Likelihood-Based Inference in Multivariate Panel Cointegration Models

Rolf Larsson*
Department of Statistics
Stockholm University

Johan Lyhagen†
Department of Economic Statistics
Stockholm School of Economics

Stockholm School of Economics
Working Paper Series in Economics and Finance, No. 331

September 21, 1999

*E-mail: rolf.larsson@stat.su.se
†Corresponding author, e-mail: stjl@hhs.se.

Abstract: This paper presents a general likelihood-based framework for inference in panel-VAR models with cointegrating restrictions. The cointegrating relations are restricted to each cross-section while the rest of the model is unrestricted. The homogenous restriction of common cointegrating space is also considered. Asymptotic distributions of parameter estimates and the test statistics for the cointegrating rank and the homogenous restriction are derived. The distribution for the cointegrating rank is shown to be the convolution of the standard distribution of the trace statistic and the χ^2 distribution. The homogenous restriction test statistic is χ^2. A Monte Carlo simulation investigates the small sample properties of the two tests. The empirical size of the test for the cointegrating rank is well above the nominal. A Bartlett corrected test statistic is shown to have size very close to the nominal. We give an empirical example for a consumption model including consumption, income and inflation.
Key Words: Cointegration; Consumption; Panel data; Rank test.
JEL-Classification: C12; C13; C15; C22; C23; D12.

1. Introduction

Compared to most previous work on panels and unit roots/cointegration (see e.g. Levin and Lin, 1992, 1993, and Im, Pesaran and Shin, 1997), this paper focuses on multivariate cointegration and extends the previous work by Larsson, Lyhagen and Löthgren (1998) and Groen and Kleibergen (1999). Consider a panel data set that consists of a sample of cross-sections where the cross-sections are e.g. industries, regions or countries. Economic theory may postulate that long run equilibriums should hold for each cross-section, but it is feasible to allow the cross-sections to depend on all the equilibriums of the cross-sections. A panel model with such cointegrating restrictions are proposed. The asymptotic distribution of the likelihood ratio test for the number of cointegrating relations is derived. This distribution may be described as the convolution of two independent variates: the first one following the well-known asymptotic distribution of the trace test (a Dickey-Fuller type distribution) and the second one being χ^2. Further, a likelihood ratio test of common cointegrating space is proposed and it is shown that the asymptotic distribution is χ^2.

A Monte Carlo simulation is performed to analyze the small sample properties of the two tests. The test for common cointegrating space has sufficiently good size and power properties while the test for cointegrating rank does not. This result makes us propose the use of a Bartlett corrected test statistic which is found to have desired properties, i.e. a size very close to the nominal one.

An empirical example concerning two groups is carried out. The groups consists of countries that are, in some sense, similar. The first consists of some larger economies (Japan, UK and US) and the second of the major Nordic countries (Denmark, Finland, Norway and Sweden). The variables are income, consumption and inflation. The result is that the first group has two cointegrating vectors while the second has only one. The test of common cointegrating space is rejected for both groups.

The paper is as follows. In the next section, the general model and the two special cases are presented while estimation is discussed in Section 3. Section 4 considers asymptotic results for the distribution of parameters and the likelihood ratio tests. To evaluate the small sample properties a Monte Carlo simulation is carried out in Section 5, and the empirical example is presented in Section 6. A conclusion ends the paper.

2. The General Model

Consider a panel data set that consists of a sample of N cross-sections (e.g. industries, regions or countries) observed over T time periods. To be able to efficiently discuss multivariate panel cointegration we need to define some notation. Let $i = 1, ..., N$ index the groups, $t = 1, ..., T$ the sample time period and $j = 1, ..., p$ the variables in each group. Then y_{ijt} denotes the ith group, the jth variable at time t. The observed p-vector for group i at time period t is given by $\mathbf{y}'_{it} = (y_{i1t}, y_{i2t}, ..., y_{ipt})'$. Define $Y_t = [\mathbf{y}'_{1t}, \mathbf{y}'_{2t}, ..., \mathbf{y}'_{Nt}]'$ as the Np-vector of the panel of observations available at time t on the p variables for the N groups.

The regression that is the basis for our work is

$$\begin{bmatrix} \Delta\mathbf{y}_{1t} \\ \Delta\mathbf{y}_{2t} \\ \vdots \\ \Delta\mathbf{y}_{Nt} \end{bmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \cdots & \Pi_{1N} \\ \Pi_{21} & \Pi_{22} & & \\ \vdots & & \ddots & \vdots \\ \Pi_{N1} & \Pi_{22} & \cdots & \Pi_{NN} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1t-1} \\ \mathbf{y}_{2t-1} \\ \vdots \\ \mathbf{y}_{Nt-1} \end{bmatrix} +$$

$$\sum_{k=1}^{m-1} \begin{bmatrix} \Gamma_{11,k} & \Gamma_{12,k} & \cdots & \Gamma_{1N,k} \\ \Gamma_{21,k} & \Gamma_{22,k} & & \\ \vdots & & \ddots & \vdots \\ \Gamma_{N1,k} & \Gamma_{22,k} & \cdots & \Gamma_{NN,k} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{y}_{1t-k} \\ \Delta\mathbf{y}_{2t-k} \\ \vdots \\ \Delta\mathbf{y}_{Nt-k} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1t} \\ \boldsymbol{\varepsilon}_{2t} \\ \vdots \\ \boldsymbol{\varepsilon}_{Nt} \end{bmatrix}, \; t = 1, ..., T, \qquad (2.1)$$

or more compactly written as

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{k=1}^{m-1} \Gamma_k \Delta Y_{t-k} + \varepsilon_t \qquad (2.2)$$

where Δ is the first difference filter $(1 - L)$, $Y_t = (\mathbf{y}'_{1t}, \mathbf{y}'_{2t}, ..., \mathbf{y}'_{Nt})'$ and $\varepsilon_t = (\boldsymbol{\varepsilon}'_{1t}, \boldsymbol{\varepsilon}'_{2t}, ..., \boldsymbol{\varepsilon}'_{Nt})'$ is of order $Np \times 1$, where ε_t is assumed multivariate normally distributed as $\varepsilon_t \sim N_{Np}(0, \Omega)$, with covariance matrix

$$\Omega = \{\Omega_{ij}\} = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1N} \\ \Omega_{21} & \Omega_{22} & & \\ \vdots & & \ddots & \vdots \\ \Omega_{N1} & \Omega_{22} & \cdots & \Omega_{NN} \end{bmatrix}, \qquad (2.3)$$

and Π and Y_{t-1} are of order $Np \times Np$ and $Np \times 1$, respectively.

As seen above, Π and Γ_k, $k = 1, ..., m-1$, can be partitioned into submatrices, Π_{ij} and Γ_{ij}, $i, j = 1, ..., N$, respectively, each of dimension $p \times p$.

To continue, we impose some structure on this model. First, we consider a reduced rank specification of the panel model where the matrix Π is of rank $\sum r_i$, $0 \leq r_i \leq p$, specified as $\Pi = AB'$, where the matrices A and B are both of order $Np \times \sum r_i$ given by

$$A = \{\alpha_{ij}\} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & & \\ \vdots & & \ddots & \vdots \\ \alpha_{N1} & \alpha_{22} & \cdots & \alpha_{NN} \end{bmatrix}, \tag{2.4}$$

and

$$B = \left\{\beta_{ij}\right\} = \begin{bmatrix} \beta_{11} & 0 & \cdots & 0 \\ 0 & \beta_{22} & & \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{NN} \end{bmatrix}. \tag{2.5}$$

A contains the short-run coefficients and B the long-run coefficients β_i each of rank r_i.

The block matrix elements of Π are given by $\Pi_{ij} = \sum_{k=1}^{N} \alpha_{ik}\beta'_{kj} = \alpha_{ij}\beta'_{jj}$ due to the restriction that $\beta_{ij} = 0 \ \forall i \neq j$. In a more compact notation, the model is written as:

$$\Delta Y_t = AB'Y_{t-1} + \sum_{k=1}^{m-1} \Gamma_k \Delta Y_{t-k} + \varepsilon_t. \tag{2.6}$$

This general model allows a simultaneous modelling of the long-run relations between several variables for a panel of groups allowing for heterogeneous long-run cointegration relations within each group. Due to the restriction $\beta_{ij} = 0 \ \forall i \neq j$, cointegrating relationships are only allowed for within each of the N groups in the panel. These cointegrating relationships are contained in the matrix $B'Y_{t-1}$ which consists of the r_i cointegrating relations for each individual, $\beta'_i \mathbf{y}_{it-1}$, $i = 1, ..., N$. However, the model allows for an important short-run dependence between the panel groups, since α_{ij} is not restricted to zero for $i \neq j$. More specifically, the off diagonal elements in $\Pi = AB'$ which are given by $\Pi_{ij} = \alpha_{ij}\beta'_j$ for $i \neq j$, represent the short-run dependencies of the changes in the series for group i that are due to long-run equilibrium deviations in group j. As in the standard single-group model the diagonal element of Π, $\Pi_{ii} = \alpha_{ii}\beta'_i$, represents the short-run adjustments in group i resulting from a deviation from long-run equilibrium in group i.

