Cointegration analysis in the presence of structural breaks in the deterministic trend

Søren Johansen[1]

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

Rocco Mosconi

Dipartimento di Economica e Produzione, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy

Bent Nielsen

Department of Economics and Nuffield College, University of Oxford, Oxford OX1 1NF, UK

18 January 2000
Preliminary version

Abstract: When analysing macro economic data it is often of relevance to allow for structural breaks in the statistical analysis. In particular cointegration analysis in the presence of structural breaks could be of interest. To do this a vector autoregressive model is proposed with known break points in the structural breaks. Within this model it is possible to test cointegration rank, restrictions on the cointegrating vector as well as restrictions on for instance the slopes of broken linear trend.

1. Introduction

In the analysis of economic time series it is often necessary to include dummies representing breaks in the deterministic components. When allowing for breaks the timing is important, this could either be known in advance or an algorithm searching for breaks could be applied. While both issues are discussed in the literature and mainly in a univariate setting, this paper focuses on cointegration analysis in a multivariate setting in the presence of breaks at known points in time.

---

[1]Comments from Maria Cristina Leali, Graziano Viganò and Anders Rahbek are gratefully acknowledged. This paper replaces a manuscript by Johansen and Nielsen from 1993 with the title 'Asymptotics for cointegration rank tests in the presence of intervention dummies - manual for the simulation program DisCo'.

The suggested approach is a slight generalisation of the likelihood-based cointegration analysis in vector autoregressive models suggested by Johansen (1988, 1996). There are only few conceptual differences and the major issue for the practitioner is that new asymptotic tables are needed.

Structural breaks have been discussed intensely in the context of univariate autoregressive time series with a unit root. An important finding is that a time series given by stationary fluctuations around a broken constant level is better described by a random walk than a stationary time series, see Perron (1989, 1990) and Rappoport and Reichlin (1989). Addressing this issue, these authors suggested various univariate models allowing for breaks in the deterministic term. In particular, Perron (1989) suggested three models: (A) 'crash model' with change in intercept but unaffected slope of the linear trend, (B) 'changing growth model' with no change in intercept but changing slope of trend function, and (C) where both intercept and slope are changed at the time of the break. The model presented here generalises model (C) and allows for testing hypotheses corresponding to model (A).

A related concern in econometrics models is parameter stability which is investigated by methods related to those for known break points. These methods typically allow for structural breaks at unknown times, and have been discussed for instance in the special issues of the *Journal of Business & Economic Statistics*, volume 10, 1990 and the *Journal of Econometrics*, volume 70, 1996. More recently a test of this type has been suggested by Inoue (1999) in connection with cointegration testing in a vector autoregressive setting. While those authors and also this paper are concerned with breaks in the deterministic terms some procedures for analysing breaks in the cointegration parameter have been presented by Kuo (1998), Seo (1998) and Hansen and Johansen (1999).

The approach taken here is to analyse cointegration in a Gaussian vector autoregressive model with a broken linear trend with known break points. Likelihood analysis of cointegration is then given in terms of reduced rank regression, a combination of least squares regression analysis and canonical correlation analysis. The model and the rank hypothesis is discussed in Section 2. Section 3 presents tests for cointegration rank for models with broken trend, broken level and outlier dummies. The asymptotic distributions have been simulated and the results described by response surface analysis. Next, in Section 4 various tests for linear restrictions on the slopes for the broken trend are given. Most of these tests are asymptotically χ^2 distributed. In Section 5 the suggested procedures are illustrated using data for purchasing power parity between Italy and Germany.

Throughout the paper the following notational convention is used. For a matrix, a, with full column rank let $\bar{a} = a(a'a)^{-1}$. Further, let $a_\perp$ satisfy $a'_\perp a = 0$ and have the property that $(a, a_\perp)$ has full rank.

2. The Model

This section defines the statistical model. The observed time series is denoted X_t, $t = 1, \ldots, T$ and divided into sub-samples according to the position of the break points. For each sub-sample a vector autoregressive model is chosen, so that the parameters of the stochastic components are the same for all sub-samples, while the deterministic trend may change between sub-samples, so that the process can have breaks. In that case the process can be given rather simple representations and interpretations in each period and statistical analysis akin to that of the usual vector autoregressive models.

2.1. *Formulation of model and rank hypothesis*

The model allow for any pre-specified number of sample periods, q say, of length $T_j - T_{j-1}$ for $j = 1, \ldots, q$ and $0 = T_0 < T_1 < T_2 < \cdots < T_q = T$. It follows that the last observation in the j-th sample is T_j and, consequently, $T_j + 1$ is the first observation in sample period number $(j + 1)$. A vector autoregressive model of order k is considered. In analogy with the usual models without structural breaks the model is formulated conditionally on the first k observations of each sub-sample, that is $X_{T_{j-1}+1}, \ldots, X_{T_{j-1}+k}$ and it is given by the equations

$$\Delta X_t = (\Pi, \Pi_j) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \mu_j + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t \qquad (2.1)$$

for $j = 1, \ldots, q$ and $T_{j-1} + k < t \leq T_j$. The innovations are assumed to be independently, identically normal distributed with mean zero and variance Ω. The parameters vary freely, so that Π, Γ_i, Ω which relate to the stochastic component of the time series are the same in all sub-samples and of dimension $(p \times p)$ with Ω being symmetric and positive definite, while the p-vectors Π_j, μ_j relate to the deterministic component and they could be different in different sample periods.

A cointegration hypothesis can be formulated in terms of the rank of either Π alone or in conjunction with $\Pi_1, \ldots, \Pi_q$. The latter gives nicer interpretations

3

and some advantageous similarity properties and are given by

$$
H_l\left(r\right): \quad \operatorname{rank}\left(\Pi, \Pi_1, \ldots, \Pi_q\right) \leq r \quad \text{or} \quad \left(\Pi, \Pi_1, \ldots, \Pi_q\right) = \alpha \begin{pmatrix} \beta \\ \gamma_1 \\ \vdots \\ \gamma_q \end{pmatrix}',
$$

where the parameters vary freely so that α, β are of dimension $(p \times r)$ and γ_j is of dimension $(1 \times r)$. The notation H_l indicates that in each sub-sample the deterministic component is linear both for non-stationary and cointegrating relations. This feature will become evident from the Granger representation below. A related hypothesis arises in case of no linear trend but a broken constant level,

$$
H_c\left(r\right): \quad \operatorname{rank}\left(\Pi, \mu_1, \ldots, \mu_q\right) \leq r \quad \text{and} \quad \Pi_1, \ldots, \Pi_q = 0.
$$

As an alternative to the models H_l and H_c a rank hypothesis could be formulated for Π alone,

$$
H_{lc}\left(r\right): \quad \operatorname{rank} \Pi \leq r, \quad \text{and} \quad \Pi_1, \ldots, \Pi_q = 0.
$$

In terms of nesting this hypothesis is related to the previous as $H_c(r) \subset H_{lc}(r) \subset H_l(r)$. However, a hypothesis of the type H_{lc} is less attractive for two reasons. First, as indicated by the sub-index, the hypothesis H_{lc} implies that the common trends have a broken linear trend while the cointegrating relation has a broken constant level. Thus, the hypothesis questions both the stochastic and the deterministic behaviour of the process. In contrast, by choosing H_l the deterministic behaviour is not questioned in the rank determination. Secondly, in Section 3.3 it will be demonstrated that the asymptotic analysis is heavily burdened with nuisance parameters. These issues are discussed in further detail by Nielsen and Rahbek (2000).

2.2. *Another formulation*

The above description involves writing q model equations of the type (2.1). In order to write these as one equation which is more conformable with standard econometric computer packages some dummy variables are introduced. Let

$$
D_{j,t} = \begin{cases} 1 & \text{for } t = T_{j-1}, \\ 0 & \text{otherwise,} \end{cases} \quad j = 1, \ldots, q; \;\; t = \ldots, -1, 0, 1, \ldots,
$$

4

so that $D_{j,t-i}$ is an indicator function for the i-th observation in the j-th period. Further,

$$E_{j,t} = \sum_{i=k+1}^{T_j - T_{j-1}} D_{j,t-i} = \begin{cases} 1 & \text{for } T_{j-1} + k + 1 \le t \le T_j, \\ 0 & \text{otherwise,} \end{cases}$$

is the effective sample of the j-th period. It is convenient to gather the sample dummies and the drift parameters for the different sample periods:

$$\begin{aligned} E_t &= (E_{1,t}, \ldots, E_{q,t})', \\ \mu &= (\mu_1, \ldots, \mu_q), \\ \gamma &= (\gamma_1', \ldots, \gamma_q')', \end{aligned}$$

of dimensions $(q \times 1)$, $(p \times q)$ and $(q \times r)$ respectively. The model equation then becomes

$$\Delta X_t = \alpha \begin{pmatrix} \beta \\ \gamma \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ tE_t \end{pmatrix} + \mu E_t + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{k} \sum_{j=2}^{q} \kappa_{j,i} D_{j,t-i} + \varepsilon_t, \quad (2.2)$$

where the dummy parameters $\kappa_{j,i}$ are p-vectors and the observations $X_1, \ldots, X_k$ are held fixed as initial observations. Note, that the dummy variables $D_{j,t-1}, \ldots, D_{j,t-k}$ corresponds to the observations $X_{T_j+1}, \ldots, X_{T_j+k}$ which are held fixed above.

2.3. Interpretation

A process satisfying the hypothesis $H_l(r)$ can be interpreted using Granger's representation theorem. That is, a linear combination of the process, given by β, cointegrates while the process exhibits a linear trend in each of the sub-samples. As usual it is necessary to assume that the process actually is an I(1) process.

Assumption 1. *Assume that the roots of the characteristic polynomial,*

$$A(z) = (1-z) I_p - \alpha \beta' z - \sum_{i=1}^{k-1} \Gamma_i (1-z) z^i,$$

are outside the complex unit circle or at 1 and that the matrices α, β have full column rank r. Further, define $\Psi = I_p - \sum_{i=1}^{k-1} \Gamma_i$ and assume full rank of the matrix:

$$\alpha_\perp' \Psi \beta_\perp. \tag{2.3}$$

The Theorem 4.2 of Johansen (1996) can be generalised as follows.

Theorem 2.1. *Granger's Representation Theorem. Suppose Assumption 1 is satisfied. Then, for each period the initial values $X_{T_{j-1}+1}, \ldots, X_{T_{j-1}+k}$ can be given a distribution such that $\beta' X_t + \gamma_j' t$ and ΔX_t are stationary processes and all linear combinations of the process have a linear trend. In particular:*

$$X_t = C \sum_{i=T_{j-1}+k+1}^{t} \varepsilon_i + Y_{j,t} + \tau_{c,j} + \tau_{l,j} t \tag{2.4}$$

for $j = 1, \ldots, q$, $T_{j-1} + k < t \leq T_j$ and $C = \beta_\perp \left\{ \alpha_\perp' \Psi \beta_\perp \right\}^{-1} \alpha_\perp'$. The processes $Y_{j,t}$ are stationary, identically distributed and have zero expectation. The slope parameters, $\tau_{l,j}$ can be expressed as

$$\tau_{l,j} = C \mu_j + (C\Psi - I_p) \overline{\beta} \gamma_j',$$

whereas the level coefficient $\tau_{c,j}$ depends on initial values in such a way that $\beta' \tau_{c,j}$ is an identified function of the parameters:

$$\beta' \tau_{c,j} = \overline{\alpha}' \left(\Psi C - I_p \right) \mu_j + \overline{\alpha}' \left(\Psi C \Psi - \Psi \right) \overline{\beta} \gamma_j' - \gamma_j'.$$

Note, that for each sample period the process $\beta' X_t + t\gamma' E_t$ is a stationary process and hence it has no trending behaviour. Further, the process $\alpha_\perp' \Psi X_t$ is a combination of a random walk, a linear trend with slope $\alpha_\perp' \mu_j$ as well as a stationary component.

