Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$ in the Gaussian Case, with Application to Likelihood-Based Cointegration Analysis

Herman J. Bierens
Pennsylvania State University

March 3, 2007

Abstract

Phillips (1988) has set forth conditions on a k-variate time series process x_t such that, with $S_t = \sum_{j=1}^{t} x_j$, $\frac{1}{T}\sum_{t=1}^{T} S_{t-1}x_t'$ converges in distribution to the stochastic matrix $\Sigma^{1/2}\left(\int_0^1 BdB'\right)\Sigma^{1/2} + \Sigma_1'$, where B is a k-variate standard Brownian motion on the unit interval, $\Sigma = \lim_{T\to\infty} E\left[\frac{1}{T}S_T S_T'\right]$, and $\Sigma_1' = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left[S_{t-1}x_t'\right]$. However, Phillips' derivation of this result is very complicated. Therefore I will give an alternative proof for the case that x_t is a covariance stationary Gaussian process, employing only fairly standard probability theory. The result will be used to explain Johansen's (1988, 1991, 1995) likelihood-based cointegration analysis, in vector error correction models without and with intercepts.

This is a lecture note rather than a research paper. Therefore, the results are not mine; they are all due to Clive Granger, Soren Johansen and Peter Phillips.

1 Introduction

Let $u_t \in \mathbb{R}^k$ be an i.i.d. white noise vector time series process with $E[u_t u_t'] = I_k$, and let $S_t = \sum_{j=1}^{t} u_j,\ t \geq 1,\ S_0 = 0$. Moreover, denote for $x \in [0,1]$, $B_T(x) = \frac{1}{\sqrt{T}} S_{[x.T]}$. Under more general conditions Phillips (1988) has shown that the random matrix $M_T = \frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1}'$ converges in distribution to a random matrix M represented by the integral $M = \int_0^1 (dB)\, B'$, where B is a k-variate standard Brownian motion on $[0,1]$ (also called a k-variate standard Wiener process).

More generally, Phillips (1988) has set forth conditions on a k-variate time series process x_t such that, with $S_t = \sum_{j=1}^{t} x_j$, $\frac{1}{T}\sum_{t=1}^{T} S_{t-1} x_t'$ converges in distribution to the stochastic matrix $\Sigma^{1/2}\left(\int_0^1 BdB'\right)\Sigma^{1/2} + \Sigma_1'$, where again B is a k-variate standard Brownian motion on the unit interval, with $\Sigma = \lim_{T\to\infty} E\left[\frac{1}{T} S_T S_T'\right]$, which is know as the long-run variance of x_t, and $\Sigma_1' = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[S_{t-1}x_t']$.

This result and its generalizations in Section 2 are especially important for cointegration theory. However, Phillips' derivation of this result is very complicated. Therefore, in this lecture note I will give an alternative proof for the case that x_t is a covariance stationary Gaussian process, employing only fairly standard probability theory. After discussing in Section 3 the cointegration phenomenon and the Granger representation theorem, the results will then be used in Section 4 to explain Johansen's (1988, 1991, 1995) likelihood-based cointegration analysis, in particular the asymptotic theory of Johansen's lambda-max and trace tests for the cointegrating rank.

2 Weak convergence to the matrix stochastic integral $\int_0^1 BdB'$

2.1 About the notation $\int_0^1 BdB'$

The reason for denoting the limiting distribution $M = \int_0^1 BdB'$ of $M_T = \frac{1}{T}\sum_{t=1}^{T} S_{t-1}u_t'$ as an integral is that we can write

$$M_T' = \frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1}' = \frac{1}{T}\sum_{t=1}^{T}(S_t - S_{t-1})S_{t-1}'$$

$$\begin{aligned}
&= \sum_{t=1}^{T} \left(B_T(t/T) - B_T((t-1)/T)\right) B_T((t-1)/T)' \\
&= \sum_{t=1}^{T} \int_{t-1}^{t} \left(B_T(\tau/T + 1/T) - B_T(\tau/T)\right) B_T(\tau/T)' d\tau \\
&= \int_{0}^{T} \left(B_T(\tau/T + 1/T) - B_T(\tau/T)\right) B_T(\tau/T)' d\tau \\
&= T \int_{0}^{1} \left(B_T(x + 1/T) - B_T(x)\right) B_T(x)' dx \\
&= \int_{0}^{1} \left(\frac{B_T(x + 1/T) - B_T(x)}{1/T}\right) B_T(x)' dx \\
&= \int_{0}^{1} \left(dB_T(x)/dx\right) B_T(x)' dx = \int_{0}^{1} \left(dB_T\right) B_T',
\end{aligned}$$

say, where $dB_T(x)/dx$ is defined as

$$dB_T(x)/dx = \frac{B_T(x + 1/T) - B_T(x)}{1/T}.$$

Note, however, that

$$dB_T(x)/dx = \sqrt{T}\left(S_{[(x+1/T)T]} - S_{[x.T]}\right) = \sqrt{T} u_{[xT+1]},$$

which does not converge weakly. Therefore, we cannot conclude from the continuous mapping theorem that[1]

$$M_T = \frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1}' = \int_{0}^{1} \left(dB_T\right) B_T' \Rightarrow \int_{0}^{1} (dB)B' = M. \tag{1}$$

Nevertheless, the convergence result (1) holds for some random matrix M.

In the case $k = 1$ the meaning of M follows from the results in Phillips (1986): Note that $S_t^2 = (u_t + S_{t-1})^2 = u_t^2 + 2u_t S_{t-1} + S_{t-1}^2$, so that

$$M_T \;=\; \frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1} = \frac{1}{2T}\sum_{t=1}^{T}\left(S_t^2 - S_{t-1}^2 - u_t^2\right)$$

[1] In the sequel, "$\Rightarrow$" denotes convergence in distribution as well as weak convergence.

$$= \frac{1}{2}\left(\left(S_T/\sqrt{T}\right)^2 - \frac{1}{T}\sum_{t=1}^{T} u_t^2\right)$$

$$= \frac{1}{2}\left(B_T(1)^2 - 1\right) - \frac{1}{T}\sum_{t=1}^{T}\left(u_t^2 - 1\right) \Rightarrow \frac{1}{2}\left(B(1)^2 - 1\right).$$

The convergence result involved follows from the central limit theorem, i.e., $B_T(1) = S_T/\sqrt{T} = \left(1/\sqrt{T}\right)\sum_{t=1}^{T} u_t \Rightarrow N(0,1) \sim B(1)$, and the law of large numbers: $p\lim_{T\to\infty}(1/T)\sum_{t=1}^{T}(u_t^2 - 1) = 0$. Thus, in the case $k = 1$,

$$M = \int_0^1 (dB)\, B = \int_0^1 B dB = \frac{1}{2}\left(B(1)^2 - 1\right).$$

Along the same lines we find that in the multivariate case,

$$\begin{aligned} M_T + M_T' &= \frac{1}{T}\sum_{t=1}^{T} S_{t-1}u_t' + \frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1}' \\ &= \frac{1}{T} S_T S_T' - \frac{1}{T}\sum_{t=1}^{T} u_t u_t' \\ &= B_T(1)B_T(1)' - I_k - \left(\frac{1}{T}\sum_{t=1}^{T} u_t u_t' - I_k\right) \\ &\Rightarrow B(1)B(1)' - I_k = M + M'. \end{aligned} \tag{2}$$

2.2 The bivariate Gaussian white noise case

Consider the Gaussian white noise case $k = 2$, i.e., $x_t = u_t$, where

$$\begin{aligned} u_t &= \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \sim i.i.d.\ N_2\left[0, I_2\right], \\ S_t &= \begin{pmatrix} S_{1,t} \\ S_{2,t} \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{t} u_{1,j} \\ \sum_{j=1}^{t} u_{2,j} \end{pmatrix}, \\ B_T(x) &= \begin{pmatrix} B_{1,T}(x) \\ B_{2,T}(x) \end{pmatrix} = \begin{pmatrix} S_{1,[xT]}/\sqrt{T} \\ S_{2,[xT]}/\sqrt{T} \end{pmatrix}. \end{aligned}$$

It follows from (2) that

$$\frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{1,t-1} = \frac{1}{2}\left(B_{1,T}(1)^2 - 1\right) - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(u_{1,t}^2 - 1\right)$$

$$\frac{1}{T}\sum_{t=1}^{T} u_{2,t}S_{2,t-1} \quad = \quad \frac{1}{2}\left(B_{2,T}(1)^2 - 1\right) - \frac{1}{2}\frac{1}{T}\sum_{t=1}^{T}\left(u_{2,t}^2 - 1\right)$$

and

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} u_{2,t}S_{1,t-1} &= B_{1,T}(1)B_{2,T}(1) - \frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{2,t-1} - \frac{1}{T}\sum_{t=1}^{T} u_{1,t}u_{2,t} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{1,t}\left(B_{2,T}(1) - \frac{S_{2,t-1}}{\sqrt{T}}\right) - \frac{1}{T}\sum_{t=1}^{T} u_{1,t}u_{2,t}\end{aligned}$$

Thus

$$\frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1}' = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{1,t-1} & \frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{2,t-1} \\ \frac{1}{T}\sum_{t=1}^{T} u_{2,t}S_{1,t-1} & \frac{1}{T}\sum_{t=1}^{T} u_{2,t}S_{2,t-1} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}\left(B_{t,T}(1)^2 - 1\right) & \frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{2,t-1} \\ B_{1,T}(1)B_{2,T}(1) - \frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{2,t-1} & \frac{1}{2}\left(B_{2,T}(1)^2 - 1\right) \end{pmatrix} \quad (3)$$

$$-\frac{1}{2}\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\left(u_{1,t}^2 - 1\right) & 0 \\ 2\frac{1}{T}\sum_{t=1}^{T} u_{1,t}u_{2,t} & \frac{1}{T}\sum_{t=1}^{T}\left(u_{2,t}^2 - 1\right) \end{pmatrix}. \quad (4)$$

The elements of the matrix (3) are functions of

$$Z_T = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{1,t} \\ \frac{1}{T}\sum_{t=1}^{T} u_{1,t}S_{2,t-1} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{2,t} \end{pmatrix} \quad (5)$$

and the matrix (4) converges in probability to a zero matrix. Therefore, we only need to show that (5) converges in distribution, as follows. Write

$$Z_T = \begin{pmatrix} 0 \\ 0 \\ B_{2,T}(1) \end{pmatrix} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{1,t}\begin{pmatrix} 1 \\ S_{2,t-1}/\sqrt{T} \\ 0 \end{pmatrix}.$$

Then conditionally on is $u_{2,1},...,u_{2,T}$, Z_T is normally distributed with expectation vector

$$\mu_T = \begin{pmatrix} 0 \\ 0 \\ B_{2,T}(1) \end{pmatrix}$$

and singular variance matrix

$$\Sigma_T = \begin{pmatrix} \Phi_{2,T} & 0 \\ 0' & 0 \end{pmatrix}$$

where

$$\Phi_{2,T} = \begin{pmatrix} 1 & \int_0^1 B_{2,T}(x)dx \\ \int_0^1 B_{2,T}(x)dx & \int_0^1 B_{2,T}(x)^2 dx \end{pmatrix}.$$

Therefore, the characteristic function of Z_T conditional on $u_{2,1}, ..., u_{2,T}$ is

$$\begin{aligned} & E\left[\exp\left(i.\xi' Z_T\right) \middle| u_{2,1}, ..., u_{2,T}\right] = \exp\left(i.\xi'\mu_T\right)\exp\left(-\frac{1}{2}\xi'\Sigma_T\xi\right) \\ = & \exp\left(i.\xi_3 B_{2,T}(1)\right)\exp\left(-\frac{1}{2}\left(\xi_1,\xi_2\right)\Phi_{2,T}\left(\xi_1,\xi_2\right)'\right) \\ = & \exp(-\frac{1}{2}\xi_1^2)\exp\left(\xi_1\xi_2\int_0^1 B_{2,T}(x)dx - \frac{1}{2}\xi_2^2\int_0^1 B_{2,T}(x)^2 dx\right) \\ & \times\exp\left(i.\xi_3 B_{2,T}(1)\right), \end{aligned}$$

where $\xi' = (\xi_1, \xi_2, \xi_3)$. Because $B_{2,T}(1)$ and $\Phi_{2,T}$ converges jointly in distribution to $B_2(1)$ and

$$\Phi_2 = \begin{pmatrix} 1 & \int_0^1 B_2(x)dx \\ \int_0^1 B_2(x)dx & \int_0^1 B_2(x)^2 dx \end{pmatrix},$$

respectively, it follows from the continuous mapping theorem that pointwise in ξ,

$$\begin{aligned} & \exp(-\frac{1}{2}\xi_1^2)\exp\left(i.\xi_3 B_{2,T}(1)\right) \\ & \times\exp\left(\xi_1\xi_2\int_0^1 B_{2,T}(x)dx - \frac{1}{2}\xi_2^2\int_0^1 B_{2,T}(x)^2 dx\right) \\ & \Rightarrow \exp(-\frac{1}{2}\xi_1^2)\exp\left(i.\xi_3 B_2(1)\right) \\ & \times\exp\left(\xi_1\xi_2\int_0^1 B_2(x)dx - \frac{1}{2}\xi_2^2\int_0^1 B_2(x)^2 dx\right) \end{aligned}$$

This result implies that

$$\begin{aligned} & \lim_{T\to\infty} E\left[\exp\left(i.\xi' Z_T\right)\right] = \exp\left(-\frac{1}{2}\xi_1^2\right) \qquad (6) \\ & \times E\left[\exp\left(i.\xi_3 B_2(1)\right)\exp\left(\xi_1\xi_2\int_0^1 B_2(x)dx - \frac{1}{2}\xi_2^2\int_0^1 B_2(x)^2 dx\right)\right], \end{aligned}$$

because convergence in distribution of bounded random variables implies convergence of their expectations. Consequently, $Z_T \Rightarrow Z = (Z_1, Z_2, Z_3)'$, where Z is a random vector with characteristic function (6). Hence,

$$\begin{aligned} M &= \int_0^1 (dB)B' = \begin{pmatrix} \frac{1}{2}\left(Z_1^2 - 1\right) & Z_2 \\ Z_1 Z_3 - Z_2 & \frac{1}{2}\left(Z_3^2 - 1\right) \end{pmatrix}, \\ M' &= \int_0^1 B dB' = \begin{pmatrix} \frac{1}{2}\left(Z_1^2 - 1\right) & Z_1 Z_3 - Z_2 \\ Z_2 & \frac{1}{2}\left(Z_3^2 - 1\right) \end{pmatrix}. \end{aligned}$$

2.3 The tri-variate Gaussian white noise case

Now suppose that $u_{2,t} \in \mathbb{R}^2$. Then

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T} u_t S_{t-1}' = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} u_{1,t} S_{1,t-1} & \frac{1}{T}\sum_{t=1}^{T} u_{1,t} S_{2,t-1}' \\ \frac{1}{T}\sum_{t=1}^{T} u_{2,t} S_{1,t-1} & \frac{1}{T}\sum_{t=1}^{T} u_{2,t} S_{2,t-1}' \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2}\left(B_{1,T}(1)^2 - 1\right) & \frac{1}{T}\sum_{t=1}^{T} u_{1,t} S_{2,t-1}' \\ B_{1,T}(1)B_{2,T}(1)' - \frac{1}{T}\sum_{t=1}^{T} u_{1,t} S_{2,t-1}' & \frac{1}{T}\sum_{t=1}^{T} u_{2,t} S_{2,t-1}' \end{pmatrix} \\ &-\frac{1}{2}\begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\left(u_{1,t}^2 - 1\right) & 0 \\ 2\frac{1}{T}\sum_{t=1}^{T} u_{1,t} u_{2,t}' & O \end{pmatrix}. \end{aligned}$$

Stack the four elements of $\frac{1}{T}\sum_{t=1}^{T} u_{2,t} S_{2,t-1}'$ in a vector ψ_T, and let

$$Z_T = \begin{pmatrix} B_{1,T}(1) \\ \frac{1}{T}\sum_{t=1}^{T} u_{1,t} S_{2,t-1} \\ \psi_T \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{1,t}\begin{pmatrix} 1 \\ \frac{1}{\sqrt{T}} S_{2,t-1} \end{pmatrix} \\ \psi_T \end{pmatrix}$$

Then

$$\begin{aligned} &E\left[\exp\left(i.\xi' Z_T\right) \mid u_{2,1}, ..., u_{2,T}\right] = \exp\left(i.\left(\xi_3' \psi_T\right)\right) \\ &\times \exp\left(-\frac{1}{2}\left(\xi_1, \xi_2'\right)\begin{pmatrix} 1 & \int_0^1 B_{2,T}(x)' dx \\ \int_0^1 B_{2,T}(x) dx & \int_0^1 B_{2,T}(x) B_{2,T}(x)' dx \end{pmatrix}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}\right), \end{aligned}$$

where $\xi = (\xi_1, \xi_2', \xi_3')'$ has been partitioned conformably with Z_T. But by the previous argument, all the random elements involved converge jointly in distribution, hence $E\left[\exp\left(i.\xi' Z_T\right)\right]$ converges to a characteristic function.

The k-variate Gaussian case follows now by induction.

Summarizing, the following result has been shown.