Larsson, Lyhagen and Löthgren (1998) consider a similar heterogeneous panel data model under cointegrating restrictions, with the added restriction that no dependencies are allowed between the panel groups. I.e., the off-diagonal block elements of the matrices A, Γ and Ω are zero. With this additional restriction the model is completely heterogeneous and the panel groups are modelled individually as

$$\Delta \mathbf{y}_{it} = \alpha_i \beta_i' \mathbf{y}_{it-1} + \sum_{k=1}^{m-1} \Gamma_{ii,k} \Delta \mathbf{y}_{it-k} + \boldsymbol{\varepsilon}_{it}, i = 1, ..., N, t = 1, ..., T. \tag{2.7}$$

Groen and Kleibergen (1999) relax the assumption of block diagonality of Ω in this model.

2.1. Homogeneity restrictions/tests

Based on the general panel model we are interested in tests of homogeneity restrictions on the model. The first basic hypothesis we consider states that all of the panel group-specific matrices Π_i $i = 1, ..., N$, have a maximum rank r:

$$H_0 : rank\left(\Pi_i\right) = r_i \leq r \text{ for all } i = 1, ..., N, \tag{2.8}$$

is tested against the alternative

$$H_1 : rank\left(\Pi_i\right) = p \text{ for all } i = 1, ..., N. \tag{2.9}$$

This null hypothesis states that the maximum cointegrating rank in the panel is given by r. Larsson, Lyhagen and Löthgren (1998) develop a likelihood-based tests for this hypothesis based on the completely heterogeneous model. In this paper we consider an extension of this test statistic to the more general model considered here.

Given the assumption of equal rank, the homogeneity hypothesis that the cointegrating vectors in the panel span the same space for each of the individual groups in the panel is natural. That is, the second homogeneity hypothesis we consider is given by:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_N = \beta, \tag{2.10}$$

against the alternative

$$H_1 : \beta_i \neq \beta_j \text{ for some } i, j. \tag{2.11}$$

Note that the homogeneous long-run coefficient β is not uniquely determined. Instead, the homogeneity hypothesis is the hypothesis that the long-run coefficients

β_i span the same space. This is seen because if $B_1 = B_2 R$ for some $Nr \times Nr$ matrix R of full rank, then we may write $AB_1' = A^* B_2'$ with $A^* = AR'$.

Under the null hypothesis of homogeneity, the matrix of long-run coefficients B can be written as $B = I_N \otimes \beta$ and the general model is given by

$$\Delta Y_t = A\left(I_N \otimes \beta'\right) Y_{t-1} + \sum_{k=1}^{m-1} \Gamma_k \Delta Y_{t-k} + \varepsilon_t. \tag{2.12}$$

3. Statistical Analysis of the Models

In this section we discuss the estimation of the model

$$\Delta Y_t = AB' Y_{t-1} + \sum_{k=1}^{m-1} \Gamma_k \Delta Y_{t-k} + \varepsilon_t, \tag{3.1}$$

with the two sets of restrictions $B = Diag\left(\beta_{ii}\right)$ and $B = \left(I_N \otimes \beta\right)$.

Observe that, for small enough T, it may not be possible to estimate the parameters of the model. For example, if the lag length m is one, the number of parameters is at most N^2p^2. As we have Np equations each equation must have Np observations to give an exactly identified system. Due to that the right hand side consists of lagged left side variables one observation is lost, hence the number of time units used must be at least $T = Np + 2$.

3.1. Individual cointegrating relations

The restriction $B = Diag\left(\beta_{ii}\right)$ may be written $B = \left(H_1^{(p)}\beta_{11}, \ldots, H_N^{(p)}\beta_{NN}\right)$ where $H_i^{(p)}$ is a $Np \times p$ matrix of zeros except in the i:th block where it is a unit matrix, i.e.

$$H_i^{(p)} = \begin{bmatrix} 0 & \ldots & 0 & I_p & 0 & \ldots & 0 \end{bmatrix}'.$$

(In the rest of this section, we will drop the superindex (p).)

Estimation of such kind of restrictions is discussed in e.g. Johansen (1995a) and Johansen (1995b). The estimation procedure is to estimate $H_1\beta_{11}$ in a reduced rank regression where $H_2\beta_{22}, \ldots, H_N\beta_{NN}$ have been concentrated out. Continue by estimating $H_2\beta_{22}$ given $H_1\beta_{11}, H_3\beta_{33}, \ldots, H_N\beta_{NN}$. When $H_N\beta_{NN}$ has been estimated, restart the estimation with the new values of $H_1\beta_{11}, \ldots, H_N\beta_{NN}$. Repeat until convergence. For starting values we propose to use the β_{ii} found when doing a standard cointegrating analysis for i:th cross-section.

3.2. Homogenous cointegrating relations

Estimating

$$\Delta Y_t = AB'Y_{t-1} + \sum_{k=1}^{m-1} \Gamma_k \Delta Y_{t-k} + \varepsilon_t. \tag{3.2}$$

using the method proposed by Johansen (1988) we get the unrestricted estimator of B which, with probability 1, does not satisfy the restriction $B = I_N \otimes \beta$, hence, it can not be used to estimate the model we are interested in. Instead we propose to use the switching method of Boswijk (1995). It is possible to numerically maximize the likelihood, but this is probably more time consuming than the switching method when large N and p are considered, although Boswijk (1995) discusses an example when Newton-Raphson will reach optimum in one step and the switching converges slowly to optimum. See also Johansen (1995a) for a discussion on optimization versus switching methods.

For ease of exposition, we consider the model in (2.12) without any short-run dynamics, which is the same as assuming that these terms have been concentrated out. Premultiply with the inverse of the square root of the covariance matrix of ε_t, i.e. with $\Omega^{-1/2}$, to get

$$\Omega^{-1/2}\Delta Y_t = \Omega^{-1/2} A \left(I_N \otimes \beta'\right) Y_{t-1} + \Omega^{-1/2}\varepsilon_t. \tag{3.3}$$

or equivalently

$$\Delta\tilde{Y}_t = \tilde{A}\left(I_N \otimes \beta'\right) Y_{t-1} + e_t. \tag{3.4}$$

where we used the notation $\tilde{Y}_t$ and $\tilde{A}$ for $\Omega^{-1/2}\Delta Y_t$ and $\Omega^{-1/2}A$ respectively, and where $E\left(e_t e_t'\right) = I_p$. Using the relation $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ we have

$$\Delta\tilde{Y}_t = \left(Y_{t-1}' \otimes \tilde{A}\right) \text{vec}\left(I_N \otimes \beta'\right) + e_t. \tag{3.5}$$

Define the matrix H of size $N^2rp \times rp$ as

$$H = \begin{bmatrix} I_p \otimes \delta_1^N \otimes I_r \\ I_p \otimes \delta_2^N \otimes I_r \\ \vdots \\ I_p \otimes \delta_N^N \otimes I_r \end{bmatrix} \tag{3.6}$$

where δ_i^N is a $N \times 1$ vector of zeros except with one in the ith position. Then $\text{vec}(I_N \otimes \beta') = H\text{vec}(\beta')$. Note that the inverse of $H'H$ exists, i.e. it is $N^{-1}I_{rp}$. Substitute this into (3.5),

$$\Delta\tilde{Y}_t = \left(Y_{t-1}' \otimes \tilde{A}\right) H\text{vec}\left(\beta'\right) + e_t. \tag{3.7}$$

The OLS estimator is then

$$\operatorname{vec}(\beta') = \left[\left(\left(Y'_{t-1} \otimes \tilde{A}\right) H\right)' \left(\left(Y'_{t-1} \otimes \tilde{A}\right) H\right)\right]^{-1} \left(\left(Y'_{t-1} \otimes \tilde{A}\right) H\right)' \Delta\tilde{Y}_t \qquad (3.8)$$

This shows that for a given value of A and Ω we may estimate β. The problem of estimating A and Ω for given values of β is much simpler, estimate A in (3.2) by regression of ΔY_t on $B'Y_{t-1}$, corrected for $(\Delta Y_{t-1}, ..., \Delta Y_{t-m+1})$. This regression also gives an estimate of Ω. The switching algorithm is that for given initial values of β estimate A and Ω, then for these estimated values estimate β. Repeat until the increase of the likelihood is sufficiently small. The mean of the β_{ii} found when doing a standard cointegrating analysis for i:th cross-section are used as starting values.