The representation shows that in each sub-sample all linear combinations of the process exhibit a linear trend, which generalises the model (C) suggested by Perron (1989). Tests for linear restrictions on the slope parameter τ_l are discussed in Section 4. The Granger Representation shows that the slope for the cointegrating vector, $\beta' \tau_l = -\gamma'$, has to be treated separately from the slope of the common trend, $\alpha_\perp' \Psi \tau_l = \alpha_\perp' \mu$. An example is the two period model, $q = 2$, with common slopes, $\tau_{l,1} = \tau_{l,2}$, which corresponds to Perron's model (A). In general these hypotheses are of the form, defining $\tau_l = \{\ldots, \tau_{l,j}, \ldots\}$,

$$H_l^\gamma (r): \qquad \beta' \tau_l G_\perp = -\gamma' G_\perp = 0 \qquad \text{or} \qquad \gamma = G\varphi \tag{2.5}$$

for the cointegrating relation and

$$H_l^\mu (r): \qquad \alpha_\perp' \Psi \tau_l M_\perp = \alpha_\perp' \mu M_\perp = 0$$

for the common trends. Here G and M are known matrices of dimension $(q \times g)$ and $(q \times m)$, respectively, with $g, m < q$ and full column rank. In particular Perron's model (A) is given by $G = M = (1, 1)'$, which means that $\gamma_1 = \gamma_2$ and $\alpha'_\perp \mu_1 = \alpha'_\perp \mu_2$.

A more subtle question is how the transition happens from one sample period to the next. The suggested conditioning on k or more initial values in each period allows for great flexibility and these transition periods can be extended if necessary. An alternative approach would be to use a specific transition function of some kind. Suppose it is of interest to model an instantaneous break. Using an unobserved components formulation one could then write

$$
\begin{aligned}
X_t &= \tau_{c,1} 1_{(t \le T_1)} + \tau_{c,2} 1_{(t > T_1)} + Z_t, \\
\Delta Z_t &= \alpha \beta' Z_{t-1} + \varepsilon_t.
\end{aligned}
$$

such that for $2 \le t \le T$,

$$
\Delta X_t = \alpha \beta' \left\{ X_{t-1} - \tau_{c,1} 1_{(t-1 \le T_1)} - \tau_{c,2} 1_{(t-1 > T_1)} \right\} + \tau_{c,1} \Delta 1_{(t \le T_1)} + \tau_{c,2} \Delta 1_{(t > T_1)} + \varepsilon_t.
$$

Using the definitions of $D_{j,t}$ and $E_{j,t}$, in particular that $D_{2,t} = 1_{(t=T_1)}$, $E_{1,t} = 1_{(t \le T_1)}$ and $E_{2,t} = 1_{(t \ge T_1+2)}$, it follows that

$$
\Delta X_t = \alpha \beta' \left(X_{t-1} - \tau_{c,1} E_{1,t} - \tau_{c,2} E_{2,t} \right) + \left\{ \tau_{c,2} - (I_p + \alpha \beta') \tau_{c,1} \right\} D_{2,t-1} + \varepsilon_t.
$$

Comparing with equation (2.2) or rather (3.6) it is found to be of the form

$$
\Delta X_t = \alpha \begin{pmatrix} \beta \\ \gamma \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ E_t \end{pmatrix} + \kappa_{2,1} D_{2,t-1} + \varepsilon_t
$$

with $\gamma'_j = -\beta' \tau_{c,j}$, for $j = 1, 2$, and $\kappa_{2,1} = \tau_{c,2} - (I_p + \alpha \beta') \tau_{c,1}$ which are restricted as

$$
\beta' \kappa_{2,1} = (I_r + \beta' \alpha) \gamma'_1 - \gamma'_2.
$$

This restriction on $\kappa_{2,1}$ is related to only one observation and is therefore difficult to test. For models of higher order the conditions for a instantaneous breaks would similarly involve all transition parameters $\kappa_{j,i}$. This issue is discussed in further detail for the univariate case by Perron (1990).

Another type of restriction of interest is co-breaking, see Hendry (1997) or Clements and Hendry (1999, p. 249-252). Co-breaking hypotheses are expressed in terms of the expected value of the process, and in particular for the cointegrating

7

relation, $\beta' X_t$, this is denoted equilibrium mean co-breaking. From the Granger's Representation Theorem 2.1 the expectation of the cointegrating vector is found to be

$$\beta' \left(\tau_c + \tau_l t \right) E_t = \left(\beta' \tau_c - \gamma' t \right) E_t.$$

If there exist an r-vector ω such that $\omega' \beta' \tau_l = -\omega' \gamma' = 0$ then the process is equilibrium mean co-breaking with ω as a co-breaking vector. Thus co-breaking is a hypothesis related to the column space of γ' or row space of γ. When $q < r$ or when the rank of γ otherwise is smaller than r then the process is indeed equilibrium mean co-breaking. In that case there are $(r- \text{rank}\, \gamma)$ co-breaking relations given by the $\{(r- \text{rank}\, \gamma) \times r\}$-matrix $\gamma_{\perp}$. The hypothesis formulated in (2.5), $\gamma = G\varphi$, relates to the column space of γ and hence it is not that helpful in finding co-breaking vectors. However, an upper bound for the rank of γ and thereby a lower bound for the number of co-breaking vectors is found. When g is smaller than r there are at least $(r - g)$ co-breaking vectors given by the $\{r \times (r- \text{rank}\, \varphi)\}$-parameter $(\varphi')_{\perp}$.

3. Test for Rank

The cointegration rank can be tested by modifying the procedures suggested by Johansen (1996). Whereas the statistical analysis is hardly unchanged the asymptotic results are related but different. New asymptotic distributions arise which are first formally described for three different cases: H_l: reduced rank of $(\Pi, \Pi_1, \ldots, \Pi_q)$, H_c: reduced rank of $(\Pi, \mu_1, \ldots, \mu_q)$ and $\Pi_1, \ldots, \Pi_q = 0$, and H_{lc}: reduced rank of Π while $\Pi_1, \ldots, \Pi_q = 0$. The analysis of the latter hypothesis is burdened with nuisance parameters and less useful than the first two. Models with outliers rather than breaks are closely related to these and will be analysed briefly. Finally, the asymptotic distributions related to H_l and H_c are described by response surface analysis which can easily be programmed.

For the suggested model the likelihood function can be maximised using canonical correlation methods as developed by Hotelling (1936), Bartlett (1937), Anderson (1951), and implemented in cointegration analysis by Johansen (1996, Section 6). In particular, in case of model H_l inference is based on squared sample canonical correlations, $1 \geq \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$, given by

$$\text{CanCor} \left\{ \Delta X_t, \left(\begin{array}{c} X_{t-1} \\ t E_t \end{array} \right) \middle| E_t, \Delta X_{t-i}, D_{j,t-i}, \begin{array}{c} i = 1, \ldots, k-1 \\ j = 2, \ldots, q \end{array} \right\},$$

where the notation indicates that ΔX_t and the vector of X_{t-1}, tE_t are first corrected for remaining terms in the model and next sample canonical correlations of the residuals are found. The likelihood ratio test statistic for the hypothesis of at most r cointegrating relations, $H_l(r)$, against $H_l(p)$ is given by

$$-T \sum_{i=r+1}^{p} \log\left(1 - \hat{\lambda}_i\right). \tag{3.1}$$

The asymptotic distribution of the test statistic is discussed below.

3.1. *Asymptotic distribution: a broken linear trend*

Inference should ideally be based on the exact distribution of the test statistic (3.1). Unfortunately this is not feasible so some kind of asymptotic distribution approximation has to be made. In order to ensure a good approximation the breaks need to be treated with care. The approach taken here is that the relative break points given by $v_j = T_j/T$ are fixed while an asymptotic argument in T is made.

For the asymptotic results the following notation is convenient. As defined above, let $v_j = T_j/T$ denote the relative break points so that $0 = v_0 < v_1 < \cdots < v_q = 1$, and let $\Delta v_j = v_j - v_{j-1}$. Further, define a q-dimensional vector of indicator functions for the sample periods, $e_u = \{\ldots, 1(v_{j-1} < u \le v_j), \ldots\}'$. This arises as the limit of E_{Tu} as T increases.

Theorem 3.1. *Suppose $H_l(r)$ and Assumption 1 are satisfied. Then the asymptotic distribution of the likelihood ratio test statistic for $H_l(r)$ against $H_l(p)$ is given by*

$$\mathrm{tr}\left\{\int_0^1 dW F' \left(\int_0^1 FF' du\right)^{-1} \int_0^1 F dW'\right\} \tag{3.2}$$

as $T \to \infty$ and for fixed relative break points, v_j. Here W is a standard Brownian motion of dimension $(p - r)$ and F is a $(p - r + q)$-dimensional process. For the j-th period, $v_{j-1} < u \le v_j$, the i-th coordinate of the process F is given by

$$F_i(u) = \begin{cases} W_i(u) - \frac{1}{\Delta v_j} \int_{v_{j-1}}^{v_j} W_i(s)\, ds & \text{for } i \le p - r, \\ u - \frac{1}{\Delta v_j} \int_{v_{j-1}}^{v_j} s\, ds & \text{for } i = p - r + j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

The asymptotic distribution has been simulated and the results analysed by a response surface analysis presented in Section 3.5.

9

For analytic reasoning and for computer simulations it is convenient to rewrite the representation of the distribution given by (3.2). This is based on two ideas. First, the distribution is invariant with respect to linear transformations of the vector process F. So if the first $(p-r)$ components are regressed on the last q components the transformed version of the matrix $(\int_0^1 FF'du)^{-1}$ is block-diagonal. It follows that the expression (3.2) can be rewritten as the sum of two terms which do not involve the levels of the Brownian motion, see (3.4) below. Secondly, when looking at a Brownian motion in two or more sub-samples while regressing on the level for each of the sub-samples, then the sub-sample Brownian motions are independent. This is also reflected in the representation (3.4) which involves q independent $(p-r)$-dimensional Brownian motions.

Theorem 3.2. *The asymptotic distribution* (3.2) *can be expressed as*

$$\mathrm{tr}\left[\left(\sum_{j=1}^{q} K_j \Delta v_j\right)' \left\{\sum_{j=1}^{q} L_j \left(\Delta v_j\right)^2\right\}^{-1} \left(\sum_{j=1}^{q} K_j \Delta v_j\right)\right] + \left(\sum_{j=1}^{q} J_j' J_j\right). \quad (3.4)$$

Here J_j, K_j, L_j *are defined as*

$$J_j = \left\{\int_0^1 \left(u - \frac{1}{2}\right)^2 du\right\}^{-1/2} \left\{\int_0^1 \left(u - \frac{1}{2}\right) \left(dW_u^{(j)}\right)\right\}$$

$$K_j = \int_0^1 \tilde{W}_u^{(j)} dW_u^{(j)},$$

$$L_j = \int_0^1 \left(\tilde{W}_u^{(j)}\right)^2 du,$$

where $W^{(1)}, \ldots, W^{(q)}$ *are independent* $(p-r)$-*dimensional standard Brownian motions and* $\tilde{W}^{(j)}$ *is* $W^{(j)}$ *corrected for a constant and a linear trend, that is*

$$\tilde{W}_u^{(j)} = W_u^{(j)} - \int_0^1 W_v^{(j)} dv - 12\left(u - \frac{1}{2}\right)\int_0^1 W_v^{(j)}\left(v - \frac{1}{2}\right) dv.$$

From the second representation of the limit distribution (3.4) it is seen that the asymptotic distribution only depends on the relative length of the sample periods, not on their ordering. For instance, in case of one break point, the asymptotic distribution is the same if $T_1 = T/3$ as if $T_1 = 2T/3$. Moreover, the first term in (3.4) is the trace of a $(p-r)$-dimensional square matrix while the second term is

the sum of inner products of $(p-r)$-dimensional vectors. This reflects the degrees of freedom arising from the matrix Π and the vectors Π_j, respectively.

This representation also shows another feature of the limit distribution. Let $DF_{q+1}(v_1, \ldots, v_q)$ denote the asymptotic in case of $q+1$ sample periods. When the length of one of sample periods tends to zero it follows that

$$\lim_{\Delta v_j \to 0} DF_{q+1}(v_1, \ldots, v_{q+1}) = DF_q(v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_{q+1}) + \chi^2(p-r), \quad (3.5)$$

where the DF_q and the χ^2 distributions are independent. The additional χ^2 term arises because the dimension of the vector (X'_{t-1}, tE'_t) is preserved although one of the relative sample length vanishes, and hence the dimension of the restrictions imposed by the rank hypothesis is unaltered. On the other hand if the dummies with the vanishing sample length are taken out of the statistical analysis, the additional χ^2-distributed element disappears.

The asymptotic distribution given above does not depend on the parameters for the deterministic component. The test is therefore asymptotically similar with respect to these parameters provided that Assumption 1 is satisfied, see also Nielsen and Rahbek (2000).