Theorem 1. *Let* u_t *be i.i.d.* $N[0, I_k]$ *and let* $S_t = \sum_{j=1}^{t} u_j$. *Then the random matrix* $(1/T) \sum_{t=1}^{T} u_t S'_{t-1}$ *converges in distribution to a random matrix, denoted by* $\int_0^1 (dB) B'$, *where* $B(.)$ *is a* k*-variate standard Brownian motion.*

2.4 The dependent case

The independence assumption is not essential for the above results, provided that some adjustments are made:

Assumption 1. *Let*

$$y_t = y_{t-1} + x_t, \tag{7}$$

where x_t *is a* k*-variate zero mean covariance stationary Gaussian vector time series process with Wold decomposition* $x_t = \sum_{m=0}^{\infty} C_m u_{t-m}$, *with* u_t *i.i.d.* $N_k(0, I_k)$. *For* $m \to \infty$ *the elements* $c_{i,j,m}$ *of the* $k \times k$ *matrices* C_m *converge to zero at an exponential rate: there exists a* $\rho \in (0, 1)$ *and a constant* $K \in (0, \infty)$ *such that* $\max_{1 \leq i \leq k, 1 \leq j \leq k} |c_{i,j,m}| < K\rho^m$.

Denote $C(L) = \sum_{m=0}^{\infty} C_m L^m$, with L the lag operator, so that

$$x_t = C(L)u_t. \tag{8}$$

We can always write

$$C(L) = C(1) + \left(\frac{C(L) - C(1)}{1 - L} \right) (1 - L) = C(1) + D(L)\,(1 - L)\,, \tag{9}$$

say, because all the elements of $C(L)-C(1)$ have root 1, hence these elements are proportional to $1-L$. This construction is known as the Beveridge-Nelson decomposition. Then,

$$x_t = C(1)u_t + D(L)\,(1 - L)\,u_t = C(1)u_t + v_t - v_{t-1}, \tag{10}$$

and

$$y_t = \sum_{j=1}^{t} x_j + y_0 = C(1) \sum_{j=1}^{t} u_j + v_t - v_0 + y_0, \tag{11}$$

where

$$v_t = D(L)u_t = \sum_{j=0}^{\infty} D_m u_{t-m}. \tag{12}$$

Assumption 1 implies that v_t is a zero-mean covariance stationary Gaussian process, and that for $m \to \infty$, $D_m \to O_{k,k}$ exponentially.

Let us derive first the limiting distribution of $\frac{1}{T}\sum_{t=1}^{T} u_t y'_{t-1}$. Using the decomposition (11), we have

$$\frac{1}{T}\sum_{t=1}^{T} u_t y'_{t-1} = \frac{1}{T}\sum_{t=1}^{T} u_t \sum_{j=1}^{t-1} u_j C(1)' + \frac{1}{T}\sum_{t=1}^{T} u_t v'_{t-1} + \frac{1}{T}\sum_{t=1}^{T} u_t \left(y_0 - v_0\right)'. \tag{13}$$

Since u_t and v_{t-1} are independent, it follows that $\frac{1}{T}\sum_{t=1}^{T} u_t v'_{t-1} = O_p\left(1/\sqrt{T}\right)$ and $\frac{1}{T}\sum_{t=1}^{T} u_t = O_p\left(1/\sqrt{T}\right)$. It follows therefore from Theorem 1 and (13) that

Theorem 2. *Under Assumption 1,* $\frac{1}{T}\sum_{t=1}^{T} u_t y'_{t-1} \Rightarrow \int_0^1 (dB)\, B'C(1)'$.

Next, consider the case $\frac{1}{T}\sum_{t=1}^{T} x_t y'_{t-1}$. Note that by Assumption 1,

$$\Sigma_{XX'_m} = E[x_t x'_{t+m}] = \begin{cases} \sum_{j=0}^{\infty} C_j C'_{j+m} & if \quad m \geq 0 \\ \sum_{j=0}^{\infty} C_{j-m} C'_j & if \quad m < 0 \end{cases} \tag{14}$$

is finite, and that

Lemma 1. *For* $|m| \to \infty$, $\Sigma_{XX'_{-m}} \to O_{k,k}$ *exponentially, and for fixed* m, $p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x'_{t+m} = \Sigma_{XX'_m}$.

The latter result is not hard to prove. See for example Bierens (2004, Ch. 7).

Since $\Sigma_{XX'_{-m}} \to O_{k,k}$ exponentially it follows that

$$\begin{aligned} \Sigma_{XY'} &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[x_t \left(y_{t-1} - y_0\right)'] \\ &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1} E[x_t x'_{t-j}] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t-1} \Sigma'_{XX'_j} \end{aligned} \tag{15}$$

$$= \sum_{j=1}^{\infty} \Sigma'_{XX'_j}$$

exists and is finite. This matrix will play a role in the generalization of Theorem 2.

Similar to Lemma 1 it follows that

$$p \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} v_t v'_{t+m} = \Sigma_{VV'_m} \quad (16)$$

$$p \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} u_t v'_{t+m} = \Sigma_{UV'_m} \quad (17)$$

where

$$\Sigma_{VV'_m} = E[v_t v'_{t+m}] = \begin{cases} \sum_{j=0}^{\infty} D_j D'_{j+m} & if \quad m \geq 0 \\ \sum_{j=0}^{\infty} D_{j-m} D'_j & if \quad m < 0 \end{cases}$$

$$\Sigma_{UV'_m} = E[u_t v'_{t+m}] = \begin{cases} D'_m & if \quad m \geq 0 \\ O_{k,k} & if \quad m < 0 \end{cases}$$

We can now write

$$\frac{1}{T}\sum_{t=1}^{T} x_t y'_{t-1} = \frac{1}{T}\sum_{t=1}^{T} x_t \left(\sum_{j=1}^{t-1} u'_j C(1)' + v_{t-1}\right) + \frac{1}{T}\sum_{t=1}^{T} x_t \left(y_0 - v_0\right)'$$

$$= \frac{1}{T}\sum_{t=1}^{T} \left(C(1)u_t + v_t - v_{t-1}\right)\left(\sum_{j=1}^{t-1} u'_j C(1)' + v_{t-1}\right) + o_p(1)$$

$$= C(1)\left(\frac{1}{T}\sum_{t=1}^{T} u_t \sum_{j=1}^{t-1} u'_j\right) C(1)' + \frac{1}{T}\sum_{t=1}^{T} (v_t - v_{t-1}) \sum_{j=1}^{t-1} u'_j C(1)'$$

$$+ \frac{1}{T}\sum_{t=1}^{T} v_t v'_{t-1} - \frac{1}{T}\sum_{t=1}^{T} v_{t-1} v'_{t-1} + C(1)\frac{1}{T}\sum_{t=1}^{T} u_t v'_{t-1} + o_p(1)$$

$$= C(1)\left(\frac{1}{T}\sum_{t=1}^{T} u_t \sum_{j=1}^{t-1} u'_j\right) C(1)' + \Sigma_{VV'_{-1}} - \Sigma_{VV'_0} - \Sigma'_{UV'_0} + o_p(1) \quad (18)$$

The last $o_p(1)$ term follows from (16), (17) and

$$\frac{1}{T}\sum_{t=1}^{T} (v_t - v_{t-1}) \sum_{j=1}^{t-1} u'_j$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} v_t \sum_{j=1}^{t-1} u_j' - \frac{1}{T}\sum_{t=1}^{T} v_{t-1} \sum_{j=1}^{t-2} u_j' - \frac{1}{T}\sum_{t=1}^{T} v_{t-1}u_{t-1}' \\
&= v_T \frac{1}{T}\sum_{j=1}^{T-1} u_j' - \frac{1}{T}\sum_{t=1}^{T} v_{t-1}u_{t-1}' = -\frac{1}{T}\sum_{t=1}^{T} v_{t-1}u_{t-1}' + O_p\left(1/\sqrt{T}\right) \\
&= -\Sigma_{UV_0'}' + o_p(1)
\end{aligned}
$$

Moreover, it follows similar to (18) that

$$
\Sigma_{XY'} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left[x_t\left(y_{t-1}-y_0\right)'\right] = \Sigma_{VV_{-1}'} - \Sigma_{VV_0'} - \Sigma_{UV_0'}'.
$$

Thus by Theorem 1,

$$
\frac{1}{T}\sum_{t=1}^{T} x_t y_{t-1}' \Rightarrow C(1)\left(\int_0^1 (dB)\,B'\right)C(1)' + \Sigma_{XY'} \tag{19}
$$

Note that

$$
C(1)B(.) \sim \left(C(1)C(1)'\right)^{1/2} B_*(.),
$$

where B_* is also a k-variate standard Brownian motion. Moreover, the matrix $C(1)C(1)'$ is known as the long-run variance matrix of x_t:

$$
\Sigma = C(1)C(1)' = \lim_{T\to\infty} E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t\right)'\right]. \tag{20}
$$

Thus, (19) also reads as

$$
\frac{1}{T}\sum_{t=1}^{T} x_t y_{t-1}' \Rightarrow \Sigma^{1/2}\left(\int_0^1 (dB_*)\,B_*'\right)\Sigma^{1/2} + \Sigma_{XY'} \tag{21}
$$

2.5 A further generalization

In cointegration analysis we will encounter stochastic matrices of the type $\frac{1}{T}\sum_{t=1}^{T} x_{t-j}y_{t-1}'$, where $j \geq 0$. We have already considered the case $j = 0$, so let us focus on the case $j \geq 1$.

The limit distribution involved can easily be derived from the equality

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} x_{t-j}y'_{t-1} &= \frac{1}{T}\sum_{t=1-j}^{T-j} x_t y'_{t-1+j} \\
&= \frac{1}{T}\sum_{t=1}^{T} x_t y'_{t-1+j} + \frac{1}{T}\sum_{t=1-j}^{0} x_t y'_{t-1+j} - \frac{1}{T}\sum_{t=T-j+1}^{T} x_t y'_{t-1+j} \\
&= \frac{1}{T}\sum_{t=1}^{T} x_t y'_{t-1} + \frac{1}{T}\sum_{t=1}^{T} x_t \left(y_{t-1+j} - y_{t-1}\right)' + O_p(1/T) \\
&\quad + O_p(1/\sqrt{T}) \\
&= \frac{1}{T}\sum_{t=1}^{T} x_t y'_{t-1} + \sum_{i=0}^{j-1}\frac{1}{T}\sum_{t=1}^{T} x_t x'_{t+i} + o_p(1) \qquad (22)
\end{aligned}$$

where the O_p terms follow from

$$\begin{aligned}
\sum_{t=1-j}^{0} x_t y'_{t-1+j} &= O_p(1), \\
\left\|\frac{1}{T}\sum_{t=T-j+1}^{T} x_t y'_{t-1+j}\right\| &\leq \frac{1}{T}\sum_{t=T-j+1}^{T} \|x_t\| \, . \, \|y_{t-1+j}\| \\
&\leq \max_{0\leq x\leq 1}\left\|y_{[xT]}/\sqrt{T}\right\| \frac{1}{\sqrt{T}}\sum_{t=T-j+1}^{T} \|x_t\|
\end{aligned}$$

where for a matrix $||.||$ denotes the maximum of the absolute values of its elements, and the fact $y_{[xT]}/\sqrt{T} \Rightarrow C(1)B(x)$ implies that

$$\max_{0\leq x\leq 1}\left\|y_{[xT]}/\sqrt{T}\right\| \Rightarrow \max_{0\leq x\leq 1}\|B(x)\| = O_p(1).$$

Moreover, it follows from Lemma 1 that

$$p\lim_{T\to\infty}\sum_{i=0}^{j-1}\frac{1}{T}\sum_{t=1}^{T} x_t x'_{t+i} = \sum_{i=0}^{j-1}\Sigma_{XX'_i}$$

Finally, note that

$$\Sigma_{XY'} + \sum_{i=0}^{j-1}\Sigma_{XX'_i} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E\left[x_{t-j}\left(y_{t-1} - y_0\right)'\right],$$

hence the general result is:

Theorem 3. *Under Assumption 1 and for $m \geq 0$,*

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} x_{t-m} y_{t-1}' &\Rightarrow C(1)\left(\int_0^1 (dB)\, B'\right) C(1)' + \Sigma_{XY'} + \sum_{i=0}^{m-1} \Sigma_{XX_i'} \quad (23) \\ &\sim \Sigma^{1/2}\left(\int_0^1 (dB_*)\, B_*'\right) \Sigma^{1/2} + \Sigma_{XY'} + \sum_{i=0}^{m-1} \Sigma_{XX_i'} \end{aligned}$$

where B and B_* are k-variate standard Brownian motions and

$$\begin{aligned} \Sigma_{XX_i'} &= E[x_t x_{t+i}'], \\ \Sigma_{XY'} &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left[x_t\,(y_{t-1} - y_0)'\right] = \sum_{j=1}^{\infty} \Sigma_{XX_j'}', \\ \Sigma &= \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T} x_t\right]. \end{aligned}$$

3 Cointegration

3.1 Vector error correction model (VECM) representation

If the time series process y_t defined by (7) in Assumption 1 is cointegrated, there exist a $k \times r$ matrix β with rank $r < k$ such that

$$\beta' C(1) = O. \tag{24}$$

Note that (24) implies that the matrix $C(1)$ is singular, with rank $k - r$. Then it follows from (11) that

$$\beta' y_t = \beta' v_t + \beta' y_0 - \beta' v_0. \tag{25}$$

Engle and Granger (1987) have set forth conditions such that then y_t can be written as an vector error correction model of order p, shortly VECM(p):

$$\Delta y_t = \pi_0 + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Pi_j \Delta y_{t-j} + e_t, \tag{26}$$

where $e_t \sim$ i.i.d. $N_k[0,\Omega]$, Δ is the difference operator $1-L$, and the lag polynomial matrix $\Pi(L) = I_k - \sum_{j=1}^{p-1} \Pi_j L^j$ is invertible: $\det(\Pi(z)) = 0$ implies $|z| > 1$. Note that (26) is actually a VAR(p) model in y_t, but with unit roots in the determinant of the VAR lag polynomial involved.

The error term e_t in (26) is of course related to u_t in (8). To see how, substitute $\Delta y_t = C(L)u_t$ and

$$\begin{aligned} y_t &= C(1)\sum_{j=1}^{t} u_j + D(L)u_t - v_0 + y_0 \\ &= C(1)\left(\sum_{j=1}^{t} L^{j-1}\right) u_t + D(L)u_t - v_0 + y_0 \end{aligned}$$

in (26). Then

$$\begin{aligned} C(L)u_t &= C_0 u_t + \sum_{j=1}^{\infty} C_j u_{t-1} = C_0 u_t + (C(L) - C_0)\, u_t \\ &= \pi_0 - \alpha\beta' (v_0 - y_0) + \alpha\beta' C(1) L \left(\sum_{j=1}^{t} L^{j-1}\right) u_t \\ &\quad + \alpha\beta' L.D(L)u_t + \sum_{j=1}^{p-1} \Pi_j L^j C(L)u_t + e_t, \end{aligned}$$

Since the only terms in this equation that relate to time t are e_t and the leading term $C_0 u_t$ of $C(L)u_t$, we must have

$$e_t = C_0 u_t, \tag{27}$$

hence

$$\Omega = C_0 C_0' \tag{28}$$

Thus, with $x_t = \Delta y_t$ we can write the VECM(p) as

$$x_t = \pi_0 + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Pi_j x_{t-j} + C_0 u_t. \tag{29}$$

3.2 Granger's representation theorem

To see how (29) is related to (8), let ϕ be a $k \times (k-r)$ matrix with rank $k-r$ such that the matrix polynomial

$$P(L) = \begin{pmatrix} \beta' D(L) \\ \phi' C(L) \end{pmatrix}$$

is invertible, and the matrix

$$\Phi = \begin{pmatrix} \beta' \\ \phi' \end{pmatrix}$$

is nonsingular. Then

$$\begin{aligned} \Phi x_t &= \begin{pmatrix} \beta' \left(C(1) + (1-L)D(L)\right) \\ \phi' C(L) \end{pmatrix} u_t = \begin{pmatrix} ((1-L)\beta' D(L)) \\ \phi' C(L) \end{pmatrix} u_t \\ &= \begin{pmatrix} (1-L)I_r & O \\ O & I_{k-r} \end{pmatrix} P(L) u_t \end{aligned}$$

hence

$$P(L)^{-1} \begin{pmatrix} I_r & O \\ O & (1-L)I_{k-r} \end{pmatrix} \Phi \Delta y_t = \Delta u_t.$$

Applying the lag operator $1 + \sum_{j=1}^{t} L^j$ to both sided of this equation yields

$$P(L)^{-1} \begin{pmatrix} I_r & O \\ O & (1-L)I_{k-r} \end{pmatrix} \Phi \left(y_t - y_0\right) = u_t - u_0,$$

hence

$$\begin{aligned} & P(L)^{-1} \begin{pmatrix} I_r & O \\ O & (1-L)I_{k-r} \end{pmatrix} \Phi y_t \\ & = u_t + P(L)^{-1} \begin{pmatrix} I_r & O \\ O & (1-L)I_{k-r} \end{pmatrix} \Phi y_0 - u_0 \\ & = u_t + P(1)^{-1} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \Phi y_0 - u_0. \end{aligned}$$

Thus, denoting

$$\begin{aligned} A(L) &= P(L)^{-1} \begin{pmatrix} I_r & O \\ O & (1-L)I_{k-r} \end{pmatrix} \Phi, \\ \mu &= P(1)^{-1} \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \Phi y_0 - u_0 = P(1)^{-1} \begin{pmatrix} \beta' \\ O \end{pmatrix} y_0 - u_0. \end{aligned}$$

we have

$$A(L)y_t = \mu + u_t.$$

This is the VAR representation of y_t. Note however that $A(L)$ is not invertible.