4. Distribution of parameters and tests

In this section we derive the distribution of the estimated parameters and the distribution of the likelihood ratio test for the cointegrating rank under the models where we have the restrictions $B = Diag\,(\beta_{ii})$ and $B = (I_N \otimes \beta)$. The rank is tested by a likelihood ratio test when the estimated model has the restriction $B = Diag\,(\beta_{ii})$. Then, given the rank, model $B = (I_N \otimes \beta)$ is tested against $B = Diag\,(\beta_{ii})$. The proofs are in the appendix.

Consider the model given in (2.2). Having observations up to time T, our object is to test

$$H_0 : rank\,(\Pi_i) = r_i \leq r \text{ for all } i = 1, ..., N, \qquad (4.1)$$

against the alternative

$$H_1 : rank\,(\Pi_i) = p \text{ for all } i = 1, ..., N, \qquad (4.2)$$

using the likelihood ratio test, Q_T. Further, define $A_\perp$as a $Np \times N\,(p-r)$ matrix (the choice of it is not unique) that fulfills the requirements $A'_\perp A = 0$, $A'A_\perp = 0$ and $(A, A_\perp)$ has full rank (Np), and similarly for $B_\perp$. Consequently, we may choose $B_\perp \equiv$diag$(\beta_{1\perp}, ..., \beta_{N\perp})$. Furthermore, letting $\Gamma \equiv I_{Np} - \sum_{k=1}^{m-1} \Gamma_k$, we need the assumption, for ruling out processes integrated of order higher than one,

Assumption A The matrix $A'_\perp \Gamma B_\perp$ has full rank.

4.1. The distribution of the parameter estimates

The asymptotics of $\widehat{B} - B$ is described in the following theorem. Following Johansen (1995b), we let $\widehat{B} - B = \overline{B}_\perp X_T$, where X_T is $N(p-r) \times Nr$, and where $\overline{B}_\perp \equiv B_\perp (B'_\perp B_\perp)^{-1}$. Furthermore, we define the $N^2 (p-r) r \times N(p-r) r$ matrix K through

$$K \equiv \left(H_1^{(r)} \otimes H_1^{(p-r)}, ..., H_N^{(r)} \otimes H_N^{(p-r)} \right), \tag{4.3}$$

where the $H_i^{(n)}$ are as defined in section 3.1. Moreover, G and W are shorthand for the processes G_t and W_t where W_t is an Np-dimensional Wiener process with covariance matrix Ω and $G_t \equiv \overline{B}'_\perp C W_t$ with $C \equiv B_\perp (A'_\perp \Gamma B_\perp)^{-1} A'_\perp$, $\Gamma \equiv I_{Np} - \sum_{i=1}^{m-1} \Gamma_i$. Note that if $r = 0$, the block diagonal structure has no meaning. Hence, in the following we will assume that $r > 0$.

Theorem 4.1. *Under assumption A and if $r > 0$, we have that as $T \to \infty$,*

$$T \operatorname{vec} X_T \xrightarrow{w} K F_1^{-1} K' \operatorname{vec} \left(\int G dW' \Omega^{-1} A \right),$$

where

$$F_1 \equiv K' \left(A' \Omega^{-1} A \otimes \int G G' \right) K.$$

4.2. Likelihood ratio test statistics

We are now ready for our first main result.

Theorem 4.2. *Under assumption A and if $r > 0$, we have that as $T \to \infty$,*

$$-2 \log Q_T \xrightarrow{w} U + V,$$

where, defining $\widetilde{W}_t$ to be an $N(p-r)$-dimensional standard Wiener process (with mean zero and unity covariance matrix),

$$U = tr \left\{ \int d\widetilde{W} \widetilde{W}' \left(\int \widetilde{W} \widetilde{W}' \right)^{-1} \int \widetilde{W} d\widetilde{W}' \right\},$$

and where V is χ^2 with $N(N-1)(p-r)r$ degrees of freedom, independent of U.

In other words, the limit distribution of our test for cointegrating rank equals the convolution of a well-known Dickey-Fuller type distribution (which arises as the asymptotic distribution for the corresponding rank test in a model without any restrictions on B, cf Johansen, 1995) and an independent χ^2 variate. It is fairly easy to simulate this distribution in the usual fashion, approximating the Wiener process with a random walk. Moreover, considering the moments of U as known (see e.g. the simulation results of Doornik, 1998), our representation provides us with a simple way of calculating the asymptotic moments of our test statistic.

4.3. Testing homogenous cointegrating relations

Our next step is to find the asymptotic distribution of the log likelihood ratio test, given the rank, of the homogeneity hypothesis

$$H_0 : \beta_1 = \beta_2 = ... = \beta_N = \beta,$$

against the alternative

$$H_1 : \beta_i \neq \beta_j \text{ for some } i, j.$$

In view of earlier literature on similar restriction tests (see e.g. Johansen, 1995b), the result that this distribution is χ^2, given in the theorem below, should not come as a surprise to the reader. The number of degrees of freedom, $(N-1)\,r\,(p-r)$, is natural because as is easily seen, this is the difference of the numbers of free parameters under the different hypotheses.

Theorem 4.3. *Under assumption A and given the rank r, the log likelihood ratio test statistic for test of $H_0 : \beta_1 = ... = \beta_N = \beta$ against $H_1 : \beta_i \neq \beta_j$ for some i, j is, under H_0 and as $T \to \infty$, asymptotically χ^2 distributed with $(N-1)\,r\,(p-r)$ degrees of freedom.*

5. A Simulation Study

It is of practical interest to evaluate how well the asymptotic distribution of the likelihood ratio test for the cointegrating rank mimics the small sample distribution. To this end, a Monte Carlo simulation is a suitable tool to use. The length of the random walk approximating the Brownian motion is 800 and the number of replicates is 100000. For the analysis of small sample properties, the sample sizes we consider are $T = 100, 200, 500$ and 1000. Due to time considerations, the

Model	Case 1		Case 3		Case 5	
r	1	2	1	2	1	2
	0.406	0.040	0.306	0.130	0.668	0.730
	0.406	0.562	0.323	0.414	0.839	0.746
	0.840	0.663	0.900	0.414	0.897	0.746
	1	0.663	1	0.601	1	0.849
	1	0.797	1	0.759	1	0.867
	1	0.797	1	0.871	1	0.896
	1	1	1	1	1	1
	1	1	1	1	1	1
	1	1	1	1	1	1

Table 5.1: Absolute value of the eigenvalues of some data generating pocesses used in the simulation study. In Case 5 only the largest one is in the table.

number of replicates is limited to 10000. The data generating process is gained by estimating the models of interest on data. The variables used are (log of) consumption, income and inflation for Japan, UK and US, i.e. $n = p = 3$, see the next section. The largest (absolute value of the) eigenvalues of the data generating processes named Case 1, Case 3 and Case 5 are shown in Table (5.1). All of them are relatively far from one, hence, the process for rank one and rank two are sufficiently separated.

The simulations are carried out in Gauss 3.2. Below are the different cases simulated for ranks one and two:

1. $B = (I_N \otimes \beta)$ with A and Ω unrestricted.
2. As in 1 but with block diagonal Ω.
3. $B = Diag\,(\beta_{ii})$ with A and Ω unrestricted.
4. As 3 but with both A and Ω block diagonal.
5. As 4 but with $m = 2$.

Cases 1, 3 and 5 are estimated from data. Case 2 is gained from restricting case 1 and case 4 is obtained from restricting case 5. Unfortunately, and contrary to the ordinary case, the convergence to the asymptotic distribution is slow. This is especially valid for Case 5 where $m = 2$. This makes it plausible to use some

Case	T=100	T=200	T=500	T=1000
1	0.293	0.149	0.089	0.072
2	0.295	0.152	0.089	0.069
3	0.314	0.173	0.091	0.076
4	0.251	0.124	0.077	0.067
5	0.846	0.408	0.156	0.097

Table 5.2: Size for small samples, 5% test and rank=1. The critical value is 97.20.

Case	T=100	T=200	T=500	T=1000
1	0.253	0.135	0.082	0.063
2	0.238	0.126	0.073	0.063
3	0.226	0.124	0.081	0.068
4	0.316	0.151	0.086	0.070
5	0.734	0.343	0.133	0.086

Table 5.3: Size for small samples, 5% test and rank=2. The critical value is 39.43.

kind of small sample asymptotics such as the Bartlett correction. Moreover, it seems that the size properties for the two different ranks considered are quite similar.

For the test of $B = (I_N \otimes \beta)$ versus $B = Diag\,(\beta_{ii})$, the size of the test is much better, see Table (5.4) and Table (5.5). Further the power is extremely good, the power is 1 even for the smallest sample size ($T = 100$).