In order to estimate the rank a sequential testing procedure is necessary. One suggestion is to test the hypotheses

$$H_l(0), H_l(1), \ldots, H_l(p-1)$$

sequentially against the unrestricted model $H_l(p)$. If $H_l(r)$ is the first hypothesis to be accepted then the cointegrating rank is estimated by r. For consistency properties of this procedure see Johansen (1996, section 12.1).

3.2. *A broken constant level*

In some applications the level of the data may change from time to time but the data do not exhibit a linear trend. Then the model is given by

$$\Delta X_t = (\Pi, \mu) \begin{pmatrix} X_{t-1} \\ E_t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{k} \sum_{j=2}^{q} \kappa_{j,i} D_{j,t-i} + \varepsilon_t. \qquad (3.6)$$

The hypothesis of reduced cointegration rank is given by $H_c(r)$: $\text{rank}(\Pi, \mu) \leq r$, or equivalently that (Π, μ) can be written as $\alpha(\beta', \gamma')$, while the likelihood ratio test statistic for $H_c(r)$ against a general alternative, $H_c(p)$, is of the form (3.1). Further

the result of Theorem 3.1 applies with F replaced by a $(p - r + q)$-dimensional process with components:

$$F_i(u) = \begin{cases} W_i(u) & \text{for } i \leq p - r, \\ 1 & \text{for } i = p - r + j, \quad v_{j-1} < u \leq v_j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

3.3. On models with unrestricted parameters for the broken trend

The hypothesis $H_c(r)$ in the model (3.6) imposes rank restrictions on the first order autoregressive parameter as well as the parameter for the broken deterministic trend. In some situations it may seem reasonable to analyse a rank hypothesis which only involves the first order autoregressive parameter:

$$H_{lc}(r): \qquad \text{rank}\,\Pi \leq r \qquad \text{or} \qquad \Pi = \alpha\beta',$$

while μ is left unrestricted. This hypothesis is analysed by correcting ΔX_t and X_{t-1} for the remaining components in the model and subsequently performing a canonical correlation analysis of the residuals. The likelihood ratio test statistic for $H_{lc}(r)$ against $H_{lc}(p) = H_c(p)$ is of the form (3.1). Its asymptotic distribution is given as follows:

Theorem 3.3. *Suppose $H_{lc}(r)$ and Assumption 1 are satisfied. Let W be a $(p-r)$-dimensional standard Brownian motion and F the $(p-r+q)$-dimensional process given in Theorem 3.1. The asymptotic distribution as $T \to \infty$ of the likelihood ratio test statistic for $H_{lc}(r)$ against $H_{lc}(p)$ depends on $\alpha'_\perp \mu$, in particular, let $n = \text{rank}\,(\alpha'_\perp \mu)$.*
   (1) Suppose $n \geq (p-r)$. Then the test is asymptotically $\chi^2\{(p-r)^2\}$ distributed.
   (2) Suppose $n = q < (p - r)$. Then the asymptotic distribution is given by:

$$\text{tr}\left\{ \int_0^1 dW F'N' \left(N \int_0^1 FF'du N' \right)^{-1} N \int_0^1 F dW' \right\} \tag{3.8}$$

where N is a $(p - r) \times (p - r + q)$-dimensional matrix given by:

$$N = \begin{pmatrix} I_{p-r-q} & 0_{(p-r-q)\times q} & 0 \\ 0 & 0 & I_q \end{pmatrix}.$$

(3) Suppose $n < q < (p - r)$. Then there exist matrices ξ, η of rank n and dimensions $\{(p - r) \times n\}$ and $\{q \times n\}$ respectively such that $\alpha'_\perp \mu = \xi \eta'$. The asymptotic distribution is then given by (3.8) with N replaced by:

$$N = \begin{pmatrix} I_{p-r-n} & 0_{(p-r-h) \times n} & 0 \\ 0 & 0 & \eta' \end{pmatrix}.$$

The test is not as attractive as the previously considered tests. The limit distribution is a function of $\alpha'_\perp \mu$. The test is therefore not asymptotically similar with respect to the slope parameters for the broken trend. Moreover, it depends on $\alpha'_\perp \mu$ in a complicated way. Although the third situation in Theorem 3.3 only occurs on a null subset of the parameter space the issue ought to be addressed in the statistical analysis. Had there been no breaks in the trend the test strategy suggested by Johansen (1996, Section 12.2) could be used. A generalisation of that idea is not simple for two reasons. First, the limit distribution depends continuously on the nuisance parameter. Secondly, in many applications it could be of interest subsequently to test hypotheses corresponding to rank restrictions on $\alpha'_\perp \mu$. Such tests are discussed in section 4.

3.4. Models with outliers

Outlier dummies, or one period dummies are often included in econometric models. These are particularly used when there are extreme residuals which would indicate a violation of the assumption about normal innovations and their timing can often be associated with particular events in economic history. Such dummies can be included in two ways, either with an unrestricted parameter or alternatively the parameter could be restricted to be included in the cointegrating vector. The asymptotic distribution of the rank test is not affected in the first case, whereas it is indeed in the second case. On the other hand the first type of dummy implies a persistent effect whereas there is no long-run effect of the restricted dummy.

Outlier dummies with unrestricted parameters could be included in the model H_l in the following way. For notational convenience suppose there is only one of these, D_b, taking the value one at time T_b and zero otherwise. Here the time T_b is required to be in the effective sample, that is, it must not fall in $T_{j-1} + 1, \ldots, T_{j-1} + k$ for any j. Starting from the model equation (2.2) this results in a model given by

$$\Delta X_t = \alpha \begin{pmatrix} \beta \\ \gamma \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ tE_t \end{pmatrix} + \mu E_t + \nu D_b + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{k} \sum_{j=2}^{q} \kappa_{j,i} D_{j,t-i} + \varepsilon_t,$$

13

where ν is an unrestricted p-vector. Granger's Representation Theorem 2.1 indicates an additional common trends component, $C\nu 1(t \geq T_b)$, showing a persistent effect of the shock. Note, that for this argument, k observations after the time T_b need to be conditioned upon. When it comes to the asymptotic theory, the common trends are normalised by $\sqrt{T}$ and hence this additional component vanishes and there is no effect to the asymptotic distribution of the rank test. This shows that the rank test is asymptotically similar with respect to the parameter ν. Correspondingly, dummies for seasonal effects can be included as long as there is an unrestricted constant in the model. These issues are briefly discussed by Johansen (1996, Section 5.8).

Outlier dummies could alternatively be restricted to the cointegrating space in a model of the following type

$$\Delta X_t = \alpha \begin{pmatrix} \beta \\ \gamma \\ \delta \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ tE_t \\ D_b \end{pmatrix} + \mu E_t + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{k} \sum_{j=2}^{q} \kappa_{j,i} D_{j,t-i} + \varepsilon_t,$$

where δ' is an unrestricted r-vector. This extention has little effect on Granger's representation Theorem 2.1, since $\nu = \alpha \delta'$ and hence $C\nu = 0$. There is an effect to the asymptotic distribution of the rank test, simply because the sample canonical correlations are now taken of vectors of greater dimension, hence the asymptotic distribution is now of the type (3.5). A proof for a low dimensional situation is given by Doornik, Hendry and Nielsen (1998, Section 12.2.2). In applications this type of dummy is not so common, but it has been applied by for instance Doornik and Hendry (1994).

3.5. *Critical values for rank tests*

Exact analytic expressions for the asymptotic distributions are not known and the quantiles have to be determined by simulation. The asymptotic distributions depend on a number of factors: number of non-stationary relations, location of break points and the trend specification. The moments of these distributions have been approximated using a large number of simulations and a subsequent response surface analysis based on these factors. Then, the quantiles can be approximated using the empirical observation that the shape of rank test distributions typically are approximated rather well by Γ distributions, see Nielsen (1997) and Doornik (1998). Since the parameters of a Γ distribution are given by the first two moments, it suffices to report adequate approximations to the asymptotic mean and

variance of the trace test distributions. The quantiles can then be determined using a numerical routine for the incomplete Γ integral or a χ^2 distribution with non-integer degrees of freedom which is available in most statistical computer packages.

In the following the cases with a broken trend or a broken level are considered with up to three sample periods, $q = 3$. The cases with $q = 1, 2, 3$ can be described jointly. Let $v_j = T_j/T$ denote the break points as percentage of the full sample. For the case $q = 3$ there are three relative sample lengths, $v_1 - 0, v_2 - v_1, 1 - v_2$. Let a and b denote the smallest and the second smallest of these. For the case $q = 2$ there are two relative sample lengths $v_1 - 0, 1 - v_1$. Let b denote the smallest of these and let $a = 0$. Finally, for $q = 1$ let $a = b = 0$.

The moments of the asymptotic distributions are unknown functions of $(p-r)$, a, b. We have found that such functions are very precisely approximated by the polynomial:

$$
\log\left(\text{moment}\right) \approx f_{\text{moment}}\left\{(p - r), a, b, T\right\}
$$

$$
= \sum_{m=0}^{2}\left(\alpha_m + \sum_{i=1}^{4}\beta_{im}x_i + \sum_{i=1}^{4}\sum_{j\geq i}\gamma_{ijm}x_ix_j + \sum_{i=1}^{4}\sum_{j\geq i}\sum_{k\geq j}\delta_{ijkm}x_ix_jx_k\right)d_m
$$

(3.9)

where $x_1 = (p - r)$, $x_2 = a$, $x_3 = b$, $x_4 = T^{-1}$, $d_m = (p - r)^{-m}$. This function is essentially a third order polynomial in $(p - r)$, a, b and T^{-1}, where the terms in $(p-r)^{-1}$ and $(p-r)^{-2}$ play the same role as the dummies for dimensions 1, 2 and 3 used in Doornik (1998), but give a better fit. Notice that the regression includes the inverse of the sample size, T^{-1}. The role of the sample size in fitting response surfaces for the trace test is discussed in Doornik (1998). The asymptotic moments are easily calculated from (3.9) by letting $T \rightarrow \infty$. Notice also that $x_1d_1 = d_0 = 1$ and $x_1d_2 = d_1$. Some of the parameters in (3.9) are not identified, and therefore are set to zero. The remaining 75 parameters have been estimated by ordinary least squares, adding an error term to (3.9) and minimising the sum of squared residuals.

The moments of the asymptotic distribution were simulated for various values of $(p - r)$, a, b and T. The involved Brownian motions can be discretised in several ways. One possibility is to mimic the representation (3.2) and generate one random walk with T steps in each simulation, and associate a percentage of this to each sample period. In order to avoid poor approximations for cases with
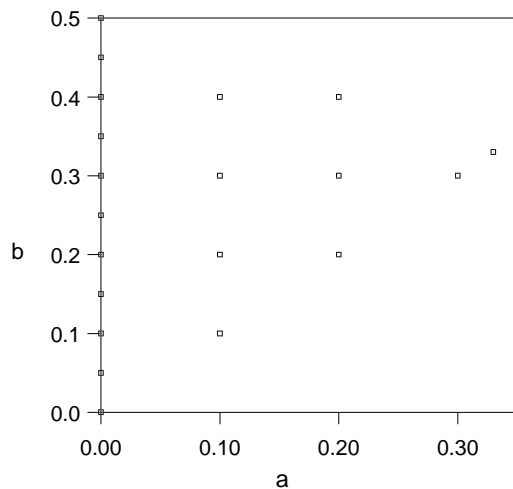
15

Figure 3.1: Values of a and b used in the simulations

	Unrestricted			Restricted		
	# Par.	R^2	σ	# Par.	R^2	σ
Model H_c, log (mean)	75	0.999997	0.00178	31	0.999996	0.00206
Model H_c, log (variance)	75	0.999963	0.00586	19	0.999936	0.00759
Model H_l, log (mean)	75	0.999998	0.00115	31	0.999996	0.00177
Model H_l, log (variance)	75	0.999940	0.00676	24	0.999894	0.00886

Table 3.1: Goodness of fit measures for the response surface

relatively short sample periods the representation (3.4) was used. The idea is to generate three random walks each with T steps and then scale them according to the relative lengths of the sample periods. The values of T were the integer part of $500/t$ for $t = 1, \ldots, 10$. The considered number of non-stationary relations was $(p - r) = 1, \ldots, 8$. Finally 20 different values of a and b were chosen as illustrated in Figure 3.1, to be representative of all pairs (a,b) such that $a < b$ and $b < (1 - a - b)$. Notice that there is a more dense sampling when $a = 0$, corresponding to one single break. This gives 1600 case which were repeated $N = 100,000$ times.