Similar to (9) we van write

$$\begin{aligned} A(L) &= A(1)L + (1-L)\,\frac{A(L) - A(1)L}{1-L} \\ &= A(1)L + (1-L)\,\Psi(L), \end{aligned}$$

say, where

$$\Psi(L) = \frac{A(L) - A(1)L}{1-L}.$$

Moreover, note that

$$\begin{aligned} A(1) &= P(1)^{-1}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}\Phi = P(1)^{-1}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}\begin{pmatrix} \beta' \\ \phi' \end{pmatrix} \\ &= P(1)^{-1}\begin{pmatrix} \beta' \\ O \end{pmatrix}. \end{aligned}$$

Thus

$$\begin{aligned} A(L)y_t &= A(1)y_{t-1} + \Psi(L)\Delta y_t \\ &= P(1)^{-1}\begin{pmatrix} \beta' \\ O \end{pmatrix} y_{t-1} + \Psi(L)\Delta y_t = \mu + u_t \end{aligned} \tag{30}$$

Multiplying (30) by $\Phi^{-1}P(0)$ and denoting

$$\Pi(L) = \Phi^{-1}P(0)\Psi(L)$$

it follows that

$$\begin{aligned} \Pi(L)\Delta y_t &= \Phi^{-1}P(0)\Psi(L)\Delta y_t = -\Phi^{-1}P(0)P(1)^{-1}\begin{pmatrix} \beta' \\ O \end{pmatrix} y_{t-1} \\ &\quad + \Phi^{-1}P(0)\mu + \Phi^{-1}P(0)u_t. \end{aligned} \tag{31}$$

Note that

$$\Phi^{-1}P(0) = \Phi^{-1}\begin{pmatrix} \beta' D_0 \\ \phi' C_0 \end{pmatrix} = \Phi^{-1}\Phi C(0) = C_0 \tag{32}$$

and

$$\Phi^{-1}P(0)\mu = C_0 P(1)^{-1}\begin{pmatrix}\beta' \\ O\end{pmatrix} y_0 - C_0 u_0$$

Moreover, observe that the leading term in $\Pi(L)$ is

$$\begin{aligned}\Pi(0) &= \Phi^{-1}P(0)\Psi(0) = \Phi^{-1}P(0)A(0) \\ &= \Phi^{-1}P(0)P(0)^{-1}\Phi = I_k.\end{aligned}$$

Finally, assume that

$$\Pi(L) = I_k - \sum_{j=1}^{p-1}\Pi_j L^j,$$

denote

$$\begin{aligned}(\alpha, \alpha_*) &= -\Phi^{-1}P(0)P(1)^{-1} = -C_0 P(1)^{-1} \\ &= -C_0\begin{pmatrix}\beta' D(1) \\ \phi' C(1)\end{pmatrix}^{-1} \qquad (33)\end{aligned}$$

where α is a $k \times r$ matrix and let

$$\begin{aligned}\pi_0 &= \Phi^{-1}P(0)\mu = \Phi^{-1}P(0)P(1)^{-1}\begin{pmatrix}\beta' \\ O\end{pmatrix} y_0 - \Phi^{-1}P(0)u_0 \qquad (34) \\ &= -\alpha\beta' y_0 - C_0 u_0.\end{aligned}$$

Then the VECM(p) model (29) follows.

An interesting special case is where π_0 takes the form

$$\pi_0 = -\alpha\phi \text{ for a vector } \phi \in \mathbb{R}^r, \qquad (35)$$

because then the (29) becomes

$$x_t = \alpha\beta'\left(y_{t-1} - \phi\right) + \sum_{j=1}^{p-1}\Pi_j x_{t-j} + C_0 u_t. \qquad (36)$$

This implies that $\beta'\left(y_{t-1} - \phi\right)$ is zero-mean stationary, because $E[x_t] = 0$. Hence, $\beta' y_{t-1}$ is stationary about a "constant" vector $\beta'\phi$. The case (35) is known as "cointegrating restrictions on the intercept parameters." If so, it follows from (34) that $\phi = \beta' y_0 + (\alpha'\alpha)^{-1} \alpha' C_0 u_0$, which is actually a random vector.

A by-product of the above argument is:

Lemma 2. *Let $\alpha_\perp$ be an orthogonal complement*[2] *of α in VECM(p) model (29). There exists an orthogonal complement $\beta_\perp$ of β such that $\beta_\perp' C(1) = (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \alpha_\perp' C_0$.*

Proof: Observe from (33) that

$$(\alpha, \alpha_*) \begin{pmatrix} \beta' D(1) \\ \phi' C(1) \end{pmatrix} = \alpha \beta' D(1) + \alpha_* \phi' C(1) = -C_0,$$

hence $\alpha_\perp' \alpha_* \phi' C(1) = \alpha_\perp' \alpha_* \phi' \beta_\perp \delta = -\alpha_\perp' C_0$. Thus, if we choose

$$\gamma' = -(\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \alpha_\perp' \alpha_* \phi'$$

then

$$\gamma' C(1) = (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \alpha_\perp' C_0.$$

Next, observe that for any orthogonal complement $\overline{\beta}_\perp$ of β, $\overline{\beta}_\perp \left(\overline{\beta}_\perp' \overline{\beta}_\perp\right)^{-1} \overline{\beta}_\perp' + \beta (\beta'\beta)^{-1} \beta' = I_k$,[3] so that

$$\gamma = \overline{\beta}_\perp \left(\overline{\beta}_\perp' \overline{\beta}_\perp\right)^{-1} \overline{\beta}_\perp' \gamma + \beta (\beta'\beta)^{-1} \beta' \gamma.$$

Then $\left(\gamma' \overline{\beta}_\perp\right) \left(\overline{\beta}_\perp' \overline{\beta}_\perp\right)^{-1} \overline{\beta}_\perp' C(1) = (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \alpha_\perp' C_0$. Taking

$$\beta_\perp = \overline{\beta}_\perp \left(\overline{\beta}_\perp' \overline{\beta}_\perp\right)^{-1} \overline{\beta}_\perp' \gamma,$$

Lemma 2 follows. Q.E.D.

This result will play a key-role in the next sections.

[2]Given a $k \times r$ matrix ξ, where $1 \leq r < k$, an orthogonal complement $\xi_\perp$ of ξ is $k \times (k-r)$ matrix with rank $k-r$ such that $\xi_\perp' \xi = O_{k-r,r}$. Note that $\xi_\perp$ is not unique, because for any nonsingular $(k-r) \times (k-r)$ matrix R_{k-r}, $\xi_\perp R_{k-r}$ is also an orthogonal complement of ξ.

[3]Because the left-hand side matrix is idempotent, with rank k.

4 Likelihood-based cointegration analysis

4.1 The VECM(p) case without intercepts

To demonstrate the Johansen (1988) approach, assume in first instance that $u_t = 0$ for $t < 1$, so that $y_0 = 0$ and thus by (34), $\pi_0 = 0$, so that

$$x_t = \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Pi_j x_{t-j} + C_0 u_t. \tag{37}$$

where $x_t = \Delta y_t$. Next, denote

$$X_{t-1} = \left(x'_{t-1}, ..., x'_{t-p+1}\right)', \ \Pi = \left(\Pi_1, ..., \Pi_{p-1}\right), \tag{38}$$

Then (37) can be written as

$$x_t = \alpha\beta' y_{t-1} + \Pi X_{t-1} + C_0 u_t. \tag{39}$$

Note that, given β, the identification of α and Π requires that

Assumption 2. $\text{Var}\left[\left(y'_{t-1}\beta, X'_{t-1}\right)'\right]$ *is nonsingular.*

The log-likelihood involved takes the form

$$\begin{aligned} &\ln L_T(\alpha, \beta, \Pi, \Omega) \\ &= -\frac{1}{2}\sum_{t=1}^{T+1} \left(x_t - \alpha\beta' y_{t-1} - \Pi X_{t-1}\right)' \Omega^{-1} \left(x_t - \alpha\beta' y_{t-1} - \Pi X_{t-1}\right) \\ &-\frac{1}{2} T \ln\left(\det \Omega\right) - T.k.\ln\left(\sqrt{2\pi}\right), \end{aligned}$$

where Ω is defined by (28).

Given α, β, and Ω, the matrix Π can be concentrated out by regressing $x_t - \alpha\beta' y_{t-1}$ on X_{t-1}:

$$\begin{aligned} \widehat{\Pi}\left(\alpha, \beta\right) &= \frac{1}{T}\sum_{t=1}^{T} \left(x_t X'_{t-1} - \alpha\beta' y_{t-1} X'_{t-1}\right) \left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1} X'_{t-1}\right)^{-1} \\ &= \left(\frac{1}{T}\sum_{t=1}^{T} x_t X'_{t-1}\right) \left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1} X'_{t-1}\right)^{-1} \\ &\quad -\alpha\beta' \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1} X'_{t-1}\right) \left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1} X'_{t-1}\right)^{-1}. \end{aligned}$$

Hence, denoting

$$R_{0,t} = x_t - \left(\frac{1}{T}\sum_{j=1}^{T} x_j X_{j-1}'\right)\left(\frac{1}{T}\sum_{j=1}^{T} X_{j-1}X_{j-1}'\right)^{-1} X_{t-1}, \quad (40)$$

$$R_{1,t} = y_{t-1} - \left(\frac{1}{T}\sum_{j=1}^{T} y_{j-1} X_{j-1}'\right)\left(\frac{1}{T}\sum_{j=1}^{T} X_{j-1}X_{j-1}'\right)^{-1} X_{t-1}, \quad (41)$$

the concentrated log-likelihood becomes

$$\ln L_T(\alpha,\beta,\Omega) = \max_{\Pi} \ln L_T(\alpha,\beta,\Pi,\Omega) \quad (42)$$

$$= -\frac{1}{2}\sum_{t=1}^{T} \left(R_{0,t} - \alpha\beta' R_{1,t}\right)' \Omega^{-1} \left(R_{0,t} - \alpha\beta' R_{1,t}\right)$$

$$-\frac{1}{2}T.\ln\left(\det\Omega\right) - T.k\ln\left(\sqrt{2\pi}\right).$$

Next, given β and Ω, α can be concentrated out by replacing $R_{0,t}-\alpha\beta' R_{1,t}$ with the OLS residual of the regression of $R_{0,t}$ on $\beta' R_{0,t}$, which yields

$$\widehat{\alpha}\left(\beta\right) = \widehat{S}_{0,1}\beta\left(\beta'\widehat{S}_{1,1}\beta\right)^{-1}$$

where

$$\widehat{S}_{0,1} = \frac{1}{T}\sum_{t=1}^{T} R_{0,t}R_{1,t}' = \frac{1}{T}\sum_{t=1}^{T} x_t y_{t-1}' \quad (43)$$

$$-\left(\frac{1}{T}\sum_{t=1}^{T} x_t X_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}y_{t-1}'\right),$$

$$\widehat{S}_{1,1} = \frac{1}{T}\sum_{t=1}^{T} R_{1,t}R_{1,t}' = \frac{1}{T}\sum_{t=1}^{T} y_{t-1} y_{t-1}' \quad (44)$$

$$-\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1} X_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}y_{t-1}'\right).$$

Hence

$$\ln L_T(\beta,\Omega) = \max_{\Pi,\alpha} \ln L_T(\alpha,\beta,\Pi,\Omega)$$

$$
\begin{aligned}
&= -\frac{1}{2}\sum_{t=1}^{T}\left(R_{0,t}-\widehat{\alpha}\left(\beta\right)\beta' R_{1,t}\right)'\Omega^{-1}\left(R_{0,t}-\widehat{\alpha}\left(\beta\right)\beta' R_{1,t}\right) \\
&\quad -\frac{1}{2}T\ln\left(\det\Omega\right)-T.k\ln\left(\sqrt{2\pi}\right).
\end{aligned}
$$

As is well known the ML estimator of Ω given β now takes the form

$$
\widehat{\Omega}\left(\beta\right)=\frac{1}{T}\sum_{t=1}^{T}\left(R_{0,t}-\widehat{\alpha}\left(\beta\right)\beta' R_{1,t}\right)\left(R_{0,t}-\widehat{\alpha}\left(\beta\right)\beta' R_{1,t}\right)'
$$

which can be further elaborated as follows:

$$
\begin{aligned}
\widehat{\Omega}\left(\beta\right) &= \frac{1}{T}\sum_{t=1}^{T}R_{0,t}R_{0,t}'-\widehat{\alpha}\left(\beta\right)\beta'\frac{1}{T}\sum_{t=1}^{T}R_{1,t}R_{1,t}'\beta\widehat{\alpha}\left(\beta\right)' \\
&= \widehat{S}_{0,0}-\widehat{\alpha}\left(\beta\right)\left(\beta'\widehat{S}_{1,1}\beta\right)\widehat{\alpha}\left(\beta\right)' \\
&= \widehat{S}_{0,0}-\widehat{S}_{0,1}\beta\left(\beta'\widehat{S}_{1,1}\beta\right)^{-1}\beta'\widehat{S}_{1,0}
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{S}_{0,0} &= \frac{1}{T}\sum_{t=1}^{T}R_{0,t}R_{0,t}'=\frac{1}{T}\sum_{t=1}^{T}x_t x_t' \qquad (45)\\
&\quad -\left(\frac{1}{T}\sum_{t=1}^{T}x_t X_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}x_t'\right), \\
\widehat{S}_{1,0} &= \widehat{S}_{0,1}' \qquad (46)
\end{aligned}
$$

Then

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left(R_{0,t}-\widehat{\alpha}\left(\beta\right)\beta' R_{1,t}\right)'\left(\widehat{\Omega}\left(\beta\right)\right)^{-1}\left(R_{0,t}-\widehat{\alpha}\left(\beta\right)\beta' R_{12,t}\right) \\
&= \mathrm{trace}\left(\left(\widehat{\Omega}\left(\beta\right)\right)^{-1}\widehat{\Omega}\left(\beta\right)\right)=\mathrm{trace}\left(I_k\right)=k,
\end{aligned}
$$

hence

$$
\begin{aligned}
&\ln L_T(\beta)=\max_{\Pi,\alpha,\Omega}\ln L_T(\alpha,\beta,\Pi,\Omega) \\
&= -\frac{1}{2}T.\ln\left(\det\left(\widehat{S}_{0,0}-\widehat{S}_{0,1}\beta\left(\beta'\widehat{S}_{1,1}\beta\right)^{-1}\beta'\widehat{S}_{1,0}\right)\right) \\
&-T.k\ln\left(\sqrt{2\pi}\right)-kT.
\end{aligned}
$$

Thus, the maximum likelihood estimator $\widehat{\beta}$ of β can be obtained by minimizing

$$\det\left(\widehat{S}_{0,0}-\widehat{S}_{0,1}\beta\left(\beta'\widehat{S}_{1,1}\beta\right)^{-1}\beta'\widehat{S}_{1,0}\right)$$

to β.

To simplify this problem, consider the easy matrix equalities

$$\begin{aligned}\begin{pmatrix} A & B \\ B' & C \end{pmatrix} &= \begin{pmatrix} A & O \\ B' & I \end{pmatrix}\begin{pmatrix} I & A^{-1}B \\ O & C-B'A^{-1}B \end{pmatrix} \\ &= \begin{pmatrix} I & B \\ O & C \end{pmatrix}\begin{pmatrix} A-BC^{-1}B' & O \\ C^{-1}B' & I \end{pmatrix},\end{aligned}$$

where A and C are nonsingular square matrices. These equalities imply that

$$\det(A)\det\left(C-B'A^{-1}B\right)=\det\left(C\right)\det\left(A-BC^{-1}B'\right).$$

Taking $A=\beta'\widehat{S}_{1,1}\beta$, $B=\beta'\widehat{S}_{1,0}$, $C=\widehat{S}_{0,0}$ it follows now that

$$\begin{aligned}&\det\left(\widehat{S}_{0,0}-\widehat{S}_{0,1}\beta\left(\beta'\widehat{S}_{1,1}\beta\right)^{-1}\beta'\widehat{S}_{1,0}\right) \\ &=\frac{\det\left(\widehat{S}_{0,0}\right)\det\left(\beta'\widehat{S}_{1,1}\beta-\beta'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\beta\right)}{\det\left(\beta'\widehat{S}_{1,1}\beta\right)}.\end{aligned} \tag{47}$$

Thus, the ML estimator $\widehat{\beta}$ of β minimizes (47). However, if $\widehat{\beta}$ is a solution then so is $c\widehat{\beta}$ for any $c\neq 0$ in the case $r=1$, and in the case $2<r<k$, we may replace $\widehat{\beta}$ by $\widehat{\beta}C_{r,r}$ with $C_{r,r}$ an arbitrary nonsingular $r\times r$ matrix. Thus, we need to normalize $\widehat{\beta}$ somehow. How to normalize $\widehat{\beta}$ will be explained below.