Case	T=100	T=200	T=500	T=1000
1	0.110	0.049	0.055	0.051
2	0.094	0.073	0.059	0.054
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1

Table 5.4: Size and power for small samples of test for common cointegrating space, four degrees of freedom, 5% test and rank=1.

Case	T=100	T=200	T=500	T=1000
1	0.120	0.074	0.055	0.055
2	0.104	0.074	0.059	0.052
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1

Table 5.5: Size and power for small samples of test for common cointegrating space, 5% test and rank=2.

5.1. Bartlett correction

The Bartlett correction was introduced by Bartlett (1937), see Cribaro-Neto and Cordeiro (1996) for a nice treatment of the subject. In cointegration, it has been used by e.g. Jacobson and Larsson (1999) with only a small improvement of the asymptotic distribution. This is probably due to the good performance of the asymptotic distribution. In our case, where the size of the test is far away from the nominal for sample sizes up to $T = 500$, the use of a Bartlett corrected statistic may be useful. Consider the statistic C_T for sample size T and let C_∞ denote the asymptotic one and E the expectation operator. Then the Bartlett corrected statistic is

$$C_T^* = EC_T \frac{C_\infty}{EC_\infty} \tag{5.1}$$

and has been found useful in practise (given that a good estimator of EC_T could be found). Jacobson and Larsson (1999) have demonstrated the difficulties to achieve a closed form expression for the likelihood ratio test for even such a simple system as one with only two variables and one cointegrating vector. In our simulations the mean of the small sample statistics are used, an approach that could be used in practise if conditioning on the estimated model. The result is that the Bartlett corrected statistic works extremely well for all sample sizes and cases considered, the size is very close to the nominal 5%, see Table (5.6) for rank one. The result for rank two is very similar, hence, not reported.

6. An Empirical Example: The Consumption Function

In this section, we estimate a standard consumption function of the type considered by Davidson et al. (1978) for two homogenous groups of OECD countries

Case	T=100	T=200	T=500	T=1000
1	0.045	0.049	0.047	0.050
2	0.045	0.048	0.047	0.048
3	0.045	0.048	0.046	0.049
4	0.049	0.050	0.047	0.051
5	0.038	0.047	0.048	0.049

Table 5.6: Size for small samples, 5% Bartlett corrected test and rank=1.

over the 35 year period $1960-1994$. The two groups are 1) Japan, UK and US, 2) Denmark, Finland, Norway and Sweden[1]. We consider the heterogeneous panel error correction model with variable vector for each country given by

$$Y_{it} = \left(c_{it},\ y_{it}^{d}, \Delta p_{it}\right)',$$

where c_{it} is the logarithm of real consumption per capita, y_{it}^{d} is the logarithm of real disposable income per capita and Δp_{it} is the rate of inflation. We follow Pesaran, Shin and Smith (1999) in the definition of the variables: Consumption is measured by the logarithm of total private consumption per-capita, inflation by the change in the logarithm of the consumption deflator and national disposable income deflated by the consumption deflator is used as measure of income. Further, the variables are demeaned and only $m = 1$ is considered. The results of the likelihood ratio tests are in Table (6.1). The Bartlett corrected critical values are gained by using the estimated model as data generating process when calculating the sample mean. A bootstrap approach like the one proposed by Gredenhoff and Jacobsson (1998) could be used but with the good size properties of the Bartlett critical values we do not think that a bootstrap is necessary. For the groups that consists of Japan, UK and US the number of cointegrating vectors is 2 when using the Bartlett corrected critical values while for the group that consists of Denmark, Finland, Norway and Sweden the number is 1. Note that if the asymptotic critical values would be used the estimated rank would be 3 for both groups. The Bartlett corrected critical value for the Denmark group and rank 2 could not be calculated due to the fact that the estimated model have roots larger than one, hence, numerical (and theoretical) problems erased. The tests of common cointegrating space gives test statistics of 40.72 and 35.19 respectively and should be

[1] The data are obtained from the OECD CD-ROM Statistical Compendium, edition 02#1997.

	Japan...			Denmark...		
H_0	As. crit.	B. crit.	$-2\log Q_T$	As. crit.	B. crit.	$-2\log Q_T$
$r = 0$	177.37	218.61	295.06	306.54	425.65	468.78
$r \leq 1$	97.12	142.24	155.74	168.91	292.32	259.91
$r \leq 2$	39.59	68.43	49.84	68.85	-	96.12

Table 6.1: Test for cointegraing rank using asymptotic and Bartrlett corrected corrected critical values for the group Japan, UK and US and the group Denmark, Finland, Norway and Sweden.

compared to $\chi^2_{0.95,df=4} = 9.49$ and $\chi^2_{0.95,df=6} = 12.59$. Hence, both groups reject the null of common cointegrating space.

7. Summary and Concluding Remarks

In this paper we have proposed a panel-VAR with cointegrating restrictions where the cointegrating relations matrix is block diagonal, each block corresponds to a cross-section, while the rest of the model is unrestricted. This model is a generalization of the models proposed by Larsson, Lyhagen and Löthgren (1998) and Groen and Kleibergen (1999). The asymptotic distribution of the estimated parameters and the two test statistics considered are derived. The first test statistic tests for the cointegrating rank while the second test the homogeneity restrictions of common cointegrating space. A Monte Carlo simulation is carried out with the purpose of analyzing the small sample properties of the two test statistics. The homogeneity test has satisfying size properties while the test for cointegrating rank has not. However, when Bartlett correcting the rank test, a size very close to the nominal is gained. An empirical example using income, consumption and inflation and two groups of countries shows that Japan, UK and US have two cointegrating relations while Denmark, Norway, Finland and Sweden have only one. It should be noted that if the asymptotic critical values instead of the Bartlett corrected ones would be used, a cointegrating rank of three would have emerged for both groups, showing that using a test with correct size is crucial for empirical work.

The present work may be extended in many interesting directions. For example, dummy variables could be included in the model. This would probably give the same type of asymptotic results. Another important issue for applications would be to investigate asymptotics as the number of individuals, or in our case countries, tends to infinity. Under suitable assumptions, we should in this case get asymptotic normality as in Larsson, Lyhagen and Löthgren (1998).

Acknowledgments

We would like to thank Mickael Löthgren and Paolo Paruolo for valuable comments. Financial support for Lyhagen from Tore Browaldhs Foundation for Scientific Research and Education and The Swedish Foundation for International Cooperation in Research and Higher Education (STINT) is gratefully acknowledged.

References

Bartlett, M.S. (1937) Properties of suffiency and statistical tests, Proceedings of the Royal Statistical Society of London, Series A, Vol 160, 268-282.

Boswijk, H.P. (1995) Identifiability of Cointegrated systems. Tinbergen Institute Discussion Paper TI 7-95-078.

Cribaro-Neto, F. and G. Cordeiro (1996) On Bartlett and Bartlett-type corrections, Econometric Reviews, 15, 339-367.

Doornik, J.A. (1998) A convenient approximation to the asymptotic distribution of cointegration tests, Nuffield College, Oxford, Mimeo.

Gredenhoff, M. and T. Jacobsson (1998) Bootstrap testing and approximate finite sample distributions for test of linear restrictions on cointegrating vectors, Working Paper No. 67, Sveriges Riksbank

Groen, J.J.J and F.R. Kleibergen (1999) Likelihood-based cointegration analysis in panels of vector error correction models, Tinbergen Institute Discussion Paper TI 99-055/4.

Im, K.S., M.H. Pesaran and Y. Shin (1997) Testing for unit roots in heterogenous panels, Trinity College, Cambridge, Mimeo.

Jacobson, T. and R. Larsson (1999) Bartlett Corrections in Cointegration Testing, Computational Statistics & Data Analysis, 31, 203-225.

Johansen S. (1988) Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control, 12, 359-386.

Johansen, S. (1995a) Identifying restrictions of linear equations - With applications to simultaneous equations and cointegration, Journal of Econometrics, 69, 111-132.

Johansen, S. (1995b) Likelihood-based inference in cointegrated vector autoregressive models, Oxford University Press, Oxford.

Larsson, R., J. Lyhagen and M. Löthgren (1998) Likelihood-Based Cointegration Tests in Heterogeneous Panels, Working Paper Series in Economics and Finance No. 250, Stockholm School of Economics.

Levin, A. and C.-F. Lin (1992) Unit root test in panel data: Asymptotics and finite-sample properties, Discussion Paper 92-23, University of California at San Diego.

Levin, A. and C.-F. Lin (1993) Unit root test in panel data: New results, Discussion Paper 93-56, University of California at San Diego.

Magnus, J. and H. Neudecker (1988) Matrix differential calculus with applications in statistics and econometrics, Wiley, Chichester.

Muirhead, R. J. (1982) Aspects of multivariate statistical theory, Wiley, New York.

