 Johansen, Søren: Determination of Cointegration Rank in the Presence of a Linear Trend.

No. 2 Johansen, Søren: A Statistical Analysis of Cointegration for I(2) Variables.