Let $\widehat{\lambda}_1\geq\widehat{\lambda}_2\geq...\geq\widehat{\lambda}_k$ be the ordered solutions of the generalized eigenvalue problem

$$\det\left(\lambda\widehat{S}_{1,1}-\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\right)=0, \tag{48}$$

and let $\widehat{q}_1,\widehat{q}_2,...,\widehat{q}_k$ be the corresponding generalized eigenvectors. Then for $j=1,..,k$,

$$\widehat{S}_{1,1}\widehat{q}_j\widehat{\lambda}_j=\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\widehat{q}_j.$$

Hence, denoting

$$\widehat{\Lambda}=\mathrm{diag}\left(\widehat{\lambda}_1,\widehat{\lambda}_2,...,\widehat{\lambda}_k\right),\ \widehat{Q}=\left(\widehat{q}_1,\widehat{q}_2,...,\widehat{q}_k\right),$$

we have

$$\widehat{S}_{1,1}\widehat{Q}\widehat{\Lambda} = \widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\widehat{Q}. \tag{49}$$

The eigenvalues $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq ... \geq \widehat{\lambda}_k$ can be obtained by solving the standard eigenvalue problem

$$\det\left(\lambda I_k - \widehat{S}_{1,1}^{-1/2}\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\widehat{S}_{1,1}^{-1/2}\right) = 0, \tag{50}$$

with corresponding orthonormal eigenvector $\widehat{q}_1^*, \widehat{q}_2^*, ..., \widehat{q}_k^*$. Thus, denoting $\widehat{Q}^* = (\widehat{q}_1^*, \widehat{q}_2^*, ..., \widehat{q}_k^*)$, we have

$$\widehat{Q}^*\widehat{\Lambda} = \widehat{S}_{1,1}^{-1/2}\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\widehat{S}_{1,1}^{-1/2}\widehat{Q}^*. \tag{51}$$

Comparing (50) and (51) we see that we may choose $\widehat{Q} = \widehat{S}_{1,1}^{-1/2}\widehat{Q}^*$, so that by the orthogonality of $\widehat{Q}^*$,

$$\widehat{Q}'\widehat{S}_{1,1}\widehat{Q} = I_k.$$

Next, let

$$\widehat{\beta} = \widehat{Q}\xi,$$

where ξ is normalized such that

$$\xi'\xi = I_r.$$

Because $\widehat{\beta} = \widehat{Q}\xi = \widehat{S}_{1,1}^{-1/2}\widehat{Q}^*\xi$ and $\widehat{Q}^*$ is orthogonal, this normalization implies that

$$\widehat{\beta}'\widehat{S}_{1,1}\widehat{\beta} = I_r.$$

Then

$$\begin{aligned}\frac{\det\left(\beta'\widehat{S}_{1,1}\beta - \beta'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\beta\right)}{\det\left(\beta'\widehat{S}_{1,1}\beta\right)} &= \det\left(I_r - \xi'\widehat{Q}'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\widehat{Q}\xi\right) \\ &= \det\left(I_r - \xi'\widehat{\Lambda}\xi\right).\end{aligned}$$

In the case $r = 1$, $\xi = (\xi_1, ..., \xi_k)' \in \mathbb{R}^k$ and $\det\left(I_r - \xi'\widehat{\Lambda}\xi\right) = 1 - \xi'\widehat{\Lambda}\xi$ $= 1 - \sum_{j=1}^k \widehat{\lambda}_k \xi_j^2$, which is minimal subject to $\sum_{j=1}^k \xi_j^2 = 1$ for $\xi = (1, 0, ..., 0)'$. In the case $r > 1$ it is not hard to shown that the solution is

$$\xi = \begin{pmatrix} I_r \\ O \end{pmatrix}.$$

Consequently,

$$\widehat{\beta} = (\widehat{q}_1, \widehat{q}_2, ..., \widehat{q}_r)$$

so that

$$\widehat{\lambda}_i \widehat{S}_{1,1} \widehat{q}_i = \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \widehat{q}_i, \ i = 1, 2, .., r, \tag{52}$$

and

$$\begin{aligned} \ln L_T(r) &= \max_{\Pi,\alpha,\beta,\Omega} \ln L_T(\alpha, \beta, \Pi, \Omega) \qquad (53) \\ &= -\frac{1}{2}T.\ln\left(\det\left(\widehat{S}_{0,0}\right)\right) - \frac{1}{2}T.\sum_{j=1}^{r} \ln\left(1 - \widehat{\lambda}_j\right) - T.k\ln\left(\sqrt{2\pi}\right) - kT. \end{aligned}$$

4.2 Tests for the cointegrating rank

Consequently, the likelihood-ratio (LR) test of the null-hypothesis that the cointegrating rank is $r < k$ against the alternative hypothesis that the cointegrating rank is $r + 1$ takes the form

$$\begin{aligned} LR_T(r|r+1) &= -2\left(\ln L_T(r) - L_T(r+1)\right) = -T.\ln\left(1 - \widehat{\lambda}_{r+1}\right) \qquad (54) \\ &= T.\widehat{\lambda}_{r+1} + o_p(1), \end{aligned}$$

where the approximation is due to the Taylor expansion of $\ln\left(1 - \widehat{\lambda}_{r+1}\right)$, provided that $T.\widehat{\lambda}_{r+1}$ converges is distribution under the null hypothesis. The latter will be shown below. Since $\widehat{\lambda}_{r+1}$ is the largest of the $k - r$ smallest solutions of the generalized eigenvalue problem (48), the LR test (54) is called by Johansen (1988) the *lambda-max test*.

Another test proposed by Johansen (1988) is the LR test

$$\begin{aligned} LR_T(r|k) &= -2\left(\ln L_T(r) - L_T(k)\right) = -T.\sum_{i=r+1}^{k} \ln\left(1 - \widehat{\lambda}_i\right) \qquad (55) \\ &= \sum_{i=r+1}^{k} T.\widehat{\lambda}_i + o_p(1), \end{aligned}$$

which has the same null hypothesis as before, but as alternative hypothesis that the cointegrating rank is k. This alternative hypothesis implies that β is a nonsingular $k \times k$ matrix, which in its turn implies that y_t is stationary, because if $\beta' y_t$ is stationary and β' is nonsingular then $y_t = (\beta')^{-1} \beta' y_t$ is stationary. This test is called by Johansen (1988) the *trace test,* because its null distribution takes the form of the trace of a random matrix.

Although these test are designed on the basis of a specific alternative hypothesis, they have power against the more general alternative that the cointegrating rank is larger than r.

4.3 Limiting distributions

Because the log-likelihood (53) is a function of the solutions of the generalized eigenvalue problem (48), which in its turn depend on the matrices $\widehat{S}_{0,0}$, $\widehat{S}_{0,1}$ and $\widehat{S}_{1,1}$, we need to derive the limiting distributions or probability limits of these matrices.

First note that

Lemma 3. *Under Assumption 1 the probability limits*

$$\begin{aligned}
\Sigma_{\beta\beta} &= p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \beta' y_{t-1} y'_{t-1} \beta, \\
\Sigma_{\beta X} &= p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \beta' y_{t-1} X'_{t-1}, \ \Sigma_{X\beta} = \Sigma'_{\beta X}, \\
\Sigma_{XX} &= p \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} X_{t-1} X'_{t-1}
\end{aligned}$$

exist. Moreover, under Assumption 2, Σ_{XX} *is non-singular, and the matrix*

$$\Sigma^*_{\beta\beta} = \Sigma_{\beta\beta} - \Sigma_{\beta X} \Sigma^{-1}_{XX} \Sigma_{X\beta}$$

is nonsingular

Proof: Only the last part is not obvious. To show this part, observe that $\Sigma^*_{\beta\beta}$ is the variance matrix of the error in the projection of $\beta' y_{t-1}$ on X_{t-1}: $\beta' y_{t-1} = \Gamma X_{t-1} + \eta_t$, where $\Gamma = \Sigma_{\beta X} \Sigma^{-1}_{XX}$, and $\Sigma^*_{\beta\beta} = \mathrm{Var}(\eta_t)$. If $\Sigma^*_{\beta\beta}$ is singular then there exist vectors $\omega \in \mathbb{R}^r$, $\upsilon \in \mathbb{R}^k$ such that $\omega' \beta' y_{t-1} = \upsilon' \Gamma X_{t-1}$, which however violates Assumption 2. Q.E.D.

4.3.1 The matrix $\widehat{S}_{0,0}$

Replacing x_t in (45) by the right-hand side of (39) yields

$$\begin{aligned}
\widehat{S}_{0,0} &= \frac{1}{T}\sum_{t=1}^{T}\left(\alpha\beta' y_{t-1} + \Pi X_{t-1} + C_0 u_t\right)\left(y'_{t-1}\beta\alpha' + X'_{t-1}\Pi' + u'_t C'_0\right) \\
&\quad - \left(\alpha\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}X'_{t-1} + \Pi\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X'_{t-1} + C_0\frac{1}{T}\sum_{t=1}^{T}u_t X'_{t-1}\right) \\
&\quad \times \left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X'_{t-1}\right)^{-1} \\
&\quad \times \left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y'_{t-1}\beta\alpha' + \frac{1}{T}\sum_{t=1}^{T}X_{t-1}X'_{t-1}\Pi' + \frac{1}{T}\sum_{t=1}^{T}X_{t-1}u'_t C'_0\right) \\
&= \alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}y'_{t-1}\beta\right)\alpha' + C_0 C'_0 \\
&\quad -\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}X'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X'_{t-1}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y'_{t-1}\beta\right)\alpha' \\
&\quad + o_p(1) \\
&= \Omega + \alpha\Sigma^*_{\beta\beta}\alpha' + o_p(1),
\end{aligned}$$

where $\Omega = C_0 C'_0$, and $\Sigma^*_{\beta\beta}$ is defined in Lemma 3.

Next, choose φ such that

$$\alpha'_\perp \Omega \varphi = O. \tag{56}$$

Then

$$\alpha'_\perp \widehat{S}_{0,0}\alpha_\perp = \alpha'_\perp\Omega\alpha_\perp + o_p(1), \tag{57}$$

$$\alpha'_\perp \widehat{S}_{0,0}\varphi = \alpha'_\perp\Omega\varphi + o_p(1) = o_p(1), \tag{58}$$

$$\varphi'\widehat{S}_{0,0}\varphi = \varphi'\Omega\varphi + \varphi'\alpha\Sigma^*_{\beta\beta}\alpha'\varphi + o_p(1) \tag{59}$$

Hence it follows from (57), (58) and (59) that

$$\begin{aligned}
&p\lim_{T\to\infty}\left(\begin{pmatrix}\alpha'_\perp \\ \varphi'\end{pmatrix}\widehat{S}_{0,0}\left(\alpha_\perp,\varphi\right)\right)^{-1} = \left(\alpha_\perp,\varphi\right)^{-1}\left(p\lim_{T\to\infty}\widehat{S}_{0,0}^{-1}\right)\begin{pmatrix}\alpha'_\perp \\ \varphi'\end{pmatrix}^{-1} \\
&= \begin{pmatrix}\left(\alpha'_\perp\Omega\alpha_\perp\right)^{-1} & O \\ O & \Psi_0\end{pmatrix}
\end{aligned} \tag{60}$$

where

$$\Psi_0 = \left(\varphi'\Omega\varphi + \varphi'\alpha\Sigma^*_{\beta\beta}\alpha'\varphi\right)^{-1}$$

Finally, note that any $k \times r$ matrix φ with rank r satisfying (56) takes the form

$$\varphi = \Omega^{-1}\alpha.R, \tag{61}$$

where R is a non-singular $r \times r$ matrix, hence

$$\begin{aligned}\Psi_0 &= \left(R'\alpha'\Omega^{-1}\alpha.R + R'\alpha'\Omega^{-1}\alpha\Sigma^*_{\beta\beta}\alpha'\Omega^{-1}\alpha.R\right)^{-1} \\ &= \left(R'\alpha'\Omega^{-1}\alpha\left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma^*_{\beta\beta}\right)\alpha'\Omega^{-1}\alpha.R\right)^{-1}\end{aligned}$$

Summarizing, it has been shown that:

Lemma 4. *Under VECM* (37) *and Assumptions 1-2,*

$$\widehat{S}_{0,0} = \Omega + \alpha\Sigma^*_{\beta\beta}\alpha' + o_p(1),$$

and

$$\begin{aligned}&p\lim_{T\to\infty} \widehat{S}^{-1}_{0,0} = \alpha_\perp\left(\alpha'_\perp\Omega\alpha_\perp\right)^{-1}\alpha'_\perp \\ &+\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma^*_{\beta\beta}\right)^{-1}\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1}.\end{aligned}$$

where $\Sigma^*_{\beta\beta}$ is defined in Lemma 3.

4.3.2 The matrix $\widehat{S}_{0,1}$

Recall that

$$\begin{aligned}\widehat{S}_{0,1} &= \frac{1}{T}\sum_{t=1}^{T} x_t y'_{t-1} \\ &-\left(\frac{1}{T}\sum_{t=1}^{T} x_t X'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}y'_{t-1}\right).\end{aligned} \tag{62}$$

Replacing x_t in (62) by the right-hand side of (39) yields

$$\begin{aligned}
\widehat{S}_{0,1} &= \alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}y_{t-1}'\right) \\
&\quad -\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}X_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y_{t-1}'\right) \\
&\quad +C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t y_{t-1}'\right) \\
&\quad -C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t X_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y_{t-1}'\right).
\end{aligned}$$

Due to (38) and Theorem 3, there exists a $k(p-1)\times k$ random matrix M_* such that

$$\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y_{t-1}' \Rightarrow M_*, \tag{63}$$

hence $\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y_{t-1}' = O_p(1)$. Moreover, it is easy to verify that under Assumption 1,

$$p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}u_t X_{t-1}' = O. \tag{64}$$

Hence,

$$\begin{aligned}
\widehat{S}_{0,1} &= \alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}y_{t-1}'\right) - \alpha\Sigma_{\beta X}\Sigma_{XX}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}X_{t-1}y_{t-1}'\right) \\
&\quad +C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t y_{t-1}'\right) + o_p(1)
\end{aligned}$$

and therefore

$$\alpha_\perp'\widehat{S}_{0,1} = \alpha_\perp' C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t y_{t-1}'\right) + o_p(1), \tag{65}$$

$$\alpha_\perp'\widehat{S}_{0,1}\beta = \alpha_\perp' C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t y_{t-1}'\beta\right) + o_p(1) = o_p(1) \tag{66}$$

$$\widehat{S}_{0,1}\beta = \alpha\Sigma_{\beta\beta}^{*} + o_p(1) \tag{67}$$

where the latter follows from Lemma 3.

Let φ be given by (61) and let $\beta_\perp$ be the $k \times (k-r)$ matrix in Lemma 2. It follows now from (65), (66), (67), Theorem 2 and Lemma 2 that

$$\begin{pmatrix} \alpha'_\perp \\ \varphi' \end{pmatrix} \widehat{S}_{0,1}(\beta_\perp, \beta) = \begin{pmatrix} \alpha'_\perp \widehat{S}_{0,1}\beta_\perp & \alpha'_\perp \widehat{S}_{0,1}\beta \\ \varphi' \widehat{S}_{0,1}\beta_\perp & \varphi' \widehat{S}_{0,1}\beta \end{pmatrix} \tag{68}$$
$$\Rightarrow \begin{pmatrix} \alpha'_\perp C_0 \int_0^1 (dB) B'\beta_\perp & O_{k-r,r} \\ \varphi' \widetilde{S}_{0,1}\beta_\perp & \varphi'\alpha\left(\Sigma_{\beta\beta} - \Sigma_{\beta X}\Sigma_{XX}^{-1}\Sigma_{X\beta}\right) \end{pmatrix}$$

where $\widetilde{S}_{0,1}$ is a random matrix such that

$$\widehat{S}_{0,1} \Rightarrow \widetilde{S}_{0,1}.$$

The latter follows from Theorem 3.

To make the right-hand side matrix in (68) block-diagonal, post-multiply (68) by

$$\Theta = \begin{pmatrix} I_{k-r} & O_{k-r,r} \\ \Theta_{21} & I_r \end{pmatrix} \tag{69}$$

where

$$\begin{aligned} \Theta_{21} &= -\left(\varphi'\alpha\Sigma^*_{\beta\beta}\right)^{-1} \varphi' \widetilde{S}_{0,1}\beta_\perp \\ &= -\left(R'\alpha'\Omega^{-1}\alpha\Sigma^*_{\beta\beta}\right)^{-1} R'\alpha'\Omega^{-1}\alpha\widetilde{S}_{0,1}\beta_\perp \\ &= (\Sigma^*_{\beta\beta})^{-1}\widetilde{S}_{0,1}\beta_\perp. \end{aligned} \tag{70}$$

Note that the second equality follows from (61). Then

$$\begin{pmatrix} \alpha'_\perp \\ \varphi' \end{pmatrix} \widehat{S}_{0,1}(\beta_\perp, \beta)\Theta$$
$$\Rightarrow \begin{pmatrix} \alpha'_\perp C_0 \int_0^1 (dB) B'C(1)'\beta_\perp & O \\ O & \varphi'\alpha\Sigma^*_{\beta\beta} \end{pmatrix} \tag{71}$$

Combining (71) and (60) it follows now that

Lemma 5. *Under VECM* (37) *and Assumptions 1-2,*

$$\Theta' \begin{pmatrix} \beta'_\perp \\ \beta' \end{pmatrix} \widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}(\beta_\perp, \beta)\Theta \Rightarrow \begin{pmatrix} \int_0^1 B_{k-r} dB'_{k-r} \int_0^1 (dB_{k-r}) B'_{k-r} & O \\ O & \Psi_* \end{pmatrix} \tag{72}$$

where $\beta_{\perp}$ *is the* $k \times (k-r)$ *matrix in Lemma 2,* Θ *is defined by* (69) *and* (70),

$$B_{k-r} = \left(\alpha'_{\perp}\Omega\alpha_{\perp}\right)^{-1/2}\alpha'_{\perp}C_0 B \tag{73}$$

is a $k-r$ *variate standard Brownian motion, and*

$$\Psi_{*} = \Sigma^{*}_{\beta\beta}\left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma^{*}_{\beta\beta}\right)^{-1}\Sigma^{*}_{\beta\beta}, \tag{74}$$

with $\Sigma^{*}_{\beta\beta}$ *defined in Lemma 3. Moreover,* $\widehat{S}_{0,1}\beta = \alpha\Sigma^{*}_{\beta\beta} + o_p(1)$.[4]

4.3.3 The matrix $\widehat{S}_{1,1}$

Since by Theorem 3, $\frac{1}{T}\sum_{j=1}^{T} X_{t-1}y'_{t-1}$ converges in distribution and therefore is of order $O_p(1)$, it follows from (44) and Lemma 3 that

$$\begin{aligned}
\widehat{S}_{1,1} &= \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y'_{t-1} \qquad (75)\\
&\quad - \left(\frac{1}{T}\sum_{j=1}^{T} y_{t-1}X'_{t-1}\right)\Sigma^{-1}_{XX}\left(\frac{1}{T}\sum_{j=1}^{T} X_{t-1}y'_{t-1}\right) + o_p(1)\\
&= \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y'_{t-1} + O_p(1)
\end{aligned}$$

Moreover,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}y'_{t-1} \Rightarrow C(1)\left(\int_0^1 B(x)B(x)'dx\right)C(1)'$$

due to the functional central limit theorem, i.e.,

$$y_{[x.T]}/\sqrt{T} \Rightarrow C(1)B(x),$$

and the continuous mapping theorem. Hence

$$\frac{1}{T}\widehat{S}_{1,1} \Rightarrow C(1)\left(\int_0^1 B(x)B(x)'dx\right)C(1)'. \tag{76}$$

[4] This is result (65), which is included in Lemma 5 for later reference.