Pesaran, M.H., Y. Shin and R.P. Smith (1999) Pooled mean group estimation of dynamic heterogeneous panels, Journal of the American Statistical Association, 94, 621-634.

8. Appendix: omitted proofs

Let us start by giving the definitions (cf Johansen, 1995b)

$$\mathrm{Var}\begin{pmatrix} \Delta Y_t \\ B'Y_{t-1} \end{pmatrix} |\Delta Y_{t-1},...,\Delta Y_{t-m+1}\Bigg) \equiv \begin{pmatrix} \Sigma_{00} & \Sigma_{0B} \\ \Sigma_{B0} & \Sigma_{BB} \end{pmatrix},$$

$G_t \equiv \overline{B}'_\perp CW_t$, $C \equiv B_\perp \left(A'_\perp \Gamma B_\perp\right)^{-1} A'_\perp$, $\Gamma \equiv I_{Np} - \sum_{i=1}^{m-1} \Gamma_i$, $\overline{B}_\perp \equiv B_\perp \left(B'_\perp B_\perp\right)^{-1}$ and $\overline{B} \equiv B\left(B'B\right)^{-1}$. Moreover, following Johansen (1995b), p. 91, we may concentrate out high order lag terms from (2.6) to obtain the auxiliary regression

$$R_{0t} = AB'R_{1t} + \widehat{\varepsilon}_t, \tag{8.1}$$

and define

$$S_{ij} \equiv T^{-1}\sum_{t=1}^{T} R_{it}R'_{jt}, \quad i,j = 0,1.$$

We then have the following lemma :

Lemma 8.1. *Under assumption A, we have that as* $T \to \infty$,

$$S_{00} \xrightarrow{P} \Sigma_{00}, \tag{8.2}$$

$$B'S_{11}B \xrightarrow{P} \Sigma_{BB}, \tag{8.3}$$

$$B'S_{10} \xrightarrow{P} \Sigma_{B0}, \tag{8.4}$$

$$T^{-1}\overline{B}'_\perp S_{11}\overline{B}_\perp \xrightarrow{w} \int GG', \tag{8.5}$$

$$\overline{B}'_\perp S_{1\varepsilon} \equiv \overline{B}'_\perp \left(S_{10} - S_{11}BA'\right) \xrightarrow{w} \int GdW', \tag{8.6}$$

$$\overline{B}'_\perp S_{11}B = O_P(1), \tag{8.7}$$

Proof. The lemma follows by a simple modification of the proof of Lemma 10.3 of Johansen (1995b). This proof builds upon the representation

$$Y_t = C\sum_{i=1}^{t} \varepsilon_i + U_t,$$

where U_t is an $I(0)$ process. ■

Before going on, we list some useful identities, to be found in e.g. Magnus and Neudecker (1988). For arbitrary matrices P, Q, R and S of dimensions such that the products below are defined, it holds that

$$\begin{aligned}
\operatorname{tr}(P) &= \operatorname{tr}(P'), && (8.8)\\
\operatorname{tr}(PQ) &= \operatorname{tr}(QP), && (8.9)\\
(P \otimes Q)(R \otimes S) &= PR \otimes QS, && (8.10)\\
\operatorname{tr}(P'Q) &= (\operatorname{vec}P)' \operatorname{vec}Q, && (8.11)\\
\operatorname{vec}(PQR) &= (R' \otimes P) \operatorname{vec}Q, && (8.12)\\
\operatorname{tr}(PQRS) &= (\operatorname{vec}S)' (P \otimes R') \operatorname{vec}Q', && (8.13)\\
(P \otimes Q)^{-1} &= P^{-1} \otimes Q^{-1}. && (8.14)
\end{aligned}$$

We will also make use of the identity

$$I = \Upsilon_{\perp}\overline{\Upsilon}_{\perp}' + \Upsilon\overline{\Upsilon}' = \overline{\Upsilon}_{\perp}\Upsilon_{\perp}' + \overline{\Upsilon}\Upsilon', \tag{8.15}$$

where I is an identity matrix. The first equality of (8.15) follows from the fact that left-hand multiplication of both sides by Υ' or by $\Upsilon_{\perp}'$ yield the same results on both sides of the equality sign. The second equality is a simple consequence of the definitions.

Proof of Theorem 4.1: As in Johansen (1995b), p. 91, concentrating out the Γ_i terms leads us to the auxiliary regression

$$R_{0t} = AB'R_{1t} + \widehat{\varepsilon}_t, \tag{8.16}$$

where the $\widehat{\varepsilon}_t$ are independent normals, each with mean zero and covariance matrix Ω. For a moment, let us assume that A and Ω are both fixed, the following arguments being applicable also when they are not, due to consistency. Then, apart from a constant, the log likelihood may be expressed as

$$\log L = -\frac{1}{2}\operatorname{tr}\left(\Omega^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_t'\right),$$

and because $d\widehat{\varepsilon}_t = -AH_i^{(r)}d\beta_i'H_i^{(p)\prime}R_{1t}$, it follows by (8.9) and (8.16) that

$$\begin{aligned}
T^{-1}d\log L &= -T^{-1}\operatorname{tr}\left(\Omega^{-1}\sum_{t=1}^{T}d\widehat{\varepsilon}_t\widehat{\varepsilon}_t'\right) = T^{-1}\operatorname{tr}\left(H_i^{(p)\prime}\sum_{t=1}^{T}R_{1t}\widehat{\varepsilon}_t'\Omega^{-1}AH_i^{(r)}d\beta_i'\right)\\
&= \operatorname{tr}\left\{H_i^{(p)\prime}\widehat{S}_{1\varepsilon}\Omega^{-1}AH_i^{(r)}d\beta_i'\right\},
\end{aligned}$$

where

$$\widehat{S}_{1\varepsilon} \equiv S_{10} - S_{11}\widehat{B}A' = S_{1\varepsilon} - S_{11}\left(\widehat{B} - B\right)A'.$$

Hence, putting the derivative w.r.t. β_i equal to zero, it follows that

$$H_i^{(p)\prime} S_{1\varepsilon}\Omega^{-1}AH_i^{(r)} = H_i^{(p)\prime} S_{11}\left(\widehat{B} - B\right)A'\Omega^{-1}AH_i^{(r)}, \tag{8.17}$$

for all i. Now, because

$$\overline{\beta}'_{i\perp}H_i^{(p)\prime} = \left(0, ..., 0, \overline{\beta}'_{i\perp}, 0, ..., 0\right) = H_i^{(p-r)\prime}\overline{B}'_{\perp}, \tag{8.18}$$

left-multiplication of (8.17) by $\overline{\beta}_{i\perp}$ and insertion of $\widehat{B} - B = \overline{B}_{\perp}X_T$ yields

$$H_i^{(p-r)\prime}\overline{B}'_{\perp}S_{1\varepsilon}\Omega^{-1}AH_i^{(r)} = H_i^{(p-r)\prime}\overline{B}'_{\perp}S_{11}\overline{B}_{\perp}X_TA'\Omega^{-1}AH_i^{(r)}.$$

To find vecX_T, we apply (8.12) to get

$$\begin{aligned} &\left(H_i^{(r)\prime} \otimes H_i^{(p-r)\prime}\right)\text{vec}\left(\overline{B}'_{\perp}S_{1\varepsilon}\Omega^{-1}A\right) \\ = \;&\left(H_i^{(r)\prime} \otimes H_i^{(p-r)\prime}\right)\left(A'\Omega^{-1}A \otimes \overline{B}'_{\perp}S_{11}\overline{B}_{\perp}\right)\text{vec}X_T \end{aligned} \tag{8.19}$$

for all i. Then, putting the (8.19) equations for each i "on top" of each other yields

$$K'\text{vec}\left(\overline{B}'_{\perp}S_{1\varepsilon}\Omega^{-1}A\right) = K'\left(A'\Omega^{-1}A \otimes \overline{B}'_{\perp}S_{11}\overline{B}_{\perp}\right)\text{vec}X_T. \tag{8.20}$$

Now, we may take

$$K_{\perp} = \left(H_{1\perp}^{(r)} \otimes H_1^{(p-r)}, ..., H_{N\perp}^{(r)} \otimes H_N^{(p-r)}\right) \tag{8.21}$$

with

$$H_{i\perp}^{(r)} = \begin{pmatrix} I_{r(i-1)} & 0_{r(i-1)\times r(N-i)} \\ 0_{r\times r(i-1)} & 0_{r\times r(N-i)} \\ 0_{r(N-i)\times r(i-1)} & I_{r(N-i)} \end{pmatrix}, \quad i = 1, ..., N.$$