The fit of (3.9) is excellent, even when the number of parameters is dramatically reduced starting from the least significant, as illustrated in Table 3.1.

	Variance		
Mean	125/1.009	125	125*1.009
80/1.002	98.99	99.08	99.17
80	99.15	99.24	99.33
80*1.002	99.31	99.40	99.49

Table 3.2: 95th Percentiles of the Gamma distribution

	Variance		
Mean	125/1.009	125	125*1.009
80/1.002	.0480	.0487	.0494
80	.0493	.0500	.0507
80*1.002	.0505	.0513	.0520

Table 3.3: Right hand tail probability of Gamma distribution for the value 99.24

Standard errors are about 0.2% for the mean and 0.9% for the variance, an order of magnitude actually very close to the Monte Carlo sampling variation in $\log$ (moment). Notice that such small errors in the moments are virtually negligible for all practical purposes when computing the quantiles and tail probabilities of the Γ distribution. This point is illustrated in Tables 3.2 and 3.3, where quantiles and tail probabilities are computed for the values mean=80 and variance=125 and small variation thereof. These values of mean and variance are approximately equal to the average of the values found in our simulations.

It is important to remark that the residuals of (3.9) are approximately homoschedastic in all cases, which means that there are no values of $(p - r)$, a, b and T in which the errors are much bigger as a percentage of the moment. This is what motivated the choice to model the log of the moments rather than the moments themselves. We also tried to model the moments directly, without taking the logs, using weighted least squares with $(p - r)^{-2}$ as weights to account for heteroschedasticity, along the lines of Doornik (1998). However, the fit appears to be slightly poorer with that specification.

The estimated coefficients are reported in Table 3.4 , where the coefficients referred to the variable $x_4 = T^{-1}$ are not reported, since they are irrelevant for computing the asymptotic moments.

Notice that when a or b or both are zero, that is when $q = 0, 1$, the asymptotic moments may still be computed by (3.9), but the formula has to be corrected

17

	H_c		H_l	
	log (mean)	log (variance)	log (mean)	log (variance)
Constant	2.80	3.78	3.06	3.97
$(p-r)$	0.501	0.346	0.456	0.314
a	1.43	0.859	1.47	1.79
b	0.399		0.993	0.256
$(p-r)^2$	-0.0309	-0.0106	-0.0269	-0.00898
$(p-r)\,a$	-0.0600	-0.0339	-0.0363	-0.0688
$(p-r)\,b$			-0.0195	
a^2	-5.72	-2.35	-4.21	-4.08
ab	-1.12			
b^2	-1.70		-2.35	
$(p-r)^3$	0.000974		0.000840	
$(p-r)\,a^2$	0.168			
a^3	6.34	3.95	6.01	4.75
ab^2	1.89			
$a^2 b$			-1.33	
b^3	1.85	-0.282	2.04	-0.587
$(p-r)^{-1}$	-2.19	-2.73	-2.05	-2.47
$a\,(p-r)^{-1}$	-0.438	0.874	-0.304	1.62
$b\,(p-r)^{-1}$	1.79	2.36	1.06	3.13
$a^2\,(p-r)^{-1}$	6.03	-2.88	9.35	-4.52
$ab\,(p-r)^{-1}$	3.08		3.82	-1.21
$b^2\,(p-r)^{-1}$	-1.97	-4.44	2.12	-5.87
$a^3\,(p-r)^{-1}$	-8.08		-22.8	
$ab^2\,(p-r)^{-1}$	-5.79		-7.15	
$b^3\,(p-r)^{-1}$		4.31	-4.95	4.89
$(p-r)^{-2}$	0.717	1.02	0.681	0.874
$b\,(p-r)^{-2}$	-1.29	-0.807	-0.828	-0.865
$a^2\,(p-r)^{-2}$	-1.52		-5.43	
$b^2\,(p-r)^{-2}$	2.87			
$a^3\,(p-r)^{-2}$			13.1	
$b^3\,(p-r)^{-2}$	-2.03		1.50	

Table 3.4: Estimated surface responses

	mean of H_c		variance of H_c		95th percentile of H_c	
$(p-r)$	(3.10)	Doornik	(3.11)	Doornik	(3.10)-(3.11)	Doornik
1	4.1	4.1	7.0	6.7	9.2	9.1
2	12.0	12.1	19.6	20.0	20.1	20.3
3	24.2	24.0	38.5	38.6	35.2	35.0
4	40.2	40.0	63.2	63.2	54.1	53.9
5	60.2	60.1	94.0	93.8	77.0	76.9
6	84.1	84.1	131.1	130.4	103.8	103.7
7	111.9	112.1	174.2	173.6	134.5	134.6
8	142.8	144.1	222.6	221.6	169.2	169.4
9	180.3	180.1	274.9	276.2	208.4	208.3
10	222.5	220.1	329.3	336.8	253.2	251.1

Table 3.5: Comparison with Doornik (1998) (a=b=q=0)

using result (3.5), as follows:

$$\text{mean} \approx \exp\left\{f_{\text{mean}}\left[(p-r),a,b,\infty\right]\right\} - (2-q)(p-r) \qquad (3.10)$$
$$\text{variance} \approx \exp\left\{f_{\text{variance}}\left[(p-r),a,b,\infty\right]\right\} - 2(2-q)(p-r) \qquad (3.11)$$

These formulae allow to make a comparison with the approximation formulae computed in Doornik (1998) for the case with no breaks. This is illustrated in Table 3.5 in the case of model H_c. The formulae, although quite different at first glance, give very similar results for the most relevant values of $(p-r)$, expecially as far as the rightmost quantiles are concerned. For $(p-r) > 10$, the differences become more relevant, and this is of course related to the fact that our approximation is based on surface responses with $(p-r) \leq 8$, and therefore should not be used for dimensions much bigger than 8. The same analysis has been done for H_l, with similar results.

4. Restrictions on the slope parameters

When the cointegrating rank is known it is usually desirable to test further restrictions on the parameters. In this section hypotheses on the slope of the deterministic trend will be considered. Recall from Theorem 2.1 that the slope parameter is given by:

$$\tau_{l,j} = C\mu_j + (C\Psi - I_p)\beta(\beta'\beta)^{-1}\gamma'_j$$

19

in the j-th sample period. In particular, the deterministic slope for the cointegrating relations, $\beta' X_t$, is therefore $\beta' \tau_{l,j} = -\gamma_j'$. In brief, the results are:

1. tests for linear restrictions on the slope for the cointegrating relation, $\beta' \tau_l$, are asymptotically χ^2-distributed.

2. tests for linear restrictions on the slope in general, τ_l, are asymptotically χ^2-distributed.

The two tests can be formed sequentially, by first imposing a restriction on the slope for the cointegrating relation, $\beta' \tau_l$, and then, if accepted, imposing the same restriction on the common trend, $\alpha_\perp' \Psi \tau_l$. In this way it is possible to impose more restrictions on the slope for the cointegrating relation, $\beta' \tau_l$ than on the slope in general, τ_l.

Note it is not straight forward to do these two tests in the opposite order. In that case the test for slope of the common trend is burdened with nuisance parameter. This is related to the issue that the common trends are not uniquely defined. Mathematically it is convenient to define common trends as either $\alpha_\perp' \Psi X_t$ or $\beta_\perp' X_t$, but it could be defined as a linear combination of any of these and the cointegrating relation and often there is no particular valuable interpretation to any of these.

4.1. *Slope of the cointegrating relation*

The slope for the linear trend in the cointegrating relation is given by the parameter γ. Linear restrictions on this parameter can be formulated as:

$$H_l^\gamma (r) : \quad \gamma = G\varphi,$$

where G is a known $(q \times g)$-matrix of rank g, where $g \leq q$, and the parameter φ is a $(g \times r)$-matrix. Under the hypothesis the slope for the cointegrating relations is therefore

$$\beta' \tau_l E_t = -\varphi' G' E_t.$$

As an example suppose $q = 2$. By the choice $G = (1, 0)'$ the linear trend is absent in the second period whereas if $G = (1, 1)'$ then the slope is not altered by the break. It is convenient to introduce the notation $\beta^\gamma = (\beta, \varphi)'$ for the extended cointegration vector. Note, that when there is no cointegration, $r = 0$, then γ vanishes, hence $H_l^\gamma(0) = H_l(0)$.

As before the likelihood is maximised by canonical correlation analysis. The squared sample canonical correlations of the residuals, $1 \geq \hat{\lambda}_1^\gamma \geq \cdots \geq \hat{\lambda}_p^\gamma \geq 0$, are given by

$$\text{CanCor} \left\{ \Delta X_t, \left(\begin{array}{c} X_{t-1} \\ tG'E_t \end{array} \right) \middle| E_t, \Delta X_{t-i}, D_{j,t-i}, \begin{array}{c} i = 1, \ldots, k-1 \\ j = 1, \ldots, q \end{array} \right\},$$

and the likelihood ratio test for the hypothesis $H_l^\gamma(r)$ in $H_l(r)$ is then given by:

$$LR\left\{ H_l^\gamma(r) \middle| H_l(r) \right\} = T \sum_{i=1}^r \log \left\{ \left(1 - \hat{\lambda}_i^\gamma \right) \middle/ \left(1 - \hat{\lambda}_i \right) \right\},$$

see Johansen (1996, Theorem 7.2). The asymptotic distribution of the test statistic is given as

Theorem 4.1. *Suppose $H_l^\gamma(r)$ and Assumption 1 are satisfied. Then the likelihood ratio test statistic for $H_l^\gamma(r)$ in $H_l(r)$ is asymptotically $\chi^2\{r(q-g)\}$-distributed.*

The restriction on the linear term of the cointegrating vector, $\gamma = G\varphi$, could be combined with for instance a linear restriction on the cointegrating vector itself, $\beta = H\psi$, where H is a known, full-rank, $(p \times h)$-matrix and the parameter ψ is of dimension $(h \times r)$. The likelihood ratio test statistic for this hypothesis is also asymptotically χ^2. The degrees of freedom is $\{r(p-h)\}$ if the alternative is $H_l^\gamma(r)$ and $\{r(p - h + q - g)\}$ if $H_l(r)$ is the alternative.

4.2. *Slope of the common trends*

The common trends is a linear combinations of the levels of the time series which does not cointegrate. They could be chosen in various ways, for instance as $\beta_\perp' X_t$ or $\alpha_\perp' \Psi X_t$, and are essentially given by the random walk component $\alpha_\perp' \sum_{i=1}^t (\varepsilon_i + \mu E_i)$ in (2.4). In correspondence with the previous sub-section it is of interest to impose linear restrictions on $\alpha_\perp' \mu$ given by a $(q \times m)$-matrix M where $m \leq q$. With the formulation

$$H_l^\mu(r): \quad \mu = \zeta M' + \alpha \delta' M_\perp'$$

where ζ, δ are of dimension $(p \times m)$ and $\{(q-m) \times r\}$ respectively, then $\alpha' \mu$ is left unrestricted whereas $\alpha_\perp' \mu$ is restricted so that $\alpha_\perp' \mu M_\perp = 0$. Now, the common trends can be chosen is various ways. On the one hand if the common trends

21

are chosen mathematically convenient as $\alpha'_\perp \Psi X_t$ then the hypothesis imposes the restriction

$$\alpha'_\perp \Psi \tau_l = \alpha'_\perp \mu = \alpha'_\perp \zeta M'.$$

On the other hand for common trends given by $\beta'_\perp X_t$ the slope is

$$\begin{aligned}
\beta'_\perp \tau_l &= \beta'_\perp C\mu + \beta'_\perp C\Psi\beta \left(\beta'\beta\right)^{-1} \gamma' \\
&= \beta'_\perp C\zeta M' + \beta'_\perp C\Psi\beta \left(\beta'\beta\right)^{-1} \varphi'G',
\end{aligned}$$

showing that the slope of the common trends is linked intimately with the slope of the cointegrating relation. Thus if span$(G) \subset$span(M) then the restriction by M applies for all definitions of common trends, in fact it applies to the entire process, so that under that condition $\tau_l M_\perp = 0$. In the following the likelihood ratio test for $H_l^\mu(r)$ in $H_l^\gamma(r)$ is discussed. For the asymptotic analysis the restriction $G \in$ span M is crucial for avoiding nuisance parameters. Note, that in the unrestricted model when there are up to p cointegrating relations then the hypothesis H_l^μ entails no restriction on the parameters, hence $H_l^\mu(p) = H_l(p)$.