Thus, with $\beta_\perp$ defined in Lemma 2

$$\frac{1}{T}\beta_\perp'\widehat{S}_{1,1}\beta_\perp \Rightarrow \beta_\perp' C(1)\left(\int_0^1 BB'\right)C(1)'\beta_\perp = \int_0^1 B_{k-r}B_{k-r}' \tag{77}$$

where B_{k-r} is defined in (73) and $\int_0^1 B_{k-r}B_{k-r}'$ is a short-hand notation for $\int_0^1 B_{k-r}(x)B_{k-r}(x)'dx$. Moreover, it follows from Lemma 3 that

$$p\lim_{T\to\infty}\beta'\widehat{S}_{1,1}\beta = \Sigma_{\beta\beta}^*$$

Furthermore, it follows from (29) than

$$\beta' y_{t-1} = (\alpha'\alpha)^{-1}\alpha' x_t - (\alpha'\alpha)^{-1}\alpha'\Pi X_{t-1} - (\alpha'\alpha)^{-1}\alpha' C_0 u_t,$$

hence by Theorems 2-3 and Lemma 3,

$$\begin{aligned}
\beta'\widehat{S}_{1,1} &= \frac{1}{T}\sum_{t=1}^T \beta' y_{t-1}y_{t-1}' - \Sigma_{\beta X}\Sigma_{XX}^{-1}\left(\frac{1}{T}\sum_{j=1}^T x_t y_{t-1}'\right) + o_p(1) \qquad (78)\\
&= \frac{1}{T}\sum_{t=1}^T \beta' y_{t-1}y_{t-1}' + O_p(1)\\
&= (\alpha'\alpha)^{-1}\alpha'\frac{1}{T}\sum_{t=1}^T x_t y_{t-1}' - (\alpha'\alpha)^{-1}\alpha'\Pi\frac{1}{T}\sum_{t=1}^T X_{t-1}y_{t-1}'\\
&\quad - (\alpha'\alpha)^{-1}\alpha' C_0\frac{1}{T}\sum_{t=1}^T u_t y_{t-1}' + O_p(1)\\
&= O_p(1)
\end{aligned}$$

Thus, for the $k\times(k-r)$ matrix $\beta_\perp$ in Lemma 2,

$$\begin{pmatrix} T^{-1/2}\beta_\perp' \\ \beta' \end{pmatrix}\widehat{S}_{1,1}\left(T^{-1/2}\beta_\perp, \beta\right) \Rightarrow \begin{pmatrix} \int_0^1 B_{k-r}B_{k-r}' & O \\ O & \Sigma_{\beta\beta}^* \end{pmatrix} \tag{79}$$

To make this result comparable with Lemma 5, observe from (69) that

$$\begin{aligned}
\Theta\begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix} &= \begin{pmatrix} I_{k-r} & O_{k-r,r} \\ \Theta_{21} & I_r \end{pmatrix}\begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix}\\
&= \begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ T^{-1/2}\Theta_{21} & I_r \end{pmatrix}
\end{aligned}$$

hence it follows from (79) that

Lemma 6. *Under VECM* (37) *and Assumptions 1-2,*

$$\begin{aligned}
&\begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix} \Theta' \begin{pmatrix} \beta'_\perp \\ \beta' \end{pmatrix} \widehat{S}_{1,1} \left(\beta_\perp, \beta\right) \Theta \begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix} \\
&= \begin{pmatrix} T^{-1/2}\beta'_\perp + T^{-1/2}\Theta'_{21}\beta' \\ \beta' \end{pmatrix} \widehat{S}_{1,1} \left(T^{-1/2}\beta_\perp + T^{-1/2}\beta\Theta_{21}, \beta\right) \\
&\Rightarrow \begin{pmatrix} \int_0^1 B_{k-r}B'_{k-r} & O \\ O & \Sigma^*_{\beta\beta} \end{pmatrix}. \qquad (80)
\end{aligned}$$

where $\beta_\perp$ *is the* $k \times (k-r)$ *matrix in Lemma 2,* Θ *is defined by* (69) *and* (70), *and* $\Sigma^*_{\beta\beta}$ *is defined in Lemma 3.*

4.3.4 Limiting distributions of the general eigenvalues and the LR test statistics

Let $\Xi = (\beta_\perp, \beta)\,\Theta$. Since Ξ is non-singular, the generalized eigenvalue problem (48) is equivalent to finding $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq ... \geq \widehat{\lambda}_k$ such that for $i = 1, ..., k$,

$$\begin{aligned}
0 \;=\; \det\Bigg(&\widehat{\lambda}_i \begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix} \Xi'\widehat{S}_{1,1}\Xi \begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O & I_r \end{pmatrix} \qquad (81) \\
&- \begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix} \Xi'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\Xi \begin{pmatrix} T^{-1/2}I_{k-r} & O_{k-r,r} \\ O_{r,k-r} & I_r \end{pmatrix}\Bigg)
\end{aligned}$$

or equivalently,

$$0 = \det\left(T\widehat{\lambda}_i T^{-1}\Xi'\widehat{S}_{1,1}\Xi - \Xi'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\Xi\right), \qquad (82)$$

It follows from Anderson, Brons and Jensen (1983, Lemma 2)[5] that the solutions $\widehat{\lambda}_i$ of (81) converge in distribution to the corresponding solutions of the generalized eigenvalue problem

$$0 \;=\; \det\left(\lambda \begin{pmatrix} \int_0^1 B_{k-r}B'_{k-r} & O \\ O & \Psi_{**} \end{pmatrix} - \begin{pmatrix} O & O \\ O & \Psi_* \end{pmatrix}\right)$$

[5] See Lemma A.1 in the Appendix.

$$
\begin{aligned}
&= \det\left(\lambda \int_0^1 B_{k-r}B_{k-r}'\right)\det\left(\lambda\Sigma_{\beta\beta}^* - \Psi_*\right) \\
&= \lambda^{k-r}\det\left(\lambda\Sigma_{\beta\beta}^* - \Sigma_{\beta\beta}^*\left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma_{\beta\beta}^*\right)^{-1}\Sigma_{\beta\beta}^*\right) \\
&\quad \times \det\left(\int_0^1 B_{k-r}B_{k-r}'\right)
\end{aligned}
$$

Note that $\det\left(\lambda\Sigma_{\beta\beta}^* - \Sigma_{\beta\beta}^*\left(\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + \Sigma_{\beta\beta}^*\right)^{-1}\Sigma_{\beta\beta}^*\right) = 0$ is equivalent to

$$
\det\left(\lambda I_r - \left(\overline{\Psi} + I_r\right)^{-1}\right) = 0 \tag{83}
$$

where

$$
\overline{\Psi} = \left(\Sigma_{\beta\beta}^*\right)^{-1/2}\left(\alpha'\Omega^{-1}\alpha\right)\left(\Sigma_{\beta\beta}^*\right)^{-1/2} \tag{84}
$$

Consequently, the solutions of (83) are all between 0 and 1. Thus, the r largest ordered solutions of (81) converge in distribution to the correspondingly ordered solutions of (83). Since the latter are non-random, and convergence in distribution to a constant implies convergence in probability to that constant, it follows that $p\lim_{T\to\infty}\left(\widehat{\lambda}_1, .., \widehat{\lambda}_k\right)' = \left(\overline{\lambda}_1, .., \overline{\lambda}_r, 0, 0, ..., 0\right)'$, where $1 > \overline{\lambda}_1 \geq ... \geq \overline{\lambda}_r > 0$ are the ordered solutions of (83).

In the case of (82) it follows from Lemmas 5 and 6 that

$$
\Xi'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\Xi \Rightarrow \begin{pmatrix} \int_0^1 B_{k-r}dB_{k-r}' \int_0^1 \left(dB_{k-r}\right)B_{k-r}' & O \\ O & \Psi_* \end{pmatrix} \tag{85}
$$

$$
T^{-1}\Xi'\widehat{S}_{1,1}\Xi \Rightarrow \begin{pmatrix} \int_0^1 B_{k-r}B_{k-r}' & O \\ O & O \end{pmatrix}, \tag{86}
$$

Because the right-hand side matrix in (86) is singular, we cannot conclude directly from Anderson, Brons and Jensen (1983, Lemma 2) that $\left(T\widehat{\lambda}_{r+1}, .., T\widehat{\lambda}_k\right)'$ converges in distribution to the corresponding solutions of

$$
\begin{aligned}
0 &= \det\left[\lambda\begin{pmatrix} \int_0^1 B_{k-r}B_{k-r}' & O \\ O & O \end{pmatrix}\right. \\
&\quad \left. - \begin{pmatrix} \int_0^1 B_{k-r}dB_{k-r}' \int_0^1 \left(dB_{k-r}\right)B_{k-r}' & O \\ O & \Psi_* \end{pmatrix}\right] \qquad (87) \\
&= \det\left(\lambda\int_0^1 B_{k-r}B_{k-r}' - \int_0^1 B_{k-r}dB_{k-r}'\int_0^1 \left(dB_{k-r}\right)B_{k-r}'\right)\det\left(-\Psi_*\right).
\end{aligned}
$$

However, generalized eigenvalue problem (82) is equivalent to

$$0=\det\left(T^{-1}\Xi'\widehat{S}_{1,1}\Xi-\rho\Xi'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\Xi\right),\text{ where }\rho=\frac{1}{T\lambda} \tag{88}$$

In this form the result of Anderson, Brons and Jensen (1983, Lemma 2) is applicable, because the right-hand side matrix in (85) is non-singular. Thus, the ordered solutions $\widehat{\rho}_1 \leq \widehat{\rho}_2 \leq ... \leq \widehat{\rho}_k$ of (88) converge in distribution to the ordered solutions of

$$0=\det\left(\left(\begin{array}{cc}\int_0^1 B_{k-r}B'_{k-r} & O\\ O & O\end{array}\right)-\rho\left(\begin{array}{cc}\int_0^1 B_{k-r}dB'_{k-r}\int_0^1\left(dB_{k-r}\right)B'_{k-r} & O\\ O & \Psi_*\end{array}\right)\right).$$

It is now easy to verify that $\left(T\widehat{\lambda}_{r+1},..,T\widehat{\lambda}_k\right)'$ converges in distribution to $\left(\widetilde{\lambda}_1,..,\widetilde{\lambda}_{k-r}\right)'$, where $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq ... \geq \widetilde{\lambda}_{k-r}$ are the solutions of (87).

Summarizing, the following result has been shown:

Theorem 4. *Under Assumptions 1-2 and VECM* (37), *the ordered solutions* $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq ... \geq \widehat{\lambda}_k$ *of the generalized eigenvalue problem* (48) *satisfy*

$$p\lim_{T\to\infty}\left(\widehat{\lambda}_1,..,\widehat{\lambda}_k\right)'=\left(\overline{\lambda}_1,..,\overline{\lambda}_r,0,0,...,0\right)',$$

where $1>\overline{\lambda}_1 \geq ... \geq \overline{\lambda}_r > 0$ *are constants*[6], *and*

$$\left(T\widehat{\lambda}_{r+1},..,T\widehat{\lambda}_k\right)'\Rightarrow\left(\widetilde{\lambda}_1,..,\widetilde{\lambda}_{k-r}\right)', \tag{89}$$

where $\widetilde{\lambda}_1 \geq ... \geq \widetilde{\lambda}_{k-r}$ *are the ordered solutions of the generalized eigenvalue problem*

$$\det\left(\lambda\int_0^1 B_{k-r}B'_{k-r}-\int_0^1 B_{k-r}dB'_{k-r}\int_0^1\left(dB_{k-r}\right)B'_{k-r}\right)=0. \tag{90}$$

Consequently, under the null hypothesis $H_0(r)$ *that the cointegrating rank is* $r<k$ *the LR statistic* $LR_T(r|r+1)$ [*see* (54)] *converges in distribution to* $\widetilde{\lambda}_1$, *whereas under the alternative hypothesis* $H_1(r_1>r)$ *that the actual cointegrating rank* r_1 *is larger than* r,

$$p\lim_{T\to\infty}LR_T(r|r+1)/T=-\ln\left(1-\overline{\lambda}_{r+1}\right)>0.$$

[6] Recall that $\overline{\lambda}_1,...,\overline{\lambda}_r$ are the solutions of eigenvalue problem (83).

Moreover, under $H_0(r)$ the trace test statistic $LR_T(r|k)$ [see (55)] converges in distribution to

$$\sum_{i=1}^{k-r} \widetilde{\lambda}_i = \text{trace}\left(\int_0^1 (dB_{k-r})\, B'_{k-r}\left(\int_0^1 B_{k-r}B'_{k-r}\right)^{-1}\int_0^1 B_{k-r}dB'_{k-r}\right), \tag{91}$$

whereas under $H_1(r_1 > r)$,

$$p\lim_{T\to\infty} LR_T(r|k)/T = -\sum_{i=r+1}^{r_1} \ln\left(1-\overline{\lambda}_i\right) > 0.$$

Remark: The proof of (89) in this lecture note is different from the original proof by Johansen. Johansen (1995, page 159) uses the fact that

$$\begin{aligned}
&\det\left[(\beta,\beta_\perp)'\, S(\rho)\,(\beta,\beta_\perp)\right] = \det\begin{pmatrix} \beta' S(\rho)\beta & \beta' S(\rho)\beta_\perp \\ \beta'_\perp S(\rho)\beta & \beta'_\perp S(\rho)\beta_\perp \end{pmatrix}\\
&= \det\left(\beta' S(\rho)\beta\right)\det\left(\beta'_\perp\left(S(\rho) - S(\rho)\beta\left(\beta' S(\rho)\beta\right)^{-1}\beta' S(\rho)\right)\beta_\perp\right)\\
&= \det\left(-\beta'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\beta + o_p(1)\right)\det\left(\rho.\beta_\perp T^{-1}\widehat{S}_{1,1}\beta_\perp - \beta'_\perp\widehat{S}_{1,0}\widehat{N}\widehat{S}_{0,1}\beta_\perp\right)
\end{aligned}$$

where

$$S(\rho) = \rho T^{-1}\widehat{S}_{1,1} - \widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1},\ \rho = T\lambda = O_p(1),$$

and

$$\begin{aligned}
\widehat{N} &= \widehat{S}_{0,0}^{-1} - \widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\beta\left(\beta'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\beta\right)^{-1}\beta'\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\\
&= \alpha_\perp\left(\alpha'_\perp\Omega\alpha_\perp\right)^{-1}\alpha'_\perp + o_p(1)
\end{aligned}$$

Since by Lemmas 5 and 6,

$$\begin{aligned}
\beta'_\perp T^{-1}\widehat{S}_{1,1}\beta_\perp &\Rightarrow \int_0^1 B_{k-r}B'_{k-r},\\
\beta_\perp\widehat{S}_{1,0}\widehat{N}\widehat{S}_{0,1}\beta_\perp &\Rightarrow \int_0^1 B_{k-r}dB'_{k-r}\int_0^1 (dB_{k-r})\, B'_{k-r},
\end{aligned}$$

the result (89) follows.

4.4 The VECM(p) case with intercepts due to initial values

The assumption $\pi_0 = 0$ is crucial for the results in Theorem 4. In the cases (29) and (36) the results of Theorem 4 change. I will only demonstrate this for the case (29). For the case (36), see Johansen (1995, Theorem 11.1). Recall from (34) that in VECM (29) the vector of intercept π_0 is due to initial values. Therefore, Δy_t is still a zero-mean stationary process. The only assumption that has to be dropped is the assumption that $u_t = 0$ for $t < 1$.