Further, observe that because X_T is block diagonal, (8.12) implies

$$K'_{\perp}\text{vec}X_T = \begin{pmatrix} H_{1\perp}^{(r)\prime} \otimes H_1^{(p-r)\prime} \\ \vdots \\ H_{N\perp}^{(r)\prime} \otimes H_N^{(p-r)\prime} \end{pmatrix}\text{vec}X_T = \begin{pmatrix} \text{vec}\left(H_1^{(p-r)\prime}X_TH_{1\perp}^{(r)}\right) \\ \vdots \\ \text{vec}\left(H_N^{(p-r)\prime}X_TH_{N\perp}^{(r)}\right) \end{pmatrix} = 0, \tag{8.22}$$

so by (8.20) and (8.15),

$$K'\text{vec}\left(\overline{B}'_{\perp}S_{1\varepsilon}\Omega^{-1}A\right) = K'\left(A'\Omega^{-1}A \otimes T^{-1}\overline{B}'_{\perp}S_{11}\overline{B}_{\perp}\right)KT\overline{K}'\text{vec}X_T.$$

Hence, using the lemma we find

$$K'\text{vec}\left(\int GdW'\Omega^{-1}A\right) = TF_1\overline{K}'\text{vec}X_T + o_P(1),$$

with F_1 as defined above, upon which, by (8.15) and (8.22),

$$T\text{vec}X_T = KT\overline{K}'\text{vec}X_T = KF_1^{-1}K'\text{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P(1),$$

as was to be proved. ■

We now need an algebraic lemma.

Lemma 8.2. *Letting* $\Phi \equiv \Sigma_{BB} - \Sigma_{B0}\Sigma_{00}^{-1}\Sigma_{0B}$, *we have*

$$\Phi^{-1} = A'\Omega^{-1}A + \Sigma_{BB}^{-1}, \qquad (8.23)$$

$$A\Sigma_{BB} - \Sigma_{0B} = 0, \qquad (8.24)$$

$$\Sigma_{0B}A' = \Sigma_{00} - \Omega, \qquad (8.25)$$

$$\Sigma_{0B}\Sigma_{BB}^{-1} = A, \qquad (8.26)$$

$$\Sigma_{0B}\Phi^{-1} = \Sigma_{00}\Omega^{-1}A, \qquad (8.27)$$

Proof. The identities (8.23)-(8.25) are given in Lemma 10.1 of Johansen (1995b). Further,

$$\Sigma_{0B}\Sigma_{BB}^{-1} = -\left(A\Sigma_{BB} - \Sigma_{0B}\right)\Sigma_{BB}^{-1} + A = A,$$

proving (8.26), and (8.27) is a simple consequence of (8.23), (8.25) and (8.26). ■

Proof of Theorem 4.2: Consider the three hypotheses $H_3 : \text{rank}(\Pi) \leq Np$, $H_2 : \Pi = AB'$ where A, B are $Np \times Nr$, of full rank and H_1 : as H_2 but where B is block-diagonal with $p \times r$-dimensional blocks. Denoting the maximum likelihood ratio between H_i and H_j ($H_i \subset H_j$) by Q_{ij}, we then have $Q_{13} = Q_{12}Q_{23}$, i.e.

$$-2\log Q_{13} = -2\log Q_{12} - 2\log Q_{23}.$$

Johansen (1995b) has showed that the asymptotic distribution of $-2\log Q_{23}$ equals the distribution of U as defined in the theorem. (The fact that B has the specific block diagonal form under our hypothesis under test does not affect this result.) Now, to prove our theorem, our plan is

1) To show the convergence of $-2\log Q_{12}$ to the χ^2 distribution.

2) To show the asymptotic independence between $-2\log Q_{12}$ and $-2\log Q_{23}$.

1) It follows as in Johansen (1995b), p. 92 that, apart from a constant, the maximum likelihood under H_1, L_1 say, fulfills

$$L_1^{-2/T} = |S_{00}| \frac{\left|\widehat{B}' M \widehat{B}\right|}{\left|\widehat{B}' S_{11} \widehat{B}\right|}, \quad M \equiv S_{11} - S_{10} S_{00}^{-1} S_{01}, \tag{8.28}$$

where $\widehat{B}$ is the ML estimate of B under H_1. Below, we will use the identity

$$\widehat{B} = B + \left(\widehat{B} - B\right) = B + \overline{B}_{\perp} X_T, \tag{8.29}$$

together with the convergence result for X_T of theorem 4.1. Similarly, for L_2, the maximum likelihood under H_2, it holds that

$$L_2^{-2/T} = |S_{00}| \frac{\left|\widetilde{B}' M \widetilde{B}\right|}{\left|\widetilde{B}' S_{11} \widetilde{B}\right|}, \tag{8.30}$$

where $\widetilde{B}$ is the ML estimate of B under H_2. As in Johansen (1995b), p. 183, U_T defined through $\widetilde{B} - B = \overline{B}_{\perp} U_T$ fulfills

$$\begin{aligned} TU_T &= \left(T^{-1} \overline{B}_{\perp}' S_{11} \overline{B}_{\perp}\right)^{-1} \overline{B}' S_{1\varepsilon}' \Omega^{-1} A \left(A' \Omega^{-1} A\right)^{-1} + o_P(1) \\ &= \left(\int GG'\right)^{-1} \int G dW' \Omega^{-1} A \left(A' \Omega^{-1} A\right)^{-1} + o_P(1). \end{aligned} \tag{8.31}$$

Observe that this shows that $\widetilde{B}$ is consistent for B also under H_1. Further, as above we have

$$\widetilde{B} = B + \left(\widetilde{B} - B\right) = B + \overline{B}_{\perp} U_T, \tag{8.32}$$

so that from (8.29),

$$D \equiv \widehat{B} - \widetilde{B} = \overline{B}_{\perp} \left(X_T - U_T\right) = O_P\left(T^{-1}\right). \tag{8.33}$$

Now, from (8.28) and (8.30)

$$Q_{12}^{-2/T} = \frac{\left|\widehat{B}' M \widehat{B}\right|}{\left|\widetilde{B} M \widetilde{B}\right|} \left(\frac{\left|\widehat{B}' S_{11} \widehat{B}\right|}{\left|\widetilde{B} S_{11} \widetilde{B}\right|}\right)^{-1}, \tag{8.34}$$

where, because $\widehat{B} \equiv \widetilde{B} + D$,

$$\frac{\left|\widehat{B}'M\widehat{B}\right|}{\left|\widetilde{B}'M\widetilde{B}\right|} = \left|I + \left(\widetilde{B}'M\widetilde{B}\right)^{-1}\left(\widetilde{B}'MD + D'M\widetilde{B} + D'MD\right)\right| \tag{8.35}$$

and similarly with S_{11} in place of M. Here, because U_T is $O_P(T^{-1})$ (cf (8.31)), we have by (8.32) and lemma 8.1 that

$$\begin{aligned} \widetilde{B}'M\widetilde{B} &= B'MB + B'M\overline{B}_\perp U_T + U_T'\overline{B}_\perp' MB + U_T'\overline{B}_\perp' M\overline{B}_\perp U_T \\ &= B'MB + o_P(1) = \Phi + o_P(1), \end{aligned} \tag{8.36}$$

and similarly,

$$\widetilde{B}'S_{11}\widetilde{B} = \Sigma_{BB}^{-1} + o_P(1). \tag{8.37}$$

Moreover, (8.33) yields

$$\widetilde{B}'MD = \left(B' + U_T'\overline{B}_\perp'\right) M\overline{B}_\perp (X_T - U_T), \tag{8.38}$$

$$D'MD = (X_T - U_T)' \overline{B}_\perp' M\overline{B}_\perp (X_T - U_T), \tag{8.39}$$

and similarly with S_{11} in place of M. Now, X_T (cf theorem 4.1) and U_T are $O_P(T^{-1})$, and furthermore, via lemma 8.1, $B'M\overline{B}_\perp$ and $B'S_{11}\overline{B}_\perp$ are $O_P(1)$ and $\overline{B}_\perp' M\overline{B}_\perp$ and $\overline{B}_\perp' S_{11}\overline{B}_\perp$ are $O_P(T)$. Hence, via (8.35)-(8.39), we see that the r.h.s. of (8.34) is of the form

$$\frac{|I + T^{-1}C_1|}{|I + T^{-1}C_2|},$$

where C_1 and C_2 are $O_P(1)$, and using the Taylor expansions (cf Johansen (1995b), p. 224)

$$\left|I + T^{-1}C_i\right| = 1 + T^{-1}\mathrm{tr}C_i + O_P\left(T^{-2}\right), \quad i = 1, 2$$

and $\log(1+x) = x + O(x^2)$, in conjunction with (8.8) and (8.9), we arrive at

$$-2\log Q_{12} = T\mathrm{tr}(2\Theta_1 + 2\Theta_2 + \Theta_3) + o_P(1), \tag{8.40}$$

where

$$\begin{aligned} \Theta_1 &\equiv D_1(X_T - U_T), \\ \Theta_2 &\equiv D_2(X_T - U_T), \\ \Theta_3 &\equiv (B'MB)^{-1}(X_T - U_T)'\overline{B}_\perp' M\overline{B}_\perp (X_T - U_T) \\ &\quad - (B'S_{11}B)^{-1}(X_T - U_T)'\overline{B}_\perp' S_{11}\overline{B}_\perp (X_T - U_T). \end{aligned}$$

with

$$
\begin{aligned}
D_1 &\equiv (B'MB)^{-1} B'M\overline{B}_\perp - (B'S_{11}B)^{-1} B'S_{11}\overline{B}_\perp, \\
D_2 &\equiv (B'MB)^{-1} U_T'\overline{B}_\perp' M\overline{B}_\perp - (B'S_{11}B)^{-1} U_T'\overline{B}_\perp' S_{11}\overline{B}_\perp.
\end{aligned}
$$