The squared sample canonical correlations of the residuals, $1 \geq \hat{\lambda}_1^\mu \geq \cdots \geq \hat{\lambda}_p^\mu \geq 0$, are now based on

$$\text{CanCor} \left\{ \Delta X_t, \begin{pmatrix} X_{t-1} \\ tG'E_t \\ M'_\perp E_t \end{pmatrix} \middle| M'E_t, \Delta X_{t-i}, D_{j,t-i}, \begin{matrix} i = 1, \ldots, k-1 \\ j = 1, \ldots, q \end{matrix} \right\}, \quad (4.1)$$

and the likelihood ratio test statistic for the hypothesis $H_l^\mu(r)$ in $H_l^\gamma(r)$ is given by

$$LR\left\{ H_l^\mu(r) | H_l^\gamma(r) \right\} = T \sum_{i=r+1}^{p} \log\left(1 - \hat{\lambda}_i^\gamma\right) / \left(1 - \hat{\lambda}_i^\mu\right),$$

see Johansen (1996, Theorem 6.2).

For the asymptotic analysis it is important that the restrictions to the common trends slope also apply to the cointegrating relations slope, that is $G \in$ span M.

Theorem 4.2. *Asymptotic distribution of likelihood ratio test statistic for $H_l^\mu(r)$ in $H_l^\gamma(r)$. Suppose $H_l^\mu(r)$ and Assumption 1 are satisfied. If in addition $G \in$ span M then the likelihood ratio test statistic for $H_l^\mu(r)$ in $H_l^\gamma(r)$ is asymptotically $\chi^2\{(p-r)(q-m)\}$.*

On the other hand if $G \notin$ span M then the asymptotic distribution of the test statistic involves nuisance parameters.

5. Empirical Illustration

This section illustrates the suggested statistical analysis, applied to a five dimensional data set with variables relevant for analysing the Uncovered Interest Parity (UIP) hypothesis between Germany and Italy. The economic model is very simple, and should be regarded as an illustration rather than a contribution to the ongoing economic debate. The analysis have been done using MALCOLM 2.4 (Mosconi, 1998), where all the techniques illustrated in this paper are implemented in a user friendly menu driven environment.

Let us consider the vector

$$Y_t = \left[\Delta p_t^I, \Delta p_t^D, \Delta e_t, i_t^I, i_t^D \right]$$

where Δp_t^I and Δp_t^D are first difference of log Consumer Price Index and represent inflation rates in Italy and Germany. The variable Δe_t represents the first difference of log nominal exchange rate between Italian Lira and German Mark (LIT/DM), while i_t^I and i_t^D are Italian and German nominal interest rates on long term treasury bonds, given as annual rates divided by 4, to make them dimensionally matching with the other variables. As for the sources, prices are from EUROSTAT (except 1973-1975, where prices are from UN - Monthly Bulletin of Statistics); notice that, after october 1990, German prices refer to unified Germany. Exchange Rates are from the Bank of Italy (average quarterly exchange rates). Interest Rates are from IMF, International Financial Statistics. The data are available from the authors.

The data, which are shown in Figure 5.1, are quarterly, ranging from 1973.2 to 1996.1 (T=92). When the UIP holds, assuming rational expectations, the following linear combination:

$$\Delta e_t - \left(i_{t-\ell}^I - i_{t-\ell}^D \right) \tag{5.1}$$

is zero for some lag $\ell \geq 0$. More weakly, we would expect it to be stationary with zero mean. Even more weakly, we might expect it to be stationary around some mean, which may be interpreted as a risk premium: a positive mean, say ρ, means that investors require $(i_{t-1}^I - i_{t-1}^D) > \rho > 0$ to move capital from Germany to Italy. The expected cointegration rank is at least equal to one, but of course it may be higher, since we do not have theoretical reasons to exclude more stationarity in the data.

To model these data, we use the model H_l, with two breaks. The last observation of the first period is 1979.4, while the last observation of the second
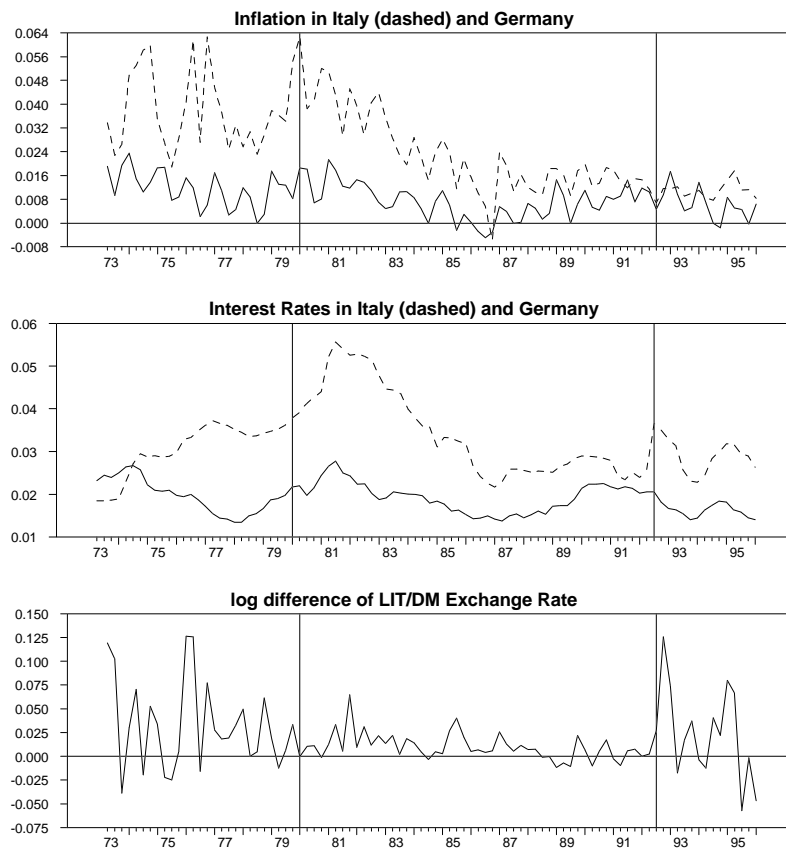
Figure 5.1: The data

24

k	Akaike	Hannan-Quinn	Schwartz	Godfrey χ^2_{ar5} (125)
1	-53.81	-53.07	-51.97	0.005
2	-54.40	-53.26	-51.57	0.244
3	-54.44	-52.90	-50.61	0.717
4	-54.83	-52.89	-50.01	0.915
5	-55.13	-52.79	-49.32	0.971

Table 5.1: Maximum lag analysis (p-value for Godfrey test)

Equation	Skewness	Kurtosis	Sk+Kur
Δp_t^I	0.352	0.581	0.557
Δp_t^D	0.149	0.155	0.129
Δe_t	0.011	0.570	0.032
i_t^I	0.346	0.130	0.204
i_t^D	0.750	0.010	0.033
System	0.198	0.018	0.021

Table 5.2: Jarque-Bera Normality tests (p-values)

period is 1992.2 ($T_1 = 27$, $T_2 = 77$; $v_1 = 0.29$, $v_2 = 0.84$; $a = 0.16$, $b = 0.29$). The first break coincides with the creation of the EMS, but it is also supposed to catch the oil shock and the modification of the US monetary policy. The second break corresponds to the exit of Italy from the EMS, but also the unification of Germany.

The analysis to determine the maximum lag k is reported in Table 5.1 . The information criteria suggest different values of k, in which case it is common practice to prefer Hannan-Quinn criterion. Therefore, $k = 2$ has been selected, since it is also the first lag to give approximately white noise residuals, according to the Godfrey test.

Jarque-Bera normality tests, reported in Table 5.2, show some problems with skewness in the Δe_t equation and kurtosis in the i_t^D equation, so that, at the system level, normality is rejected. Due to the illustrative aim of this analysis we did not try to analyse these problems any further. Notice, however, that all residuals based misspecification tests, like Godfrey and Jarque-Bera, should be modified in the present setting to take into account that the first k residuals of each period are set to zero by the presence of the dummies $D_{j,t-i}$, whose purpose is to condition upon the first k observations of each period. This might partly explain the problems with kurtosis.

Hypothesis	Test	p-value
$r = 0$	274.73	0.000
$r \leq 1$	145.37	0.000
$r \leq 2$	75.64	0.022
$r \leq 3$	28.36	0.692
$r \leq 4$	8.82	0.861

Table 5.3: Rank Tests

Root	$r = 5$	$r = 3$
1	$0.76 + 0.15i$	1.00
2	$0.76 - 0.15i$	1.00
3	$0.06 - 0.76i$	$0.06 - 0.76i$
4	$0.06 + 0.76i$	$0.06 + 0.76i$
5	0.69	$0.47 - 0.20i$

Table 5.4: Characteristic roots of the models

The tests for cointegration rank are reported in Table 5.3. The analysis supports cointegration rank of $r = 3$, which is consistent with our prior expectation. Therefore, we estimate the model with $r = 3$. Table 5.4 reports the five largest characteristic roots for both the unrestricted and the restricted models, which seems to be consistent with the I(1) assumption rank $(\alpha'_\perp \Psi \beta_\perp) = (p - r)$.

Before trying to set up identifying restriction on the cointegration space, let us illustrate some interesting tests on the deterministic components. The slope of the deterministic trend for the cointegrating relation, γ, is a (3×3)-matrix in our example, whose i-th column represents the trend coefficients of the i-th stationary relation in the three different periods. Consider the matrices

$$G_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad G_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad , \quad G_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad , \quad G_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

and the restrictions

$$H_{l,i}^{\gamma}(3): \quad \gamma = G_i \varphi.$$

The first three set the trend coefficient to zero in all stationary relations in period 1, 2 and 3 respectively. The fourth restriction means that there was no break in any of the stationary components, that is, the trend slope did not change. Notice

Hypothesis	$\chi^2(n)$	n	p-value
$H_{l,1}^{\gamma}(3)$	5.11	3	0.164
$H_{l,2}^{\gamma}(3)$	7.07	3	0.070
$H_{l,3}^{\gamma}(3)$	31.44	3	0.000
$H_{l,4}^{\gamma}(3)$	35.28	6	0.000
$H_{l}^{\gamma\mu}(3)$			
$H_{1}^{I}(3)$	10.52	6	0.105
$H_{2}^{I}(3)$	14.34	13	0.351

Table 5.5: Test statistics on several hypotheses

that these hypotheses may be also written as

$$H_{l,i}^{\gamma}(3): \quad \beta^{\gamma} = \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & G_i \end{bmatrix} \begin{bmatrix} \beta \\ \varphi \end{bmatrix}$$

which is easily implemented in standard cointegration software. The results are given in Table 5.5, which shows that there is no evidence of a trend in stationary components in the first and second period (joint test of $H_{l,1}^{\gamma}(3)$ and $H_{l,2}^{\gamma}(3)$ also accepts: $\chi^2(6) = 8.8$, p-value $= 0.185$), but the trend can not be removed from the third period. Also, since $H_{l,4}^{\gamma}(3)$ is strongly rejected, there is evidence that the trend in the stationary components did change significantly between periods.

Let us now illustrate a joint hypothesis about the commond trends and the stationary components, namely

$$H_{l}^{\gamma\mu}(3) = \begin{cases} \mu = \zeta M' + \alpha\delta'M'_{\perp} \\ \gamma = M\varphi \end{cases}$$

with

$$M = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

When $H_{l}^{\gamma\mu}(3)$ holds, then $\tau_l M_{\perp} = 0$ (i.e. $\tau_l = \psi'M'$), so that the slope of both the stationary and the non stationary components remains the same in all periods. It is important to remark that, even under $H_{l}^{\gamma\mu}(3)$, breaks may occur in the constant term τ_c in both the stationary and non stationary components. Part of $H_{l}^{\gamma\mu}(3)$, involving the stationary components, has been already tested and rejected with $H_{l,4}^{\gamma}(3)$: as shown in Table 5.5, the joint test also ends up in a rejection.