To demonstrate how the results in Theorem 4 change in the case (29) with $\pi_0 \neq 0$, write this model as

$$\begin{aligned} x_t &= \alpha\beta' y_{t-1} + (\pi_0, \Pi)\begin{pmatrix} 1 \\ X_{t-1} \end{pmatrix} + C_0 u_t \\ &= \alpha\beta' y_{t-1} + \Pi_* \widetilde{X}_{t-1} + C_0 u_t \end{aligned}$$

where

$$\Pi_* = (\pi_0, \Pi)\,,\ \widetilde{X}_{t-1} = \left(1, X'_{t-1}\right)',$$

with X_{t-1} and Π defined by (38).

The main changes occur in the limiting distributions of the matrices $\widehat{S}_{0,1}$ and $\widehat{S}_{1,1}$.

4.4.1 The matrix $\widehat{S}_{0,1}$

The matrix (43) now becomes

$$\begin{aligned} \widehat{S}_{0,1} &= \alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}y'_{t-1}\right) \\ &\quad -\alpha\left(\frac{1}{T}\sum_{t=1}^{T}\beta' y_{t-1}\widetilde{X}'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}y'_{t-1}\right) \\ &\quad +C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t y'_{t-1}\right) \\ &\quad -C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t\widetilde{X}'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}y'_{t-1}\right) \end{aligned} \qquad (92)$$

The problem is that (63) does no longer hold if we replace X_{t-1} by $\widetilde{X}_{t-1}$, so that the last term in (92) is no longer of order $o_p(1)$. Instead, we now have that

Lemma 7. *Under Assumption 1-2 and VECM* (29),

$$\left(\frac{1}{T}\sum_{t=1}^{T}u_t\widetilde{X}'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}y'_{t-1}\right)$$
$$=\frac{1}{T}\sum_{t=1}^{T}u_t\frac{1}{T}\sum_{t=1}^{T}y'_{t-1}+o_p(1)\Rightarrow B(1)\int_0^1 B(x)'dxC(1)'.$$

Proof: It follows from the easy convergence result (see the Appendix)

$$\frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y_{t-1}\Rightarrow C(1)\int_0^1 B(x)dx \tag{93}$$

and (63) that

$$\begin{pmatrix} 1/\sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}y'_{t-1} = \begin{pmatrix} \frac{1}{T\sqrt{T}}\sum_{t=1}^{T}y'_{t-1} \\ \frac{1}{T}\sum_{t=1}^{T}X_{t-1}y'_{t-1} \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^1 B(x)'dxC(1)' \\ M_* \end{pmatrix}$$

Moreover, it is easy to verify that under Assumption 1,

$$\frac{1}{T}\sum_{t=1}^{T}u_t\widetilde{X}'_{t-1}\begin{pmatrix} \sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix} = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_t,\frac{1}{T}\sum_{t=1}^{T}u_tX'_{t-1}\right) \Rightarrow \left(B(1),O_{k,k(p-1)}\right).$$

Furthermore,

$$\begin{pmatrix} 1/\sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)^{-1}\begin{pmatrix} \sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix} \tag{94}$$
$$=\left(\begin{pmatrix} 1/\sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix}\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\begin{pmatrix} \sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix}\right)^{-1}$$
$$=\begin{pmatrix} 1 & \frac{1}{T}\sum_{t=1}^{T}X'_{t-1} \\ \frac{1}{T}\sum_{t=1}^{T}X_{t-1} & \frac{1}{T}\sum_{t=1}^{T}X_{t-1}X'_{t-1} \end{pmatrix}^{-1},$$

Since under Assumption 1, $p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}X_{t-1} = 0$ and $p\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}X_{t-1}X_{t-1}' = \Sigma_{XX}$, it follows from (94) that

$$p\lim_{T\to\infty}\begin{pmatrix} 1/\sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}_{t-1}'\right)^{-1}\begin{pmatrix} \sqrt{T} & 0' \\ 0 & I_{k(p-1)} \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0' \\ 0 & \Sigma_{XX}^{-1} \end{pmatrix}.$$

Lemma 7 follows straightforwardly from these convergence results. Q.E.D.

The previous result (65) now becomes

$$\alpha_{\perp}'\widehat{S}_{0,1} = \alpha_{\perp}'C_0\left(\frac{1}{T}\sum_{t=1}^{T}u_t\left(y_{t-1}-\frac{1}{T}\sum_{j=1}^{T}y_{j-1}\right)'\right)+o_p(1),$$

where

$$\frac{1}{T}\sum_{t=1}^{T}u_t\left(y_{t-1}-\frac{1}{T}\sum_{j=1}^{T}y_{j-1}\right)' \Rightarrow \left(\int_0^1(dB)\,B' - B(1)\int_0^1 B(x)'dx\right)C(1)'.$$

Note that, with $S_{t-1}=\sum_{j=1}^{t-1}u_j$ and $\overline{S}_{-1}=(1/T)\sum_{t=1}^{T}S_{t-1}$,

$$\frac{1}{T}\sum_{t=1}^{T}u_t\left(S_{t-1}-\overline{S}_{-1}\right)' \Rightarrow \int_0^1(dB)\,B' - B(1)\int_0^1 B(x)'dx \tag{95}$$

Since

$$\left(S_{[xT]-1}-\overline{S}_{-1}\right)/\sqrt{T} \Rightarrow B(x)-\int_0^1 B(y)dy = \overline{B}(x), \tag{96}$$

say, where $\overline{B}(x)$ is known as a demeaned k-variate standard Brownian motion, the right-hand side of (95) will be denoted by

$$\int_0^1(dB)\,\overline{B}' \equiv \int_0^1(dB)\,B' - B(1)\int_0^1 B(x)'dx. \tag{97}$$

With this change of notation, (72) reads

$$\Theta'\begin{pmatrix} \beta_{\perp}' \\ \beta' \end{pmatrix}\widehat{S}_{1,0}\widehat{S}_{0,0}^{-1}\widehat{S}_{0,1}\left(\beta_{\perp},\beta\right)\Theta \Rightarrow \tag{98}$$
$$\begin{pmatrix} \beta_{\perp}C(1)\int_0^1\overline{B}dB'C_0'\alpha_{\perp}\left(\alpha_{\perp}'\Omega\alpha_{\perp}\right)^{-1}\alpha_{\perp}'C_0\int_0^1(dB)\,\overline{B}'C(1)'\beta_{\perp} & O \\ O & \Psi_* \end{pmatrix}$$

where Ψ_* is a nonrandom $r\times r$ matrix, similar (but not equal) to (74), and Θ defined similar to (69).

4.4.2 The matrix $\widehat{S}_{1,1}$

The matrix (44) now becomes

$$\begin{aligned}\widehat{S}_{1,1} &= \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-1}' \\ &- \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\widetilde{X}_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}y_{t-1}'\right).\end{aligned}$$

Again, the result (76) does no longer hold, because similar to Lemma 7 we have that

Lemma 8. *Under Assumption 1 and VECM* (29),

$$\begin{aligned}&\frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\widetilde{X}_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}_{t-1}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}y_{t-1}'\right) \\ &\Rightarrow C(1)\int_0^1 B(x)dx\int_0^1 B(x)'dxC(1)'.\end{aligned}$$

Hence, (76) now becomes,

$$\begin{aligned}\frac{1}{T}\widehat{S}_{1,1} &\Rightarrow C(1)\left(\int_0^1 B(x)B(x)'dx - \int_0^1 B(x)dx\int_0^1 B(x)'dx\right)C(1)' \\ &= C(1)\int_0^1\left(B(x)-\int_0^1 B(y)dy\right)\left(B(x)-\int_0^1 B(y)dy\right)'dxC(1)' \\ &= C(1)\int_0^1 \overline{B}(x)\overline{B}(x)'dxC(1)'.\end{aligned}$$

4.4.3 Limiting distributions of the general eigenvalues and the LR test statistics

It is now easy to verify that

Theorem 5. *If we change* (90) *to*

$$\det\left(\lambda\int_0^1 \overline{B}_{k-r}\overline{B}_{k-r}' - \int_0^1 \overline{B}_{k-r}dB_{k-r}'\int_0^1\left(dB_{k-r}\right)\overline{B}_{k-r}'\right) = 0$$

and (91) *to*

$$\sum_{i=1}^{k-r}\widetilde{\lambda}_i = \text{trace}\left(\int_0^1 (dB_{k-r})\,\overline{B}'_{k-r}\left(\int_0^1 \overline{B}_{k-r}\overline{B}'_{k-r}\right)^{-1}\int_0^1 \overline{B}_{k-r}dB'_{k-r}\right),$$

where[7]

$$\overline{B}_{k-r}(x) = B_{k-r}(x) - \int_0^1 B_{k-r}(y)dy,$$

then the results of Theorem 4 carry over to VECM (29).

5 Asymptotic properties of the ML estimators of α, β and Ω

The partially concentrated log-likelihood (42) can be written as

$$\begin{aligned}
&\ln L_T(\alpha,\beta,\Omega) = \max_{\Pi}\ln L_T(\alpha,\beta,\Pi,\Omega)\\
&= -\frac{1}{2}T.\text{trace}\left(\Omega^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(R_{0,t}-\alpha\beta' R_{1,t}\right)\left(R_{0,t}-\alpha\beta' R_{1,t}\right)'\right)\\
&-\frac{1}{2}T.\ln\left(\det\Omega\right) - T.k\ln\left(\sqrt{2\pi}\right)\\
&= -\frac{1}{2}T.\text{trace}\left(\Omega^{-1}\left(\widehat{S}_{0,0}-\alpha\beta'\widehat{S}_{1,0}-\widehat{S}_{0,1}\beta\alpha'+\alpha\beta'\widehat{S}_{1,1}\beta\alpha'\right)\right) \qquad (99)\\
&-\frac{1}{2}T.\ln\left(\det\Omega\right) - T.k\ln\left(\sqrt{2\pi}\right)
\end{aligned}$$

with corresponding ML estimators

$$\left(\widehat{\alpha},\widehat{\beta},\widehat{\Omega}\right) = \arg\max_{\alpha,\beta,\Omega}\ln L_T(\alpha,\beta,\Omega) \qquad (100)$$

Although $\widehat{\alpha}$ and $\widehat{\beta}$ themselves are not unique, $\widehat{\alpha}\widehat{\beta}'$ is unique, and therefore $\widehat{\Omega}$ is unique. The same applies to α and β, of course. Nevertheless, after suitable normalization these ML estimators are consistent:

[7] Again, $\int_0^1 \overline{B}_{k-r}\overline{B}'_{k-r}$ is a short-hand notation for $\int_0^1 \overline{B}_{k-r}(x)\overline{B}_{k-r}(x)'dx$.

Theorem 6. *Let* $\left(\widehat{\alpha}, \widehat{\beta}, \widehat{\Omega}\right)$ *be ML estimators of* (α, β, Ω). *Without loss of generality we may assume that*

$$\widehat{\beta}'\widehat{\beta} = O_p(1),\ \left(\widehat{\beta}'\widehat{\beta}\right)^{-1} = O_p(1),\ \beta'\widehat{\beta} = O_p(1), \tag{101}$$

and that the columns of $\widehat{\beta}$ *and* β *have been rescaled such that*

$$\det\left(\widehat{\beta}'\widehat{\beta}\right) = \det\left(\beta'\beta\right) = 1. \tag{102}$$

Then under Assumptions 1-2 and VECM (29),

$$\left(\det\left(\beta'\widehat{\beta}\right)\right)^2 = 1 + O_p\left(T^{-1}\right). \tag{103}$$

Consequently, $\widetilde{\beta} = \widehat{\beta}\left(\beta'\widehat{\beta}\right)^{-1}(\beta'\beta)$ *exists with probability converging to* 1. *The matrix* $\widetilde{\beta}$ *of normalized estimated cointegrating vectors is super consistent:*

$$\widetilde{\beta} - \beta = \beta_{\perp}U_T, \text{with } U_T = O_p\left(T^{-1}\right), \tag{104}$$

Moreover, $\widetilde{\alpha} = \widehat{\alpha}\left(\beta'\beta\right)^{-1}\left(\beta'\widehat{\beta}\right)$ *is a consistent estimator of* α *and* $\widehat{\Omega}$ *is a consistent estimator of* Ω.

Proof: Appendix

Note that $\left(\widetilde{\alpha}, \widetilde{\beta}, \widehat{\Omega}\right)$ also maximizes the log-likelihood (99). The first-order conditions involved are

$$\begin{aligned} O_{r,k} &= \widetilde{\alpha}'\widehat{\Omega}\left(\widehat{S}_{0,1} - \widetilde{\alpha}\widetilde{\beta}'\widehat{S}_{1,1}\right) & (105)\\ O_{k,r} &= \left(\widehat{S}_{0,1} - \widetilde{\alpha}\widetilde{\beta}'\widehat{S}_{1,1}\right)\widetilde{\beta} & (106)\\ \widehat{\Omega} &= \widehat{S}_{0,0} - \widetilde{\alpha}\widetilde{\beta}'\widehat{S}_{1,0} - \widehat{S}_{0,1}\widetilde{\beta}\widetilde{\alpha}' + \widetilde{\alpha}\widetilde{\beta}'\widehat{S}_{1,1}\widetilde{\beta}\widetilde{\alpha}' & (107) \end{aligned}$$

This is not too hard to verify for the case $r = 1$, but these conditions hold for $1 \leq r < k$ as well.

The limiting distribution of $T\left(\widetilde{\beta} - \beta\right)$ is given in Theorem 7:

Theorem 7. *Let $\widetilde{\beta}$ and $\beta_\perp$ be the defined as in Theorem 6 and Lemma 2, respectively, and let Assumptions 1-2 hold. Then in the case of VECM* (37),

$$\begin{aligned} T\left(\widetilde{\beta}-\beta\right) &\Rightarrow \beta_\perp \left(\beta_\perp' C(1)C(1)'\beta_\perp\right)^{-1/2} \\ &\quad \times \left(\int_0^1 B_{k-r}B_{k-r}'\right)^{-1} \int_0^1 B_{k-r}dB_\alpha' \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \end{aligned}$$

whereas in the case of VECM (29),

$$\begin{aligned} T\left(\widetilde{\beta}-\beta\right) &\Rightarrow \beta_\perp \left(\beta_\perp' C(1)C(1)'\beta_\perp\right)^{-1/2} \\ &\quad \times \left(\int_0^1 \overline{B}_{k-r}\overline{W}_{k-r}'\right)^{-1} \int_0^1 \overline{B}_{k-r}dB_\alpha' \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \end{aligned}$$

where B_α is an r-variate standard Brownian motion which is independent of B_{k-r} and $\overline{B}_{k-r}$

Proof: Appendix

6 Drift

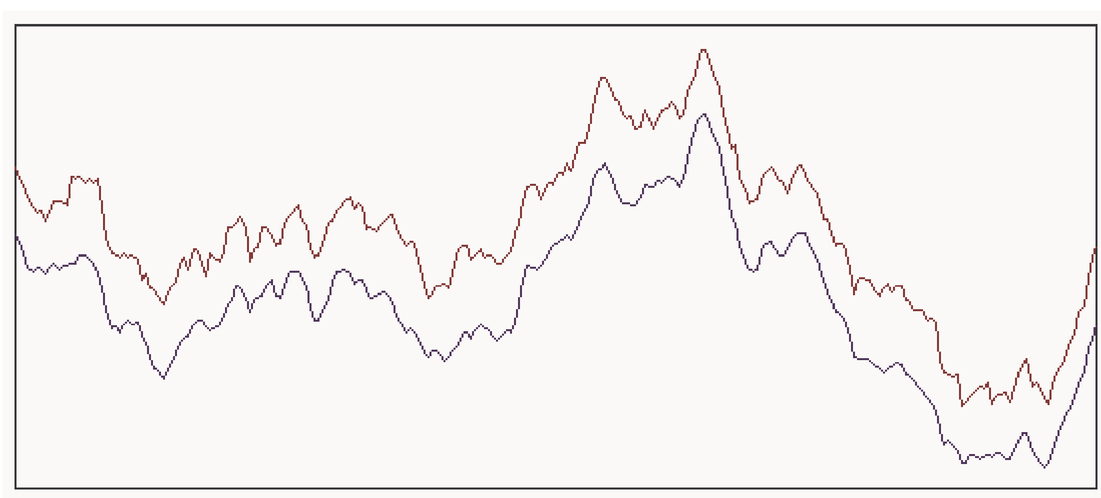

Figure 1. Cointegration when there is no drift

The two cases considered so far only differ regarding the treatment of initial values; the assumption that Δy_t is a zero-mean stationary process satisfying the conditions in Assumption 1 has been maintained in both cases. Recall from (25) that in this case the cointegrating relationship takes the form $\beta' y_t = \beta'(y_0 - v_0) + \beta' v_t$, where $\beta' v_t$ is a zero-mean stationary process. In the bivariate case $k = 2$ the time series look like in Figure 1 above. In this case the time series run parallel and approximately horizontal, where the distance between the time series is due to the initial values $\beta'(y_0 - v_0)$.