We will now show that D_1 and D_2 cancel out each other asymptotically. To this end, lemma 8.1 yields

$$
D_1 = \Phi^{-1}\left(B'S_{11} - \Sigma_{B0}\Sigma_{00}^{-1}S_{01}\right)\overline{B}_\perp - \Sigma_{BB}^{-1}B'S_{11}\overline{B}_\perp + o_P(1).
$$

Moreover, by lemma 8.2,

$$
\begin{aligned}
\Phi^{-1} &= A'\Omega^{-1}A + \Sigma_{BB}^{-1}, \\
\Phi^{-1}\Sigma_{B0}\Sigma_{00}^{-1} &= A'\Omega^{-1},
\end{aligned}
$$

so that

$$
\begin{aligned}
D_1 &= -\Phi^{-1}\Sigma_{B0}\Sigma_{00}^{-1}S_{01}\overline{B}_\perp + A'\Omega^{-1}AB'S_{11}\overline{B}_\perp + o_P(1) \\
&= -A'\Omega^{-1}\left(S_{01} - AB'S_{11}\right)\overline{B}_\perp + o_P(1) \\
&= -A'\Omega^{-1}S_{\varepsilon 1}\overline{B}_\perp + o_P(1) \\
&= -A'\Omega^{-1}\left(\int GdW'\right)' + o_P(1), \qquad (8.41)
\end{aligned}
$$

where the last equality follows from lemma 8.1. Furthermore, because by lemma 8.1, $\overline{B}_\perp' M\overline{B}_\perp$ and $\overline{B}_\perp' S_{11}\overline{B}_\perp$ asymptotically both behave like $T\int GG'$, it follows as above and from (8.31) that

$$
\begin{aligned}
D_2 &= \left(\Phi^{-1} - \Sigma_{BB}^{-1}\right)\left(A'\Omega^{-1}A\right)^{-1}A'\Omega^{-1}\left(\int GdW'\right)'\left(\int GG'\right)^{-1}\int GG' + o_P(1) \\
&= A'\Omega^{-1}\left(\int GdW'\right)' + o_P(1),
\end{aligned}
$$

which behaves like $-D_1$, as asserted. Hence, (8.40) simplifies into

$$
-2\log Q_{12} = T\mathrm{tr}\Theta_3 + o_P(1).
$$

As for Θ_3, we at first find in a similar manner as above that

$$
T\Theta_3 = \left(A'\Omega^{-1}A\right)T\left(X_T - U_T\right)'\left(\int GG'\right)T\left(X_T - U_T\right) + o_P(1),
$$

and (8.13) implies

$$-2\log Q_{12} = T\left\{\mathrm{vec}\left(X_T - U_T\right)\right\}' JT\mathrm{vec}\left(X_T - U_T\right) + o_P\left(1\right), \tag{8.42}$$

where

$$J \equiv A'\Omega^{-1}A \otimes \int GG'.$$

Further, we get via (8.12), (8.14) and (8.31) that

$$T\mathrm{vec}U_T = J^{-1}\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P\left(1\right), \tag{8.43}$$

which, combined with theorem 4.1, yields

$$T\mathrm{vec}\left(X_T - U_T\right) = -P\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P\left(1\right), \tag{8.44}$$

where

$$\begin{aligned} P &\equiv J^{-1} - KF_1^{-1}K' = J^{-1} - K\left(K'JK\right)^{-1}K' \\ &= J^{-1}K_\perp\left(K_\perp'J^{-1}K_\perp\right)^{-1}K_\perp'J^{-1}. \end{aligned} \tag{8.45}$$

(The last equality holds because left-hand multiplication by $K'J$ or $K_\perp'$ yields the same result on both sides.) Hence, because $P'JP = P$, (8.42) becomes

$$\begin{aligned} &-2\log Q_{12} \\ =\ & \mathrm{vec}\left(\int GdW'\Omega^{-1}A\right)' P\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P\left(1\right) \\ =\ & \mathrm{vec}\left(\int GdW'\Omega^{-1}A\right)' J^{-1}K_\perp\left(K_\perp'J^{-1}K_\perp\right)^{-1}K_\perp'J^{-1}\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right) \\ & +o_P\left(1\right). \end{aligned} \tag{8.46}$$

Now, let us for a while condition on G. Then, $\mathrm{vec}(\int GdW'\Omega^{-1}A)$ has covariance matrix J. This is seen because $\Omega^{-1/2}W$ is a process with unit covariance matrix, and so (8.12) and (8.10) imply

$$\begin{aligned} & \mathrm{E}\left[\left\{\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right)\right\}\left\{\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right)\right\}'\right] \\ =\ & \int\left(A'\Omega^{-1/2}\otimes G\right)\left(\Omega^{-1/2}A\otimes G'\right) = J. \end{aligned}$$

Hence, conditional on G, $K_\perp' J^{-1} \text{vec}(\int GdW'\Omega^{-1}A)$ is normal with expectation 0 and covariance matrix $K_\perp' J^{-1} K_\perp$. Therefore, the leading term of (8.46) is χ^2, and since this distribution is independent of G, this property holds also unconditionally. Thus, convergence to a χ^2 distribution is shown. Moreover, the number of degrees of freedom equals the dimension of $K_\perp' J^{-1} K_\perp$, i.e. the number of columns of $K_\perp$, which via (8.21) is seen to be $N(N-1)(p-r)r$.

2) From Johansen (1995b), p. 158-160, we deduce the representation

$$-2\log Q_{23} = \text{tr}\left\{\left(\int GG'\right)^{-1} \int GdW'A_\perp \left(A_\perp'\Omega A_\perp\right)^{-1} A_\perp' \left(\int GdW'\right)'\right\} + o_P(1), \tag{8.47}$$

where W and G are as above. We need to show that the main terms of (8.46) and (8.47), M_1 and M_2 say, are independent. Now, by (8.9), (8.13) and (8.14), (8.47) may be re-written as

$$\begin{aligned} & -2\log Q_{23} \\ = \;& \text{tr}\left\{\left(A_\perp'\Omega A_\perp\right)^{-1} A_\perp' \left(\int GdW'\right)' \left(\int GG'\right)^{-1} \int GdW'A_\perp\right\} + o_P(1) \\ = \;& \text{vec}\left(\int GdW'A_\perp\right)' \left(A_\perp'\Omega A_\perp \otimes \int GG'\right)^{-1} \text{vec}\left(\int GdW'A_\perp\right) + o_P(1). \end{aligned}$$

Let us again condition on G. Then, $\int GdW'\Omega^{-1}A$ and $\int GdW'A_\perp$ are both normals, each with expectation zero, and the covariance between them is

$$\text{E}\left\{\int GdW'\Omega^{-1}A\left(\int GdW'A_\perp\right)'\right\} = 0,$$

showing that $\int GdW'\Omega^{-1}A$ and $\int GdW'A_\perp$ are conditionally independent given G. Hence, M_1 and M_2 must also be conditionally independent given G. Furthermore, as we saw earlier, M_1 is independent of G. Hence we get, denoting the densities for M_1 and M_2 by f_1 and f_2, their simultaneous density by $f_{1,2}$, the density of G by f_G and the corresponding conditional densities by $f_{1|G}$ etcetera,

$$f_{1,2} = \int f_{1,2|G} f_G = \int f_{1|G} f_{2|G} f_G = f_1 \int f_{2|G} f_G = f_1 f_2,$$

where the integrals are over the support of the G density. This shows the independency between M_1 and M_2, and we are done. ■