27

In order to (over)-identify the cointegration space, we suggest the following stationary linear combinations:

$$
\begin{aligned}
z_{1t} &= \Delta e_t - \left(i_t^I - i_t^D\right) \\
z_{2t} &= \left(i_t^D - \Delta p_t^D\right) \\
z_{3t} &= \left(i_t^I - \Delta p_t^I\right) - \left(i_t^D - \Delta p_t^D\right)
\end{aligned}
$$

The first represents the UIP hypothesis: notice that the lag ℓ of (5.1) is not in this formula, but this is not important since Δi_t^I and Δi_t^D are stationary processes and, for example, $i_{t-1}^I = i_t^I - \Delta i_t^I$. The second linear combination represent the German real interest rate, while the third represents the real interest rate differential. Notice that, if these linear combinations are stationary, then also

$$
\begin{aligned}
z_{4t} &= z_{1t} + z_{3t} = \Delta e_t - \left(\Delta p_t^I - \Delta p_t^D\right) \\
z_{5t} &= z_{2t} + z_{3t} = \left(i_t^I - \Delta p_t^I\right)
\end{aligned}
$$

are stationary, and could be used to find an alternative and equivalent basis of the cointegration space. Identifying restrictions may be written as

$$
H_i^I(3) = \beta^\gamma = \left[\begin{array}{c} \beta \\ \gamma \end{array} \right] = [B_{1,i}b_1, B_{2,i}b_2, B_{3,i}b_3].
$$

In order to test the local trend stationarity of z_{1t}, z_{2t} and z_{3t}, we set up the following identifying restrictions:

$$
B_{1,1} = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}, \quad
B_{2,1} = \begin{bmatrix}
0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}, \quad
B_{3,1} = \begin{bmatrix}
-1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
$$

while, in order to place some plausible restrictions on the deterministic part, we

use the following restriction matrices:

$$
B_{1,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ -1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad B_{2,2} = \begin{bmatrix} 0 & 0 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad B_{3,2} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}
$$

which exclude the linear trend from z_{1t} and z_{2t} in the first and second periods, and from z_{3t} in all periods. As shown in Table 5.5, both $H_1^I(3)$ and $H_2^I(3)$ can not be rejected. Figure 5.2 represent z_{1t}, z_{2t} and z_{3t}, together with their deterministic components estimated under $H_2^I(3)$.

This shows that in the period $(79.1, 92.2)$, in which Italy belongs to the EMS, both z_{1t} and z_{3t} are approximately zero on average. The mean of z_{1t}, which is the distance from the UIP, is positive and quite large in the first period, meaning that the "risk premium" on Italy before the EMS was around 4% on annual basis. In the third period, the mean of z_{1t} is again positive, and impressively high immediately after Italy went out the EMS, but is negatively trending, reaching zero in 1995.
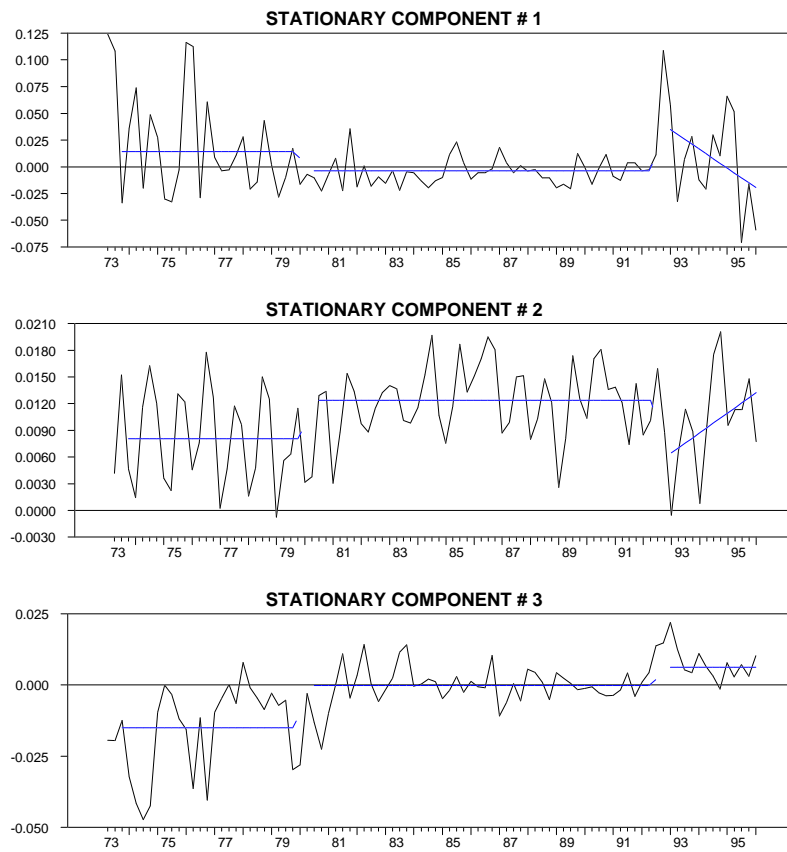
Figure 5.2: The stationary components

30

A. Appendix: Mathematical Details

The techniques in the proofs follow are closely related to those presented in the monograph by Johansen (1996) - which we in the following will refer to as (J). One major difference is that the monograph focuses on models, H_{lc}, where the parameters for the deterministic terms are unrestricted under the rank hypothesis, whereas here the focus is on models, H_l, where the deterministic component is not affected by the rank hypothesis. In the following, modified versions of the theorems and lemmas of (J) will be given and differences in the proofs will be pointed out.

A.1. *Proof of Theorem 2.1*

Each sub-sample can be considered separately because of the conditioning on the first k observations in each period. The representation (2.4) therefore follows from Theorem 4.2 (J). It is a consequence of his proof that the stochastic behaviour of the process does not depend on the deterministic terms and, hence, the stationary components of (2.4), $Y_{j,t}$, are identically distributed with zero expectation. To find the representation of the slope parameter $\tau_{l,j}$ and the intercept parameter note that the processes $\beta'(X_t - \tau_c E_t - \tau_l E_t t)$ and $\Delta X_t - \tau_{l,t}$ have zero expectation taking expectations in (2.2) with these expressions. In order to obtain a model with zero level it must be that

$$\left(\begin{array}{c} \mu_j \\ \gamma_j' \end{array} \right) = \left(\begin{array}{cc} \Psi + \alpha\beta' & -\alpha \\ -\beta' & 0 \end{array} \right) \left(\begin{array}{c} \tau_{l,j} \\ \beta'\tau_{c,t} \end{array} \right).$$

The expressions in the Theorem then follows by noting that

$$\left(\begin{array}{cc} C & (C\Psi - I_p)\overline{\beta} \\ \overline{\alpha}'(\Psi C - I_p) & \overline{\alpha}'(\Psi C\Psi - \Psi)\overline{\beta} - I_r \end{array} \right) \left(\begin{array}{cc} \Psi + \alpha\beta' & -\alpha \\ -\beta' & 0 \end{array} \right) = I_{p+r}.$$

□

A.2. *Some asymptotic results*

The asymptotic properties of the residual product moment matrices are discussed under the hypothesis $H_l^\gamma(r)$ by modifying the Lemmae 10.2 and 10.3 of (J). One difference is that here an explicit study of the asymptotic behaviour of the process X is avoided. Further, the definitions of the relevant residuals and the necessary transformation matrices are somewhat different.

The statistical analysis is based on canonical correlation analysis. It is important to note that the sample canonical correlations are invariant with respect to linear transformations of the data, whereas this property is not shared by the corresponding canonical vectors which are used for estimating the cointegrating vector.

Since the data exhibit a linear trend according to Theorem 2.1 it is convenient to detrend the levels of the process, that is $(X'_{t-1}, tE'_t G)' = Q_T\{(X'_{t-1}|tE'_t G), tE'_t G\}'$ where

$$Q_T = \begin{pmatrix} I_p & \left(\sum_{t=1}^T X'_{t-1} E'_t G t\right) \left(\sum_{t=1}^T t G' E_t E'_t G t\right)^{-1} \\ 0 & I_g \end{pmatrix}.$$

Hence define the residual process

$$\begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix} = \left\{ \begin{pmatrix} \Delta X_t \\ X_{t-1}|t G' E_t \\ t G' E_t \end{pmatrix} \middle| E_t, \Delta X_{t-i}, D_{j,t-i}, \begin{array}{l} i = 1, \ldots, k-1 \\ j = 1, \ldots, q \end{array} \right\} \quad \text{(A.1)}$$

and the residual product moment matrices

$$\begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix} = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix} \begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix}'.$$

The squared sample canonical correlations, $1 \geq \lambda_1 \geq \cdots \geq \lambda_p \geq 0$ and $\lambda_i = 0$ for $i = p+1, \ldots, p+g$ of R_0 and R_1 are then given as solutions to the eigenvalue problem

$$\left| \lambda S_{11} - S_{10} S_{00}^{-1} S_{01} \right| = 0.$$

The corresponding eigenvectors, v_i, satisfy

$$S_{11} \lambda_i v_i = S_{10} S_{00}^{-1} S_{01} v_i.$$

The matrix of the r first eigenvectors is denoted $\hat{\beta}^0$ and the parameters β, φ are then estimated from the equation

$$Q'_T \hat{\beta}^\gamma = Q'_T \begin{pmatrix} \hat{\beta} \\ \hat{\varphi} \end{pmatrix} = \hat{\beta}^0.$$

A corresponding asymptotic relation can be established for the parameters using the representation in Theorem 2.1

$$Q'_T \beta^\gamma = Q'_T \begin{pmatrix} \beta \\ \varphi \end{pmatrix} \xrightarrow{\mathcal{P}} \begin{pmatrix} \beta \\ 0 \end{pmatrix} = \beta^0. \quad \text{(A.2)}$$

In the asymptotic analysis under $H_l(r)$ the residuals, $R_{1,t}$, are decomposed using three orthogonal matrices

$$\beta^0 = \begin{pmatrix} \beta \\ 0 \end{pmatrix}, \qquad B_1 = \begin{pmatrix} \beta_\perp \\ 0 \end{pmatrix}, \qquad B_2 = \begin{pmatrix} 0 \\ I_g \end{pmatrix}.$$

Under the hypothesis $H_l^\gamma(r)$ the asymptotic results also depends on the $\{(p-r) \times (q-g)\}$-dimensional parameter $\alpha'_\perp \mu \overline{G}_\perp$. If this has reduced rank, n say, there exist matrices ξ, η of dimension $\{(p-r) \times n\}$ and $\{(q-g) \times n\}$ respectively and with rank n such that $\beta'_\perp \tau_l \overline{G}_\perp = \xi \eta'$ and $\alpha'_\perp \mu \overline{G}_\perp = \alpha'_\perp \Psi \beta_\perp \xi \eta'$. Thus define

$$B_{11} = B_1 \xi_\perp \left(\xi'_\perp \beta'_\perp C \Omega C' \beta_\perp \xi_\perp \right)^{-1/2}, \qquad B_{12} = B_1 \overline{\xi}$$

which are orthogonal if $\beta_\perp$ is normalised so that $\beta'_\perp \beta_\perp = I_{p-r}$. Further, define

$$B_T = \left(B_{11}, B_{12} T^{-1/2}, B_2 T^{-1/2} \right), \qquad A_T = \left(\beta^\gamma, B_T T^{-1/2} \right),$$

where A_T has full rank and $\overline{B}_T = (\overline{B}_1, B_2 T^{1/2})$.

The asymptotic results are expressed in terms of two independent standard Brownian motions V, W of dimension r and $(p-r)$ respectively. These are the limits of the independent random walks $(\alpha' \Omega^{-1} \alpha)^{-1/2} \alpha' \Omega^{-1} \sum_{s=1}^t \varepsilon_s / T^{1/2}$ and $(\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \alpha'_\perp \sum_{s=1}^t \varepsilon_s / T^{1/2}$.

Lemma A.1. *Asymptotic properties of $B'_T R_1$. (J, Lemma 10.2)*
Suppose the hypothesis $H_l^\gamma(r)$ and the Assumption 1, are satisfied. Let F be the $(p-r+q)$-dimensional process given by (3.3) and define the $\{(p-r+g) \times (p-r+q)\}$-matrix

$$N = \begin{pmatrix} I_{p-r-n} & 0_{(p-r-n) \times n} & 0 \\ 0 & 0 & \eta' G'_\perp \left\{ I_q - ZG \left(G'ZG \right)^{-1} G' \right\} \\ 0 & 0 & G' \end{pmatrix}, \qquad (A.3)$$

where $G = \int_0^1 u^2 e_u e'_u du$. Then as $T \to \infty$

$$\frac{1}{\sqrt{T}} B'_T R_{1,[Tu]} \xrightarrow{D} N F_u. \qquad (A.4)$$

Note, that under the hypothesis $H_l(r)$ then $G = I_g$, $n = 0$, hence $N = I_{p-r+g}$. Under $H_{lc}(r)$ then $G = 0$, hence N reduces to the matrices given in Theorem 3.3.