However, this pattern is rare for macro-economic time series. Most macro-economic time series have a trending pattern, due to drift: $E[\Delta y_t] = \mu \neq 0$, where μ is the vector of drift parameters. Thus, suppose that instead of (7) in Assumption 1,

$$y_t - y_{t-1} = \mu + C(L)u_t. \tag{108}$$

Then (11) becomes

$$\begin{aligned} y_t &= \sum_{j=1}^{t} (\mu + x_j) + y_0 = \mu t + \sum_{j=1}^{t} x_j + y_0 \\ &= (y_0 - v_0) + \mu t + C(1) \sum_{j=1}^{t} u_j + v_t \end{aligned} \tag{109}$$

Thus the expectation vector $\mu = E[\Delta y_t]$ now becomes a vector of trend parameters! Moreover,

$$\beta' y_t = \beta' (y_0 - v_0) + \beta' \mu t + \beta' v_t \tag{110}$$

so that $\beta' y_t$ is now trend stationary. If so, in the bivariate case $k = 2$ the time series look like in Figure 2 below. In this case the two time series drift apart, due to the time trend in the cointegrating relationship, and both are (upwards)[8] sloping due to the drift parameters.

[8] Drift can also generate downwards sloping patterns, but that never happens for macro-economic time series.

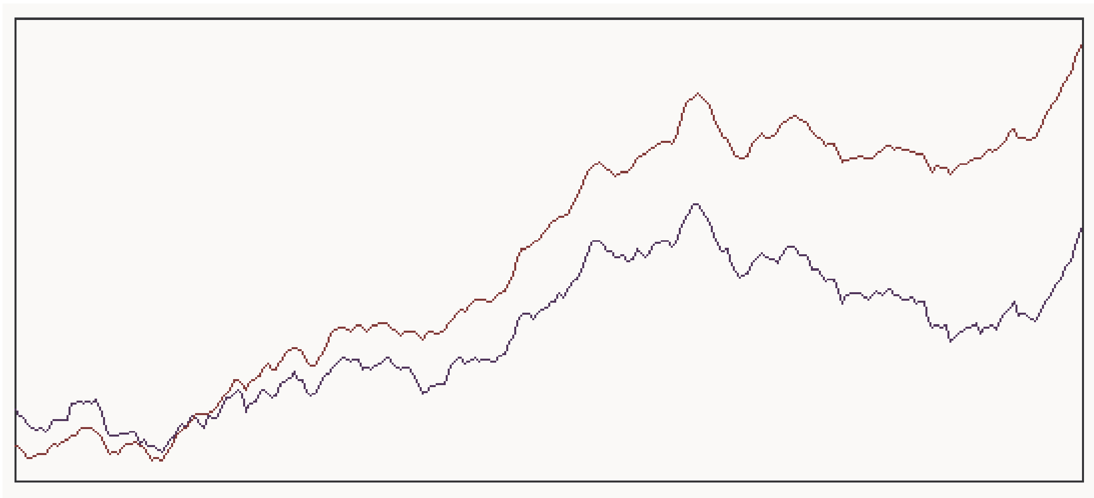

Figure 2: Drift, with trend in the cointegrating relationship

However, a more common pattern is displayed in Figure 3.

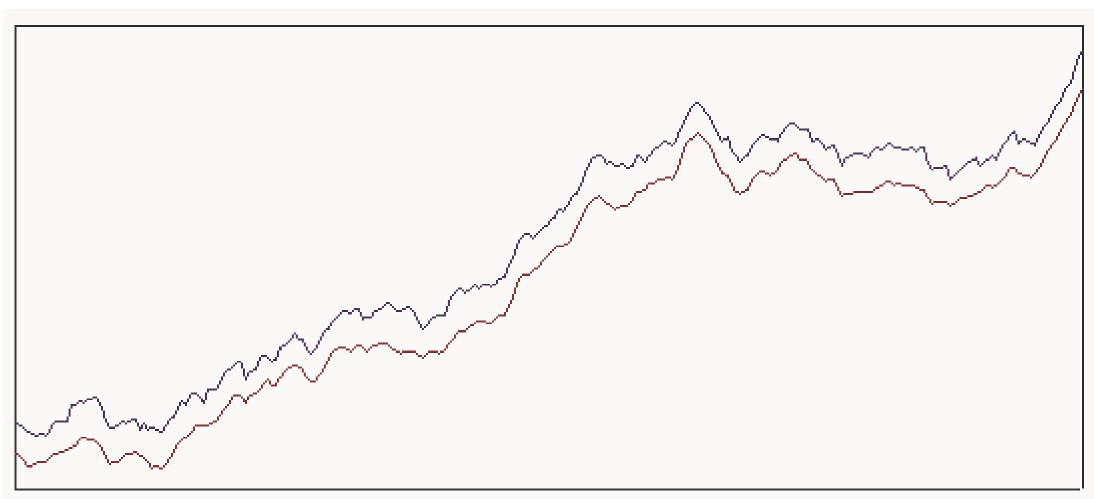

Figure 3: Drift, with constant in the cointegrating relationship

In this case the two time series run approximately parallel, but still (upwards) sloping due to the drift parameters. This pattern is only possible if in (110),

$$\beta' \mu = 0 \tag{111}$$

so that the cointegrating relationship becomes

$$\beta' y_t = \beta' (y_0 - v_0) + \beta' v_t. \tag{112}$$

It is easy to verify that the Granger representation theorem carries over if we replace y_t in (26) with $y_t - \mu.t$, which then gives rise to a VECM(p) model of the form

$$\Delta y_t - \mu = \pi_0 + \alpha\beta'(y_{t-1} - (t-1)\mu) + \sum_{j=1}^{p-1} \Pi_j (\Delta y_{t-j} - \mu) + C_0 u_t, \tag{113}$$

or equivalently,

$$\Delta y_t = \pi_{00} + \alpha\beta'(y_{t-1} - (t-1)\mu) + \sum_{j=1}^{p-1} \Pi_j \Delta y_{t-j} + C_0 u_t, \tag{114}$$

where

$$\pi_{00} = \pi_0 + \left(I_k - \sum_{j=1}^{p-1} \Pi_j \right) \mu.$$

This VECM(p) corresponds to the case (108) as displayed in Figure 2.

Under the restriction (111) the model becomes

$$\Delta y_t = \pi_{00} + \alpha\beta' y_{t-1} + \sum_{j=1}^{p-1} \Pi_j \Delta y_{t-j} + C_0 u_t, \tag{115}$$

which at first sight looks the same as (26). However, the crucial difference is that now the vector π_{00} of intercepts depends on the drift parameters, which generates the sloping patterns as in Figure 3, whereas in the case (26) the vector of intercepts π_0 depends only on initial values. These initial values are not able to generate drift, as illustrated in Figure 1.

Due to the drift the results in Theorem 5 change, in different ways depending on whether condition (111) is imposed or not. See Johansen (1995, Theorem 11.1).

7 Appendix

7.1 Convergence of generalized eigenvalues

The following lemma is a corollary of Lemma 2 of Anderson, Brons and Jensen (1983):

Lemma A.1. *Let A_T be a positive definite $m \times m$ matrix and let B_T be a symmetric $m \times m$ matrix, satisfying $A_T \Rightarrow A$ and $B_T \Rightarrow B$, where $\det(A) > 0$. Let $\lambda_{1,T} \geq \lambda_{2,T} \geq ... \geq \lambda_{m,T}$ be the ordered solutions of the generalized eigenvalue problem $\det(\lambda A_T - B_T) = 0$, and let $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$ be the ordered solutions of the generalized eigenvalue problem $\det(\lambda A - B) = 0$. Then $(\lambda_{1,T}, \lambda_{2,T}, ..., \lambda_{m,T})' \Rightarrow (\lambda_1, \lambda_2, ..., \lambda_m)'$.*

7.2 Derivation of (93)

Recall from (11) that $y_t = C(1) \sum_{j=1}^{t} u_t + v_t + y_0 - v_0$, so that

$$\begin{aligned}
\frac{1}{T\sqrt{T}} \sum_{t=1}^{T} y_{t-1} &= C(1) \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{t-1} u_t \right) + \frac{1}{T\sqrt{T}} \sum_{t=1}^{T} v_{t-1} \\
&\quad + \frac{1}{\sqrt{T}} (y_0 - v_0) \\
&= C(1) \frac{1}{T} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{t-1} u_t \right) + O_p\left(T^{-1/2}\right) \\
&= C(1) \frac{1}{T} \sum_{t=1}^{T} \int_{t-1}^{t} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{[z]} u_t \right) dz + O_p\left(T^{-1/2}\right) \\
&= C(1) \int_{0}^{T} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{[z]} u_t \right) d\left(z/T\right) + O_p\left(T^{-1/2}\right) \\
&= C(1) \int_{0}^{1} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{[xT]} u_t \right) dx + O_p\left(T^{-1/2}\right) \\
&\Rightarrow C(1) \int_{0}^{1} B(x) dx
\end{aligned}$$

The latter follows from

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{[xT]} u_t \Rightarrow B(x)$$

7.3 Proof of Theorem 6

Proof of (103)

Recall from (52) that the columns $\widehat{\beta}_1, ..., \widehat{\beta}_r$ of $\widehat{\beta}$ are the eigenvectors corresponding to the r largest solutions of (48): $\widehat{\lambda}_i \widehat{S}_{1,1} \widehat{\beta}_i = \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \widehat{\beta}_i$, $i = 1, 2, .., r$. Hence, denoting $\widehat{\Lambda}_r = \text{diag}\left(\widehat{\lambda}_1, ..., \widehat{\lambda}_r\right)$, it follows that

$$\widehat{S}_{1,1} \widehat{\beta} \widehat{\Lambda}_r = \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \widehat{\beta} \tag{116}$$

Next, multiply equation (116) from the left side by $\left(\beta, T^{-1/2}\beta_\perp\right)$, and use the fact that

$$\left(\beta, T^{-1/2}\beta_\perp\right)^{-1} \widehat{\beta} = \begin{pmatrix} \overline{\beta}' \widehat{\beta} \\ T^{1/2} \overline{\beta}'_\perp \widehat{\beta} \end{pmatrix} \tag{117}$$

Then (116) becomes

$$\begin{aligned} &\begin{pmatrix} \beta' \widehat{S}_{1,1} \beta & T^{-1/2} \beta' \widehat{S}_{1,1} \beta_\perp \\ T^{-1/2} \beta'_\perp \widehat{S}_{1,1} \beta & T^{-1} \beta'_\perp \widehat{S}_{1,1} \beta_\perp \end{pmatrix} \begin{pmatrix} \overline{\beta}' \widehat{\beta} \\ T^{1/2} \overline{\beta}'_\perp \widehat{\beta} \end{pmatrix} \widehat{\Lambda}_r \\ &= \begin{pmatrix} \beta' \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta & T^{-1/2} \beta' \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta_\perp \\ T^{-1/2} \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta & T^{-1} \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta_\perp \end{pmatrix} \begin{pmatrix} \overline{\beta}' \widehat{\beta} \\ T^{1/2} \overline{\beta}'_\perp \widehat{\beta} \end{pmatrix} \end{aligned}$$

which implies

$$\begin{aligned} &\beta' \widehat{S}_{1,1} \beta \left(\overline{\beta}' \widehat{\beta}\right) \widehat{\Lambda}_r + T^{-1/2} \beta' \widehat{S}_{1,1} \beta_\perp T^{1/2} \overline{\beta}'_\perp \widehat{\beta} \widehat{\Lambda}_r \\ &= T^{-1/2} \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta \left(\overline{\beta}' \widehat{\beta}\right) + T^{-1/2} \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta_\perp \left(\overline{\beta}'_\perp \widehat{\beta}\right) \end{aligned}$$

and

$$\begin{aligned} &T^{-1/2} \beta'_\perp \widehat{S}_{1,1} \beta \left(\overline{\beta}' \widehat{\beta}\right) \widehat{\Lambda}_r + T^{-1} \beta'_\perp \widehat{S}_{1,1} \beta_\perp T^{1/2} \overline{\beta}'_\perp \widehat{\beta} \widehat{\Lambda}_r \\ &= T^{-1/2} \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta \left(\overline{\beta}' \widehat{\beta}\right) + T^{-1/2} \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta_\perp \overline{\beta}'_\perp \widehat{\beta} \end{aligned}$$

The latter equality can be rewritten as

$$\begin{aligned} T \overline{\beta}'_\perp \widehat{\beta} &= \left(T^{-1} \beta'_\perp \widehat{S}_{1,1} \beta_\perp\right)^{-1} \\ &\times \left(\beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta \left(\overline{\beta}' \widehat{\beta}\right) \widehat{\Lambda}_r^{-1} + \beta'_\perp \widehat{S}_{1,0} \widehat{S}_{0,0}^{-1} \widehat{S}_{0,1} \beta_\perp \overline{\beta}'_\perp \widehat{\beta} \widehat{\Lambda}_r^{-1} \right. \\ &\qquad \left. - \beta'_\perp \widehat{S}_{1,1} \beta \left(\overline{\beta}' \widehat{\beta}\right)\right) \end{aligned} \tag{118}$$

It follows now from Lemmas 5 and 6 and Theorem 4 that the right-hand side of (118) is of order $O_p(1)$, hence $T\overline{\beta}'_{\perp}\widehat{\beta} = O_p(1)$ and thus

$$\beta'_{\perp}\widehat{\beta} = O_p(T^{-1}). \tag{119}$$

We can always write $\widehat{\beta}$ as

$$\widehat{\beta} = \beta\left(\beta'\beta\right)^{-1}\left(\beta'\widehat{\beta}\right) + \beta_{\perp}\left(\beta'_{\perp}\beta_{\perp}\right)^{-1}\left(\beta'_{\perp}\widehat{\beta}\right), \tag{120}$$

Hence

$$I_r = \left(\widehat{\beta}'\widehat{\beta}\right)^{-1/2}\left(\widehat{\beta}'\beta\right)\left(\beta'\beta\right)^{-1}\left(\beta'\widehat{\beta}\right)\left(\widehat{\beta}'\widehat{\beta}\right)^{-1/2} + O_p(T^{-1}), \tag{121}$$

where the $O_p(T^{-1})$ term follows from (101) and (119). Now (103) follows easily from (121) and the normalizations (102).

Proof of (104)

Recall that $\left(\beta'_{\perp}\widehat{\beta}\right)^{-1} = \left(\det\left(\beta'_{\perp}\widehat{\beta}\right)\right)^{-1}\left(\beta'_{\perp}\widehat{\beta}\right)_{Adj}$, where $\left(\beta'_{\perp}\widehat{\beta}\right)_{Adj}$ is the adjoint of $\beta'_{\perp}\widehat{\beta}$, i.e., the matrix of cofactors of $\left(\beta'_{\perp}\widehat{\beta}\right)'$. Since $\beta'_{\perp}\widehat{\beta} = O_p(1)$ implies $\left(\beta'_{\perp}\widehat{\beta}\right)_{Adj} = O_p(1)$, it follows from (103) that

$$\left(\beta'_{\perp}\widehat{\beta}\right)^{-1} = O_p(1).$$

It follows now from (119) and (120) that

$$\widetilde{\beta} = \widehat{\beta}\left(\beta'\widehat{\beta}\right)^{-1}\left(\beta'\beta\right) = \beta + \beta_{\perp}U_T,$$

where

$$U_T = \left(\beta'_{\perp}\beta_{\perp}\right)^{-1}\left(\beta'_{\perp}\widehat{\beta}\right)\left(\beta'\widehat{\beta}\right)^{-1}\left(\beta'\beta\right) = O_p(T^{-1}).$$

Consistency of $\widetilde{\alpha}$ and $\widehat{\Omega}$

Recall from (78) and Lemma 8 that $\beta'\widehat{S}_{1,1} = O_p(1)$, and from the proofs of Lemmas 5 and 7 that $\widehat{S}_{0,1} = O_p(1)$. Hence it follows from Lemma 9 and (104) that in the case (37),

$$\begin{aligned}
\widetilde{\beta}'\widehat{S}_{1,1}\widetilde{\beta} &= \left(\beta' + O_p\left(T^{-1}\right)\right)\widehat{S}_{1,1}\left(\beta + O_p\left(T^{-1}\right)\right) \\
&= \beta'\widehat{S}_{1,1}\beta + O_p\left(T^{-1}\right) = \Sigma^*_{\beta\beta} + o_p(1) &\quad (122)\\
\widehat{S}_{0,1}\widetilde{\beta} &= \widehat{S}_{0,1}\left(\beta + O_p\left(T^{-1/2}\right)\right) = \widehat{S}_{0,1}\beta + O_p\left(T^{-1}\right) \\
&= \alpha\Sigma^*_{\beta\beta} + o_p(1) &\quad (123)
\end{aligned}$$

where the last equality in (122) follows from Lemma 6 and the last equality in (123) follows from Lemma 5. It follows therefore from (106) that

$$\widetilde{\alpha} = \widehat{S}_{0,1}\widetilde{\beta}\left(\widetilde{\beta}'\widehat{S}_{1,1}\widetilde{\beta}\right)^{-1} = \alpha + o_p(1) \tag{124}$$

This result carries over to the case of VECM (29).

It follows now from (107), (122), (123), (124) and Lemma 4 that

$$\widehat{\Omega} = \widehat{S}_{0,0} - \alpha\Sigma^*_{\beta\beta}\alpha' + o_p(1) = \Omega + o_p(1) \tag{125}$$

Again, this result carries over to the case of VECM (29).