Proof of Theorem 4.3: Denote by B^* the ML estimate of B under the present H_0. Under this H_0, we find as in the proof of Theorem 4.1 that

$$T^{-1} d \log L = \mathrm{tr} \left\{ \sum_{i=1}^{N} H_i^{(p)\prime} S_{1\varepsilon}^* \Omega^{-1} A H_i^{(r)} d\beta' \right\},$$

where $S_{1\varepsilon}^* \equiv S_{10} - S_{11} B^* A'$. But via (8.8), (8.9), (8.11) and (8.12), we get

$$\begin{aligned} T^{-1} d \log L &= \sum_{i=1}^{N} \mathrm{tr} \left\{ \left(H_i^{(p)\prime} S_{1\varepsilon}^* \Omega^{-1} A H_i^{(r)} \right)' d\beta \right\} \\ &= \sum_{i=1}^{N} \left\{ \mathrm{vec} \left(H_i^{(p)\prime} S_{1\varepsilon}^* \Omega^{-1} A H_i^{(r)} \right) \right\}' \mathrm{vec}\,(d\beta) \\ &= \sum_{i=1}^{N} \left\{ \left(H_i^{(r)\prime} \otimes H_i^{(p)\prime} \right) \mathrm{vec} \left(S_{1\varepsilon}^* \Omega^{-1} A \right) \right\}' \mathrm{vec}\,(d\beta) \\ &= \left\{ \mathrm{vec} \left(S_{1\varepsilon}^* \Omega^{-1} A \right) \right\}' K_{r,p}^* \mathrm{vec}\,(d\beta)\,, \end{aligned}$$

where the $N^2 pr \times pr$ matrix $K_{r,p}^*$ is defined through

$$K_{r,p}^* \equiv \sum_{i=1}^{N} \left(H_i^{(r)} \otimes H_i^{(p)} \right).$$

Hence, because $S_{1\varepsilon}^* \equiv S_{1\varepsilon} - S_{11} \left(B^* - B \right) A'$, we get the equality

$$K_{r,p}^{*\prime} \mathrm{vec} \left(S_{1\varepsilon} \Omega^{-1} A \right) = K_{r,p}^{*\prime} \mathrm{vec} \left\{ S_{11} \left(B^* - B \right) A' \Omega^{-1} A \right\},$$

or, writing $B^* - B = \overline{B}_\perp X_T^*$ and using (8.12),

$$K_{r,p}^{*\prime} \mathrm{vec} \left(S_{1\varepsilon} \Omega^{-1} A \right) = K_{r,p}^{*\prime} \left(A' \Omega^{-1} A \otimes S_{11} \overline{B}_\perp \right) \mathrm{vec} X_T^*. \tag{8.48}$$

But applying (8.18) with β in place of β_i, it follows via (8.10) that

$$\left(I_r \otimes \overline{\beta}_\perp' \right) K_{r,p}^{*\prime} = \sum_{i=1}^{N} \left(H_i^{(r)\prime} \otimes H_i^{(p-r)\prime} \overline{B}_\perp' \right) = K_{r,p-r}^{*\prime} \left(I_r \otimes \overline{B}_\perp' \right),$$

and so, left-multiplying (8.48) by $I_r \otimes \overline{\beta}_\perp'$ and using (8.10) and (8.12), we find (from now on, we put $K^* \equiv K_{r,p-r}^*$)

$$K^{*\prime} \mathrm{vec} \left(\overline{B}_\perp' S_{1\varepsilon} \Omega^{-1} A \right) = K^{*\prime} \left(A' \Omega^{-1} A \otimes \overline{B}_\perp' S_{11} \overline{B}_\perp \right) \mathrm{vec} X_T^*. \tag{8.49}$$

Now, note that by construction,

$$K^* = K\left(1_N \otimes I_{r(p-r)}\right), \tag{8.50}$$

where 1_N denotes an N-dimensional vector with elements 1. Hence, we may define

$$K_\perp^* = K_\perp\left(1_{N(N-1)} \otimes I_{r(p-r)}\right),$$

so that via (8.22) with X_T^* in place of X_T, $K_\perp^{*\prime}\text{vec}X_T^* = 0$, and (8.15) and (8.49) yield

$$K^{*\prime}\text{vec}\left(\overline{B}_\perp' S_{1\varepsilon}\Omega^{-1}A\right) = K^{*\prime}\left(A'\Omega^{-1}A \otimes T^{-1}\overline{B}_\perp' S_{11}\overline{B}_\perp\right) K^* T\overline{K}^{*\prime}\text{vec}X_T^*,$$

and consequently, by Lemma 8.1,

$$T\text{vec}X_T^* = K^* T\overline{K}^{*\prime}\text{vec}X_T^* = K^* F_2^{-1} K^{*\prime}\text{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P(1), \tag{8.51}$$

where

$$F_2 \equiv K^{*\prime} J K^*, \quad J = A'\Omega^{-1}A \otimes \int GG'.$$

Now, let H_1 be the alternative hypothesis, i.e. the hypothesis that $\Pi = AB'$ with B block diagonal. (This is the same H_1 as in the proof of the previous theorem.) Our idea is to relate the maximum likelihood ratio between H_0 and H_1, Q_{01} say, to Q_{02} and Q_{12}, which are defined accordingly with H_2 being $\Pi = AB'$ with B unrestricted with rank Nr as in the previous proof, via $Q_{01} = Q_{02}/Q_{12}$, i.e.

$$-2\log Q_{01} = -2\log Q_{02} - (-2\log Q_{12}). \tag{8.52}$$

We already know $-2\log Q_{12}$ from the previous proof. Moreover, $-2\log Q_{02}$ may be derived in a similar fashion as $-2\log Q_{12}$ was found there. The analogy takes us as far as to the representation (cf (8.42))

$$-2\log Q_{02} = T\left\{\text{vec}\left(X_T^* - U_T\right)\right\}' JT\text{vec}\left(X_T^* - U_T\right) + o_P(1). \tag{8.53}$$

Here, (8.43) and (8.51) yield

$$T\text{vec}\left(X_T^* - U_T\right) = -P^*\text{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P(1), \tag{8.54}$$

where

$$P^* = J^{-1} - K^* F_2^{-1} K^{*\prime}.$$

Now, in analogy with $P'JP = P$ (cf (8.45)), we also have $P^{*\prime}JP^* = P^*$, so from (8.44), (8.42), (8.54), (8.53) and (8.52) we find

$$-2\log Q_{01} = \mathrm{vec}\left(\int GdW'\Omega^{-1}A\right)' Z\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right) + o_P(1), \qquad (8.55)$$

where

$$Z \equiv P^{*\prime}JP^* - P'JP = P^* - P = KF_1^{-1}K' - K^*F_2^{-1}K^{*\prime}.$$

Furthermore, putting $x \equiv 1_N \otimes I_{r(p-r)}$, we have by (8.50) that $K^* = Kx$, $F_2 = x'F_1x$, and so

$$Z = K\left\{F_1^{-1} - x\left(x'F_1x\right)^{-1}x'\right\}K'.$$

Now, similar to (8.45), we have

$$F_1^{-1} - x\left(x'F_1x\right)^{-1}x' = F_1^{-1}x_\perp\left(x'_\perp F_1^{-1}x_\perp\right)^{-1}x'_\perp F_1^{-1},$$

and so, we find

$$Z = KF_1^{-1}x_\perp\left(x'_\perp F_1^{-1}x_\perp\right)^{-1}x'_\perp F_1^{-1}K'.$$

Hence, letting

$$Y \equiv K'\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right),$$

we may re-write (8.55) as

$$-2\log Q_{01} = Y'F_1^{-1}x_\perp\left(x'_\perp F_1^{-1}x_\perp\right)^{-1}x'_\perp F_1^{-1}Y + o_P(1).$$

The result that $-2\log Q_{01}$ is χ^2 follows in the same way as in part 2) of the proof of the previous theorem, if we can prove that, conditioning on G, Y is normal with expectation 0 and covariance matrix F_1. To this end, note that from the definition of K we have $Y' = \left(Y_1', ..., Y_N'\right)$ with

$$Y_i \equiv \left(H_i^{(r)\prime} \otimes H_i^{(p-r)\prime}\right)\mathrm{vec}\left(\int GdW'\Omega^{-1}A\right).$$

But, as we saw in part 2) of the proof of the previous theorem, conditioned on G, $\mathrm{vec}(\int GdW'\Omega^{-1}A)$ has covariance matrix J. Hence,

$$\mathrm{E}\left(Y_iY_j'\right) = \left(H_i^{(r)\prime} \otimes H_i^{(p-r)\prime}\right)J\left(H_j^{(r)} \otimes H_j^{(p-r)}\right),$$

i.e.

$$\mathrm{E}\left(YY'\right) = K'JK = F_1,$$

as was to be shown.

Finally, we note that the number of degrees of freedom of our χ^2 distribution equals the dimension of $x'_\perp F_1^{-1}x_\perp$, which is $(N-1)\,r\,(p-r)$. ■