The *proof* relies on the representation of the process X given in Theorem 2.1

$$\beta'_\perp X_{t-1} = \beta'_\perp C \sum_{i=T_{j-1}+k+1}^{t} \varepsilon_i E_{j,t} + t\beta'_\perp \tau_l E_t + O_\mathcal{P}(1) \tag{A.5}$$

for $T_{j-1} + k < t \le T_j$. Using the identity $I_g = \overline{G}G' + \overline{G}_\perp G'_\perp$ and the definition $\beta'_\perp \tau_l \overline{G}_\perp = \xi\eta'$ it follows that

$$\beta'_\perp X_{t-1} = \beta'_\perp C \sum_{i=T_{j-1}+k+1}^{t} \varepsilon_i E_{j,t} + t\xi\eta' G'_\perp E_t + t\beta'_\perp \tau_l \overline{G}G' + O_\mathcal{P}(1).$$

The result then follows by regressing on $tG'E_t$ and normalising this expression appropriately. $\qquad\square$

For each periode the processes ΔX_t and $\beta' X_{t-1} + t\varphi' G' E_t$ can be given stationary initial distributions, see Theorem 2.1. Apart from a changing level the stationary distributions are identical. Thus define

$$\begin{pmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{pmatrix} = \mathrm{Var}\left(\begin{array}{c|c} \Delta X_t & \begin{array}{cc} & i = 1,\ldots,k-1 \\ \Delta X_{t-i}, D_{j,t-i}, & j = 1,\ldots,q \end{array} \end{array} \right).$$

Lemma A.2. *Asymptotic behaviour of S_{ij}. (J, Lemma 10.3)*
Suppose the hypothesis $H_l^\gamma(r)$ and the Assumption 1 are satisfied. Then as $T \to \infty$

$$\begin{pmatrix} S_{00} & S_{01}\beta^0 \\ \beta^{0'}S_{10} & \beta^{0'}S_{11}\beta^0 \end{pmatrix} \xrightarrow{\mathcal{P}} \begin{pmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{pmatrix}. \tag{A.6}$$

This asymptotic covariance matrix satisfies the identity

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}\left(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right)^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1} = \alpha_\perp\left(\alpha'_\perp \Omega \alpha_\perp\right)^{-1}\alpha'_\perp. \tag{A.7}$$

Moreover, in the non-stationary directions the product moment matrices satisfy

$$T^{-1}B'_T S_{11} B_T \xrightarrow{\mathcal{D}} N \int_0^1 F_u F'_u du N', \tag{A.8}$$

$$\begin{pmatrix} (\alpha'\Omega^{-1}\alpha)^{-1/2}\alpha'\Omega^{-1} \\ (\alpha'_\perp \Omega \alpha_\perp)^{-1/2}\alpha'_\perp \end{pmatrix} S_{\varepsilon 1} B_T \xrightarrow{\mathcal{D}} \int_0^1 \begin{pmatrix} dV_u \\ dW_u \end{pmatrix} F'_u N', \tag{A.9}$$

$$\left(B'_T S_{10}, B'_T S_{11}\beta^0\right) = O_\mathcal{P}(1), \tag{A.10}$$

where $S_{\varepsilon 1} = S_{01} - \alpha\beta^{\gamma'}Q_T S_{11}$.

The *proof* of the equations (A.8)-(A.10) mimicks that of Lemma 10.3 (J), where Lemma A.5 is used instead of Lemma 10.2 (J). The equation (A.6) follows by noting that for $t \in E_{j,t}$, the representation (2.4) implies:

$$\begin{aligned} \Delta X_t &= C\varepsilon_t + \Delta Y_{j,t} + \tau_{l,j}, \\ \beta' X_{t-1} + t\varphi' G' E_t &= \beta'\left(Y_{j,t} + \tau_c E_t + t\tau_l E_t\right). \end{aligned}$$

The distribution of $Y_{j,t}$ is the same in all periods. Consequently, the processes $\beta' X_{t-1} + t\varphi' G' E_t$ and ΔX_t are stationary and the conditional variance, (A.6), is the same in all periods. Further

$$\left(\beta' X_{t-1} \mid tG' E_t\right) = \left(\beta' X_{t-1} + t\varphi' G' E_t \mid tG' E_t\right)$$

and hence the limit in (A.6) is expressed in terms of the long-run variance of $\beta' X_{t-1} + t\varphi' G' E_t$. Finally, the equation (A.7) follows from Theorem 10.1 (J). $\square$

A.3. *Proof of Theorem 3.1*

The proof follows that of Theorem 11.1 (J). The only difference is that Lemma A.2 with $G = I_g$ is used instead of Lemma 10.3 (J) and that some of the notation is changed. That is, the terms β, G, B, W should be replaced with the terms β^γ, F, W, $\Omega^{1/2} W$. Note that the arguments between equations (11.20), (11.21) (J) are redundant with the present model formulation and that the proof requires that the identity (A.7) is satisfied. $\square$

A.4. *Proof of Theorem 3.3*

The proof resembles that of Theorem 3.1 with the difference that $G = 0$, so that $g = 0$ and $G_\perp = I_q$. The asymptotic theory therefore depends on the rank of $\alpha'_\perp \mu \overline{G}_\perp = \alpha'_\perp \mu$. In particular if $\mathrm{rank}(\alpha'_\perp \mu) = (p - r)$ then the function NF is deterministic, see (A.4). The asymptotic distribution is therefore χ^2. $\square$

A.5. *Asymptotic properties of $\hat{\beta}$ under $H_l^\gamma(r)$*

The proof of Theorem 4.1 uses the asymptotic properties of $\hat{\beta}^0$. To discuss these it is convenient to consider a normalisation depending on the unknown parameter β^0

$$\tilde{\beta}^0 = \hat{\beta}^0 \left(\overline{\beta}^{0\prime} \hat{\beta}^0\right)^{-1}.$$

Since the matrix (β^0, B_T) has full rank and orthogonal blocks, $\beta^{0\prime} B_T = 0$, it follows that $I_{p+g} = \beta^0 \overline{\beta^0}' + B_T \overline{B_T}'$. Consequently

$$\tilde{\beta}^0 = \beta^0 + B_T \overline{B_T}' \tilde{\beta}^0 = \beta^0 + B_T U_T \qquad \text{where} \qquad U_T = \overline{B_T}' \tilde{\beta}^0.$$

Correspondingly, define $\tilde{\alpha} = \hat{\alpha} \hat{\beta}^{0\prime} \overline{\beta}^0$ so that $\tilde{\alpha} \tilde{\beta}^{0\prime} = \hat{\alpha} \hat{\beta}^{0\prime}$. Modified versions of the results of Lemma 13.1,13.2, Theorem 13.8 (J) then hold.

Lemma A.3. *Asymptotic behaviour of S_{ij} (J, Lemma 13.1)*
Suppose the hypothesis $H_l^{\gamma}(r)$ and the Assumption 1 are satisfied. Then the estimators under the hypotheses, $\tilde{\beta}^0$, $\tilde{\alpha}$, $\hat{\Omega}$, are consistent and

$$\begin{aligned}
\tilde{\beta}^{0\prime} S_{11} \tilde{\beta}^0 &= \beta^{0\prime} S_{11} \beta^0 + o_{\mathcal{P}}(1), \\
\tilde{\beta}^{0\prime} S_{10} &= \beta^{0\prime} S_{10} + o_{\mathcal{P}}(T^{-1/2}).
\end{aligned}$$

The *proof* mimicks that of (J). $\qquad\qquad\square$

Lemma A.4. *Asymptotic distribution $\tilde{\beta}^0$. (J, Lemma 13.2)*
The estimator $\tilde{\beta}^0$ is mixed Gaussian in the sense

$$TU_T = \overline{B_T}' \left(\tilde{\beta}^0 - \beta^0 \right) \xrightarrow{\mathcal{D}} \left(N \int_0^1 F_u F_u' du N' \right)^{-1} N \int_0^1 F_u dV_u' \left(\alpha' \Omega^{-1} \alpha \right)^{-1/2}.$$

where F,N is given by (3.3),(A.3).

The *proof* is unaltered. Note that $(\alpha'\Omega^{-1}\alpha)^{1/2} V_\alpha = (\alpha'\Omega^{-1}\alpha)^{-1/2} \alpha'\Omega^{-1} W$ in (J) is denoted V here. $\qquad\qquad\square$

In the Lemma A.6 below the test for a simple hypothesis on β^γ is discussed. For the proof both of the parameters β^γ and β^0 are referred to. Thus recall the above definition of the residual product moment matrices S_{ij} where the levels of the process are detrended. Correspondingly let S_{ij}^γ be the residual product moment matrices where the levels are not detrended, hence,

$$\begin{pmatrix} S_{00}^\gamma & S_{01}^\gamma \\ S_{10}^\gamma & S_{11}^\gamma \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ 0 & Q_T \end{pmatrix} \begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & Q_T' \end{pmatrix}.$$

36

Lemma A.5. *Detrending of levels residuals.*
Suppose the hypothesis $H_l^\gamma(r)$ and the Assumption 1 are satisfied. Then

$$\hat{\beta}^{\gamma\prime} S_{10}^\gamma = \hat{\beta}^{0\prime} S_{10}, \qquad \hat{\beta}^{\gamma\prime} S_{11}^\gamma \hat{\beta}^\gamma = \hat{\beta}^{0\prime} S_{11} \hat{\beta}^0, \qquad (A.11)$$

$$\beta^{\gamma\prime} S_{10}^\gamma = \beta^{0\prime} S_{10} + o_{\mathcal{P}}(1), \qquad \beta^{\gamma\prime} S_{11}^\gamma \beta^\gamma = \beta^{0\prime} S_{11} \beta^0 + o_{\mathcal{P}}(1). \qquad (A.12)$$

The *proof.* The identities (A.11) follows from the eigenvalue problem described between equations A.1),(A.2). For (A.12) combine the approximation (A.2) and the results of Lemma A.2. For instance,

$$\beta^{\gamma\prime} S_{10}^\gamma = \beta^{\gamma\prime} Q_T S_{10} = \left\{ Q_T' \beta^\gamma - \beta^0 \right\}' S_{10} + \beta^{0\prime} S_{10}$$

where $Q_T' \beta^\gamma - \beta^0$ converges in probability to zero by (A.2) and S_{10} is of stochastic order one by (A.6) and (A.10). $\qquad \square$

Lemma A.6. *Test for simple hypothesis on β^γ. (J, Lemma 13.8)*
The likelihood ratio test statistic for a simple hypothesis on β^γ against the hypothesis $H_l(r)$ is asymptotically distributed as the random variable:

$$\mathrm{tr} \left\{ \int_0^1 dV F' N' \left(N \int_0^1 FF' du N' \right)^{-1} N \int_0^1 F dV' \right\}.$$

This variable has a $\chi^2 \{r(p - r + g)\}$ distribution.

The *proof* is slighly different from that given in (J). First, as in (J) the likelihood function is expanded around β^γ. This is done in terms of $\tilde{\beta}^\gamma$ and S_{ij}^γ. Using Lemma A.5 is then possible to replace β^γ, $\tilde{\beta}^\gamma$ and S_{ij}^γ with β^0, $\tilde{\beta}^0$ and S_{ij} without changing the asymptotic results. The remaining of the arguments of (J) then apply. The degrees of freedom is given by the product of $\dim V = r$ and $\dim(NF) = p - r + g$.

A.6. *Proof of Theorem 4.1*

The proof follows that of Theorem 13.9 (J) using Lemma A.6 instead of Lemma 13.8 (J) and where the (random) matrices

$$N' \left(\int_0^1 FF' \right)^{1/2}, \qquad \int_0^1 dV F \left(\int_0^1 FF' \right)^{-1/2}$$

replace his matrices M' and Z. The degrees of freedom is given by the product of $\dim V = r$ and $\dim(N_\perp F) = (q - g)$. $\qquad \square$

A.7. *Asymptotic results under the hypothesis $H_l^\mu(r)$*

For the asymptotic analysis of $H_l^\mu(r)$ the assumption $G \in \operatorname{span} M$ is important. Had this not been satisfied nuisance parameters would appear. The easiest example of that phenomenen is seen in a model without break, $q = 1$, two lags, $k = 2$, a linear trend for the cointegrating relation, $G = I_1 = 1$. Tests of the hypothesis that the common trends slope is absent, $M = 0$, leads to inference burdened with nuisance parameters.