7.4 Proof of Theorem 7

It follows from (105) that

$$\begin{aligned} O_{r,k-r} &= \widetilde{\alpha}'\widehat{\Omega}^{-1}\left(\widehat{S}_{0,1} - \widetilde{\alpha}\widetilde{\beta}'\widehat{S}_{1,1}\right)\beta_\perp \\ &= \widetilde{\alpha}'\widehat{\Omega}^{-1}\left(\widehat{S}_{0,1} - \alpha\beta'\widehat{S}_{1,1} - (\widetilde{\alpha}-\alpha)\widetilde{\beta}'\widehat{S}_{1,1} - \alpha\left(\widetilde{\beta}-\beta\right)'\widehat{S}_{1,1}\right)\beta_\perp \\ &= \widetilde{\alpha}'\widehat{\Omega}^{-1}\left(\widehat{S}_{0,1} - \alpha\beta'\widehat{S}_{1,1}\beta_\perp\right)\beta_\perp - \widetilde{\alpha}'\widehat{\Omega}^{-1}(\widetilde{\alpha}-\alpha)\left(\widetilde{\beta}-\beta\right)'\widehat{S}_{1,1}\beta_\perp \\ &\quad -\widetilde{\alpha}'\widehat{\Omega}^{-1}(\widetilde{\alpha}-\alpha)\beta'\widehat{S}_{1,1}\beta_\perp - \widetilde{\alpha}'\widehat{\Omega}^{-1}\alpha\left(\widetilde{\beta}-\beta\right)'\widehat{S}_{1,1}\beta_\perp \end{aligned}$$

Substituting (104) in this equation and multiplying from the right side by $\left(\widetilde{\alpha}'\widehat{\Omega}^{-1}\alpha\right)^{-1}$ yield

$$\begin{aligned} & U_T'\left(\beta_\perp'\widehat{S}_{1,1}\beta_\perp\right) = \left(\widetilde{\alpha}'\widehat{\Omega}^{-1}\alpha\right)^{-1}\widetilde{\alpha}'\widehat{\Omega}^{-1}\left(\widehat{S}_{0,1} - \alpha\beta'\widehat{S}_{1,1}\right)\beta_\perp \\ & - \left(\widetilde{\alpha}'\widehat{\Omega}^{-1}\alpha\right)^{-1}\widetilde{\alpha}'\widehat{\Omega}^{-1}(\widetilde{\alpha}-\alpha)\,TU_T'\left(T^{-1}\beta_\perp'\widehat{S}_{1,1}\beta_\perp\right) \\ & - \left(\widetilde{\alpha}'\widehat{\Omega}^{-1}\alpha\right)^{-1}\widetilde{\alpha}'\widehat{\Omega}^{-1}(\widetilde{\alpha}-\alpha)\beta'\widehat{S}_{1,1}\beta_\perp \end{aligned}$$

Since by (78), $\beta'\widehat{S}_{1,1}\beta_\perp = O_p(1)$, it follows from Lemma 9 that the last two terms are of order $o_p(1)$. Moreover, $\widehat{S}_{0,1}\beta_\perp - \alpha\beta'\widehat{S}_{1,1}\beta_\perp = O_p(1)$ and $\left(\widetilde{\alpha}'\widehat{\Omega}^{-1}\alpha\right)^{-1}\widetilde{\alpha}'\widehat{\Omega}^{-1} = \left(\alpha'\Omega^{-1}\alpha\right)^{-1}\alpha'\Omega^{-1} + o_p(1)$. Thus

$$\begin{aligned} T.U_T &= \left(T^{-1}\beta_\perp'\widehat{S}_{1,1}\beta_\perp\right)^{-1}\beta_\perp'\left(\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha'\right)\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1} \\ &\quad + o_p(1) \end{aligned}$$

If follows from Lemma 2, Lemma 6 and Lemma A.2 below that in the case of VECM (37),

$$T.U_T \Rightarrow \left(\int_0^1 B_{k-r}B'_{k-r}\right)^{-1} \int_0^1 B_{k-r}dB'_\alpha \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2}$$

where

$$B_\alpha = \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \alpha'\Omega^{-1}C_0 B$$

is an r variate standard Brownian motion, which is independent of $B_{k-r} = (\alpha'_\perp\Omega\alpha_\perp)^{-1/2}\alpha'_\perp C_0 B$, whereas in the case of VECM (29),

$$\begin{aligned} T.U_T &\Rightarrow \left(\int_0^1 \overline{B}_{k-r}\overline{B}'_{k-r}\right)^{-1} \left(\int_0^1 B_{k-r}(x)dB_\alpha(x)' - \int_0^1 B_{k-r}(x)dxB_\alpha(1)'\right) \\ &\quad \times \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \\ &= \left(\int_0^1 \overline{B}_{k-r}\overline{B}'_{k-r}\right)^{-1} \left(\int_0^1 \overline{B}_{k-r}dB'_\alpha\right) \left(\alpha'\Omega^{-1}\alpha\right)^{-1/2} \end{aligned}$$

where $\overline{B}_{k-r}$ is defined in Theorem 5 and the equality follows from

$$\begin{aligned} \int_0^1 \overline{B}_{k-r}dB'_\alpha &= \int_0^1 \left(B_{k-r}(x) - \int_0^1 B_{k-r}(y)dy\right) dB_\alpha(x)' \\ &= \int_0^1 B_{k-r}(x)dB_\alpha(x)' - \int_0^1 B_{k-r}(y)dy \int_0^1 dB_\alpha(x)' \\ &= \int_0^1 B_{k-r}(x)dB_\alpha(x)' - \int_0^1 B_{k-r}(y)dyB_\alpha(1)' \end{aligned}$$

Theorem 7 now follows from (104).

Lemma A.2. *Under Assumptions 1-2,*

$$\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha' \Rightarrow C(1)\left(\int_0^1 BdB'\right)C'_0 \tag{126}$$

in the case of VECM (37), *and*

$$\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha' \Rightarrow C(1)\left(\int_0^1 BdB' - \int_0^1 B(x)dxB(1)'\right)C'_0 \tag{127}$$

in the case of VECM (29).

Proof: First, consider the case of VECM (37): $x_t = \alpha\beta' y_{t-1} + \Pi X_{t-1} + C_0 u_t$.

Recall that $\widehat{S}_{0,1} = \frac{1}{T}\sum_{t=1}^{T} R_{0,t}R'_{1,t}$ and $\widehat{S}_{1,1} = \frac{1}{T}\sum_{t=1}^{T} R_{1,t}R'_{1,t}$, so that

$$\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha' = \frac{1}{T}\sum_{t=1}^{T} R_{1,t}\left(R'_{0,t} - R'_{1,t}\beta\alpha'\right),$$

where $R_{1,t}$ is the residual of the regression of y_{t-1} on X_{t-1}: $R_{1,t} = y_{t-1} - \widehat{\Delta}X_{t-1}$, with

$$\widehat{\Delta} = \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}X'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1} = O_p(1)$$

and $R_{0,t}$ is the residual of the regression of x_t on X_{t-1}: $R_{0,t} = x_t - \widehat{\Upsilon}X_{t-1}$, with

$$\widehat{\Upsilon} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t X'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1} = O_p(1)$$

Hence, $R_{0,t} - \alpha\beta' R_{1,t}$ is the residual of the regression of $x_t - \alpha\beta' y_{t-1} = \Pi X_{t-1} + C_0 u_t$ on X_{t-1}. But the latter regression has the same residual as the regression of $C_0 u_t$ on X_{t-1}:

$$R_{0,t} - \alpha\beta' R_{1,t} = C_0 u_t - \widehat{\Gamma}X_{t-1}$$

where

$$\widehat{\Gamma} = C_0\left(\frac{1}{T}\sum_{t=1}^{T} u_t X'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X'_{t-1}\right)^{-1} = O_p\left(T^{-1/2}\right) \tag{128}$$

Thus

$$\begin{aligned}
\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha' &= \frac{1}{T}\sum_{t=1}^{T}\left(y_{t-1} - \widehat{\Delta}X_{t-1}\right)\left(C_0 u_t - \widehat{\Gamma}X_{t-1}\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T} y_{t-1}u'_t C'_0 - T^{-1/2}\widehat{\Delta}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}u'_t\right)C'_0
\end{aligned}$$

$$
\begin{aligned}
& -\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}X_{t-1}'\right)\widehat{\Gamma}' + \widehat{\Delta}\left(\frac{1}{T}\sum_{t=1}^{T} X_{t-1}X_{t-1}'\right)\widehat{\Gamma}' \\
= & \frac{1}{T}\sum_{t=1}^{T} y_{t-1}u_t'C_0' + O_p\left(T^{-1/2}\right)
\end{aligned}
$$

where the $O_p\left(T^{-1/2}\right)$ is due to (128), $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}u_t' = O_p(1)$, Theorem 3 and Lemma 3. It follows now from Theorem 2 that $\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha' \Rightarrow C(1)\left(\int_0^1 BdB'\right)C_0'$.

Next, consider the case of VECM (29) with $\pi_0 \neq 0$.

Recall that this model can be written as $x_t = \alpha\beta' y_{t-1} + \Pi_* \widetilde{X}_{t-1} + C_0 u_t$ where $\Pi_* = (\pi_0, \Pi)$, $\widetilde{X}_{t-1} = \left(1, X_{t-1}'\right)'$. Then $R_{1,t} = y_{t-1} - \widetilde{\Delta}\widetilde{X}_{t-1}$, where

$$
\begin{aligned}
\widetilde{\Delta} &= \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\widetilde{X}_{t-1}'\right)\left(\frac{1}{T}\sum_{t=1}^{T} \widetilde{X}_{t-1}\widetilde{X}_{t-1}'\right)^{-1} \\
&= \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}, \frac{1}{T}\sum_{t=1}^{T} y_{t-1}X_{t-1}'\right)\begin{pmatrix} 1 & \overline{X}_{-1}' \\ \overline{X}_{-1} & \widehat{\Sigma}_{XX} \end{pmatrix}^{-1}
\end{aligned}
$$

where

$$
\overline{X}_{-1} = \frac{1}{T}\sum_{t=1}^{T} X_{t-1},\ \widehat{\Sigma}_{XX} = \frac{1}{T}\sum_{t=1}^{T} X_{t-1}X_{t-1}',
$$

It is a standard linear algebra exercise to verify that

$$
\begin{pmatrix} 1 & \overline{X}_{-1}' \\ \overline{X}_{-1} & \widehat{\Sigma}_{XX} \end{pmatrix}^{-1} = \begin{pmatrix} \widehat{\sigma} & -\widehat{\sigma}\overline{X}_{-1}'\widehat{\Sigma}_{XX}^{-1} \\ -\widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1} & \widehat{\Sigma}_{XX}^{-1} + \widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}\overline{X}_{-1}'\widehat{\Sigma}_{XX}^{-1} \end{pmatrix}
$$

where

$$
\widehat{\sigma} = \left(1 - \overline{X}_{-1}'\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}\right)^{-1}
$$

hence

$$
\begin{aligned}
\widetilde{\Delta} = & \left(\widehat{\sigma}\frac{1}{T}\sum_{t=1}^{T} y_{t-1} - \widehat{\sigma}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}X_{t-1}'\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1},\right. \\
& \left. -\widehat{\sigma}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\left(\overline{X}_{-1}'\widehat{\Sigma}_{XX}^{-1}\right) + \frac{1}{T}\sum_{t=1}^{T} y_{t-1}X_{t-1}'\left(\widehat{\Sigma}_{XX}^{-1} + \widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}\overline{X}_{-1}'\widehat{\Sigma}_{XX}^{-1}\right)\right) \\
= & \sqrt{T}\widetilde{\Delta}_1 + \widetilde{\Delta}_2 \qquad (129)
\end{aligned}
$$

where

$$\widetilde{\Delta}_1 = \left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\right)\left(\widehat{\sigma}, -\widehat{\sigma}\overline{X}'_{-1}\widehat{\Sigma}_{XX}^{-1}\right)$$
$$\widetilde{\Delta}_2 = \left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}X'_{t-1}\right)\left(-\widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}, \widehat{\Sigma}_{XX}^{-1} + \widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}\overline{X}'_{-1}\widehat{\Sigma}_{XX}^{-1}\right)$$

Moreover, $R_{0,t} - \alpha\beta' R_{1,t}$ is now the residual of the regression of $C_0 u_t$ on $\widetilde{X}_{t-1}$: $R_{0,t} - \alpha\beta' R_{1,t} = C_0 u_t - \widetilde{\Gamma} X_{t-1}$, where similar to (129),

$$\begin{aligned}\widetilde{\Gamma} &= C_0\left(\frac{1}{T}\sum_{t=1}^{T} u_t\widetilde{X}'_{t-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)^{-1} \\ &= T^{-1/2}\widetilde{\Gamma}_1 + T^{-1/2}\widetilde{\Gamma}_2\end{aligned}$$

with

$$\widetilde{\Gamma}_1 = C_0\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right)\left(\widehat{\sigma}, -\widehat{\sigma}\overline{X}'_{-1}\widehat{\Sigma}_{XX}^{-1}\right)$$
$$\widetilde{\Gamma}_2 = C_0\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t X'_{t-1}\right)\left(-\widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}, \widehat{\Sigma}_{XX}^{-1} + \widehat{\sigma}\widehat{\Sigma}_{XX}^{-1}\overline{X}_{-1}\overline{X}'_{-1}\widehat{\Sigma}_{XX}^{-1}\right)$$

Note that by Assumption 1 and Lemma 3,

$$\begin{aligned}p\lim_{T\to\infty}\overline{X}_{-1} &= E\left[X_{t-1}\right] = 0 \\ p\lim_{T\to\infty}\widehat{\Sigma}_{XX} &= E\left[X_{t-1}X'_{t-1}\right] = \Sigma_{XX} \\ p\lim_{T\to\infty}\widehat{\sigma} &= 1\end{aligned}$$

and recall that

$$\begin{aligned}\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{t-1} &\Rightarrow C(1)\int_0^1 B(x)dx^9 \\ \frac{1}{T}\sum_{t=1}^{T} y_{t-1}X'_{t-1} &= O_p(1) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t &\Rightarrow B(1)\end{aligned}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t X'_{t-1} \quad = \quad O_p(1)$$

so that

$$\begin{aligned}
\widetilde{\Delta}_1 &= \left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\right)\left(1, O_{1,(p-1)k}\right) + o_p(1) \Rightarrow C(1)\int_0^1 B(x)dx \\
\widetilde{\Delta}_2 &= O_p(1) \\
\widetilde{\Gamma}_1 &= C_0\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t\right)\left(1, O_{1,(p-1)k}\right) + o_p(1) \Rightarrow C_0 B(1) \\
\widetilde{\Gamma}_2 &= C_0\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t X'_{t-1}\right)\begin{pmatrix} 1 & O_{1,k} \\ O_{k,1} & \Sigma_{XX}^{-1} \end{pmatrix}(O_{1,k}, I_k) + o_p(1)
\end{aligned}$$

It follows from the previous part that

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}u'_t \Rightarrow C(1)\int_0^1 BdB'$$

Moreover, it is not too hard to verify that

$$\begin{aligned}
\widetilde{\Delta}_1\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{X}_{t-1}u'_t\right)C'_0 &\Rightarrow C(1)\int_0^1 B(x)dxB(1)'C'_0, \\
\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\widetilde{X}'_{t-1}\right)\widetilde{\Gamma}'_1 &\Rightarrow C(1)\int_0^1 B(x)dxB(1)'C'_0 \\
\widetilde{\Delta}_1\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)\widetilde{\Gamma}'_1 &\Rightarrow C(1)\int_0^1 B(x)dxB(1)'C'_0 \\
\widetilde{\Delta}_1\left(\frac{1}{T}\sum_{t=1}^{T}\widetilde{X}_{t-1}\widetilde{X}'_{t-1}\right)\widetilde{\Gamma}'_2 &= o_p(1) \\
\left(\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\widetilde{X}'_{t-1}\right)\widetilde{\Gamma}'_2 &= o_p(1)
\end{aligned}$$

It follows therefore straightforwardly that

$$\widehat{S}_{1,0} - \widehat{S}_{1,1}\beta\alpha' \quad = \quad \frac{1}{T}\sum_{t=1}^{T}\left(y_{t-1} - T^{1/2}\widetilde{\Delta}_1\widetilde{X}_{t-1} - \widetilde{\Delta}_2\widetilde{X}_{t-1}\right)$$

$$\times \left(u_t' C_0' - T^{-1/2}\widetilde{X}_{t-1}'\widetilde{\Gamma}_1' - T^{-1/2}\widetilde{X}_{t-1}'\widetilde{\Gamma}_2' \right)$$
$$\Rightarrow \quad C(1) \left(\int_0^1 B dB' C_0' - \int_0^1 B(x) dx B(1)' \right) C_0'$$

8 References

Anderson, S. A., H. K. Brons and S. T. Jensen (1983), "Distribution of Eigenvalues in Multivariate Statistical Analysis", *Annals of Statistics* 11, 392-415.

Bierens, H. J. (1994), *Topics in Advanced Econometrics: Estimation, Testing and Specification of Cross-Section and Time Series Models,* Cambridge University Press.

Bierens, H. J. (2004), *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press.

Engle, R. F. and C. W. J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing", *Econometrica* 55, 251-276.

Johansen, S. (1988), "Statistical Analysis of Cointegrated Vectors", *Journal of Economic Dynamics and Control* 12, 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrated Vectors in Gaussian Vector Autoregressive Models", *Econometrica* 59, 1551-1580.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Phillips, P. C. B. (1986), "Time Series Regression with a Unit Root", *Econometrica* 55, 277-301.

Phillips, P. C. B. (1988), "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 B dB'$", *Journal of Multivariate Time Series Analysis* 24, 252-264.