The likelihood ratio test statistic for $H_l^\mu(r)$ in $H_l^\gamma(r)$ is based on the sample canonical correlations given in (4.1). Due to the invariance properties of canonical correlations it is equivalent to consider

$$
\operatorname{CanCor} \left\{ \Delta X_t, \begin{pmatrix} X_{t-1}|tG'E_t, M_\perp'E_t \\ tG'E_t|M_\perp'E_t \\ M_\perp'E_t \end{pmatrix} \middle| M'E_t, \Delta X_{t-i}, D_{j,t-i}, \begin{matrix} i = 1, \ldots, k-1 \\ j = 1, \ldots, q \end{matrix} \right\}
$$

and corresponding sample product moment matrices, S_{ij}^μ say. The asymptotic behaviour of S_{ij}^μ is partly described in Lemma A.2. To see this define the residuals

$$
\begin{pmatrix} r_{0,t} \\ r_{1,t} \end{pmatrix} = \left\{ \begin{pmatrix} \Delta X_t \\ M_\perp'E_t \end{pmatrix} \middle| M'E_t, \Delta X_{t-i}, D_{j,t-i}, \begin{matrix} i = 1, \ldots, k-1 \\ j = 1, \ldots, q \end{matrix} \right\}
$$

and the product moment matrices

$$
s_{01} = \frac{1}{T} \sum_{t=1}^{T} r_{0,t} r_{1,t}', \qquad s_{11} = \frac{1}{T} \sum_{t=1}^{T} r_{1,t} r_{1,t}'.
$$

It follows that

$$
\begin{pmatrix} S_{00}^\mu & S_{01}^\mu \\ S_{10}^\mu & S_{11}^\mu \end{pmatrix} = \left\{ \begin{matrix} S_{00} + s_{01} s_{11}^{-1} s_{10} & \begin{pmatrix} S_{01} & s_{01} \end{pmatrix} \\ \begin{pmatrix} S_{10} \\ s_{10} \end{pmatrix} & \begin{pmatrix} S_{11} & 0 \\ 0 & s_{11} \end{pmatrix} \end{matrix} \right\}.
$$

The asymptotic properties of S_{ij} are given in Lemma A.2, whereas s_{ij} has the following properties:

Lemma A.7. *Asymptotics for s_{ij}.*
Suppose the hypothesis $H_l^\mu(r)$ and the Assumption 1 are satisfied. Further, assume that either $G \in \operatorname{span} M$ or $k = 1$. Define the $(q - m)$-dimensional function

$$
f_u = M_\perp'e_u - M_\perp' \int_0^1 e_u e_u' du M \left(M' \int_0^1 e_u e_u' du M \right)^{-1} M'e_u.
$$

Then

$$T^{1/2} s_{01} \quad = \quad O_{\mathcal{P}}(1), \tag{A.13}$$

$$s_{11} \xrightarrow{\mathcal{P}} \int_0^1 f_u f_u' du, \tag{A.14}$$

$$T^{1/2} s_{10} \alpha_\perp (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \xrightarrow{\mathcal{D}} \int_0^1 f_u dW_u'. \tag{A.15}$$

It then follows that $S_{00}^\mu \xrightarrow{\mathcal{P}} \Sigma_{00}$ *and hence*

$$(S_{00}^\mu)^{-1} - (S_{00}^\mu)^{-1} S_{01}^\mu \beta^\mu \left\{ \beta^{\mu\prime} S_{10}^\mu (S_{00}^\mu)^{-1} S_{01}^\mu \beta^\mu \right\}^{-1} \beta^{\mu\prime} S_{10}^\mu (S_{00}^\mu)^{-1}$$

$$\xrightarrow{\mathcal{P}} \alpha_\perp (\alpha_\perp' \Omega \alpha_\perp)^{-1} \alpha_\perp', \tag{A.16}$$

where β^μ *is the* $\{(p+g+q-m) \times r\}$*-matrix* $(\beta', 0)'$*. Note, that when* $G \notin \operatorname{span} M$ *and the slope parameter satisfies* $\tau_l M_\perp \neq 0$ *then* $s_{01} = O_{\mathcal{P}}(1)$ *and the limit in (A.16) involves nuisance parameters.*

For the *proof* the main observations is that for a stationary process, Z_t, with zero mean:

$$\sum_{t=1}^T E_t E_t' = O(T), \qquad \sum_{t=1}^T E_t Z_t' = O_{\mathcal{P}}\left(\sqrt{T}\right). \tag{A.17}$$

In order to use those results the differenced process need to be analysed carefully. The Theorem 2.1 and H_l^μ imply the representation

$$\Delta X_t = C\varepsilon_t + \Delta Y_t E_t + \tau_l E_t,$$

where

$$\tau_l = C\zeta M' + (C\Psi - I_p)\overline{\beta}\varphi'G'.$$

If $G \in \operatorname{span} M$ then $G'M_\perp = 0$ and $\tau_l = \theta M'$ for some parameter θ. It follows that the residuals of ΔX_t corrected for $M'E_t$ equal the residuals of the zero mean stationary process $C\varepsilon_t + \Delta Y_t E_t$ corrected for $M'E_t$, that is

$$(\Delta X_t | M'E_t) = (C\varepsilon_t + \Delta Y_t E_t | M'E_t). \tag{A.18}$$

For the asymptotic analysis of s_{11} the regression on lagged differences can be ignored which is seen from (A.17)-(A.18), hence

$$s_{11} = \frac{1}{T} \sum_{t=1}^T (M_\perp' E_t | M'E_t, D_{j,t-i}) (M_\perp' E_t | M'E_t, D_{j,t-i})' + o_{\mathcal{P}}(1). \tag{A.19}$$

For asymptotic purposes also the regression on the dummies can be ignored and (A.14) follows. For the asymptotic analysis of s_{10} note that (A.18) implies that

$$s_{10} = \sum_{t=1}^{T} \left(M_\perp' E_t \right) \left(C\varepsilon_t + \Delta Y_t E_t \middle| M' E_t, C\varepsilon_{t-i} + \Delta Y_{t-i} E_{t-i}, D_{j,t-i} \right)'. \qquad \text{(A.20)}$$

Then (A.17) shows that s_{10} is of order $T^{-1/2}$. Moreover, using the model equation

$$\begin{aligned} & \left(\alpha_\perp' \Delta X_t \middle| M' E_t, C\varepsilon_{t-i} + \Delta Y_{t-i} E_{t-i}, D_{j,t-i} \right) \\ = \ & \left(\alpha_\perp' \varepsilon_t \middle| M' E_t, C\varepsilon_{t-i} + \Delta Y_{t-i} E_{t-i}, D_{j,t-i} \right). \end{aligned}$$

Now, for a stationary process Z_t adapted to the natural filtration for ε_s it follows that $\sum_{t=1}^{T} Z_{t-1} \varepsilon_t = \mathrm{O}_\mathcal{P}(\sqrt{T})$. Together with (A.17) this implies

$$T^{1/2} s_{10} \alpha_\perp = T^{-1/2} \sum_{t=1}^{T} \left(M_\perp' E_t \right) \left(\alpha_\perp' \varepsilon_t \middle| M' E_t, D_{j,t-i} \right)' + \mathrm{o}_\mathcal{P}(1), \qquad \text{(A.21)}$$

and this leads to (A.15). For the convergence (A.16) note that when s_{10} is of order $T^{-1/2}$ while S_{00}, s_{11} are of order one then $S_{00}^\mu = S_{00} + \mathrm{o}_\mathcal{P}(1)$. Using that $\beta^{\mu\prime} S_{10}^\mu = \beta^{0\prime} S_{10}$ it then follows that

$$\begin{aligned} & \left(S_{00}^\mu \right)^{-1} - \left(S_{00}^\mu \right)^{-1} S_{01}^\mu \beta^\mu \left\{ \beta^{\mu\prime} S_{10}^\mu \left(S_{00}^\mu \right)^{-1} S_{01}^\mu \beta^\mu \right\}^{-1} \beta^{\mu\prime} S_{10}^\mu \left(S_{00}^\mu \right)^{-1} \\ = \ & S_{00}^{-1} - S_{00}^{-1} S_{01} \beta^0 \left\{ \beta^{0\prime} S_{10} S_{00}^{-1} S_{01} \beta^0 \right\}^{-1} \beta^{0\prime} S_{10} S_{00}^{-1} + \mathrm{o}_\mathcal{P}(1) \end{aligned}$$

which by (A.6)-(A.7) is equivalent to $\alpha_\perp \left(\alpha_\perp' \Omega \alpha_\perp \right)^{-1} \alpha_\perp'$.

In the general case, when $\tau_l M_\perp' \neq 0$ and hence $G \notin \mathrm{span}\, M$ and $\beta' \tau_l M_\perp' \neq 0$, then (A.18) and (A.20) fail and s_{10} is of order one. In particular $\beta' s_{01} = \beta' \tau_l s_{11} + \mathrm{o}_\mathcal{P}(1)$, which is of order one. □

A.8. *Proof of Theorem 4.2*

The case $G \in \mathrm{span}\, M$. The proof follows in two steps. First it is established that the likelihood ratio test statistic for $H_l^\mu(r)$ against $H_l^\mu(p) = H_l(p)$ converges in distribution to

$$\mathrm{tr} \left[\int_0^1 dW \begin{pmatrix} NF \\ f \end{pmatrix}' \left\{ \int_0^1 \begin{pmatrix} NF \\ f \end{pmatrix} \begin{pmatrix} NF \\ f \end{pmatrix}' du \right\} \int_0^1 \begin{pmatrix} NF \\ f \end{pmatrix} dW' \right].$$

The proof is the same as that of Theorem 3.1 with B_T replaced by

$$B_T^\mu = \begin{pmatrix} B_T & 0 \\ 0 & T^{1/2}I_{q-m} \end{pmatrix}.$$

Next, mimicking the proof Corollary 11.2 (J) it is argued that the likelihood ratio test statistic for $H_l^\mu(r)$ against $H_l^\gamma(r)$ converges in distribution to

$$\text{tr}\left\{ \int_0^1 dW f' \left(\int_0^1 f f' du \right)^{-1} \int_0^1 f dW' \right\}.$$

It then follows that the test is asymptotically χ^2 with $\dim W = (p-r)$ times $\dim f = (q-m)$ degrees of freedom.

The case $G \notin \text{span } M$. In general, the result (A.16) fails and the above argument does not hold. It follows that nuisance parameters are involved in the asymptotic distribution. $\square$

B. References

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

Bartlett, M. S., 1938. Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society* 34, 33-40.

Clements, M.P., and Hendry, D.F., 1999. *Forecasting non-stationary economic time series.* MIT press, Cambridge MA.

Doornik, J.A, 1998. Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12, 573-593.

Doornik, J.A., and Hendry, D.F., 1994. Modelling linear dynamic econometric systems. *Scottish Journal of Political Economy* 41, 1-33.

Doornik, J.A., Hendry, D.F. and Nielsen, B., 1998. Inference in cointegrating models: UK M1 revisited. *Journal of Economic Surveys* 12, 533-572.

Hansen, H. and Johansen, S., 1999. Some tests for parameter constancy in cointegrated VAR models. To appear in *Econometrics Journal.*

Hendry, D.F., 1997. The econometrics of macroeconomic forecasting. *Economic Journal* 107, 1330-1357.

Hotelling, H., 1936. Relations between two sets of variates. *Biometrika* 28, 321-77.

Inoue, A., 1999. Tests of cointegrating rank with a trend-break. *Journal of Econometrics* 90, 215-237.

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

Johansen, S., 1996. *Likelihood-based inference in cointegrated vector autoregressive models.* 2nd printing. Oxford University Press.

Kuo, B., 1998. Test for partial parameters stability in regressions with I(1) processes. *Journal of Econometrics* 86, 337-368.

Mosconi, R., 1998. *MALCOLM: The theory of practice of cointegration analysis in RATS.* Venice: Ca' Foscarina.

Nielsen, B., 1997. Bartlett correction of the unit root test in autoregressive models. *Biometrika* 84, 500-504.

Nielsen, B., and Rahbek, A., 2000. Similarity issues in cointegration models. To appear in *Oxford Bulletin of Economics and Statistics* 62.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361-1401. Erratum, 1993, *Econometrica* 61, 248-249.

Perron, P., 1990. Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics* 8, 153-162. Corrections and Extensions by Perron, P., and Vogelsang, T., 1992, *Journal of Business & Economic Statistics* 10, 467-470.

Rappoport, P., and Reichlin, L., 1989. Segmented trends and non-stationary time series. *Economic Journal* 99 supplement, 168-177.

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