# Implementing Pesaran-Shin-Smith:

## *based on Pesaran et al. (2000): Structural analysis of vector error correction models with exogenous I(1) variables*


Stefan Zeugner*

*zeugner@ihs.ac.at*

Supervisor:
Martin Wagner
*mawagner@ihs.ac.at*

Institute for Advanced Studies
Stumpergasse 56
A-1060 Vienna


October 18, 2006


**Abstract**

This first year paper is based on Pesaran et al. (2000) who generalise the cointegration tests introduced by Johansen to include exogenous I(1) variables in a VECM model. It reiterates the proofs for their central test statistics and presents them in a less dense format: Following Pesaran et al. (2000), this paper focuses on the derivation of the corresponding cointegrating rank tests, by first introducing a VAR model, subsequently deriving the likelihood for the cointegration parameters and, finally, the test statistics and their asymptotic distributions. The final section introduces tests on whether the required exogeneity restrictions hold. In addition, this paper is concerned with implementing the mentioned test statistics in a MATLAB routine. The respective outlines on implementation and usage are in the appendix.


*DISCLAIMER: This paper is a reformulation of several proofs in Pesaran et al. (2000) in an extended and consolidated manner for purely pedagogical purposes. Any academic contribution stems from Pesaran et al. (2000) or other indicated references.  The author assumes no responsibility for any consequence arising of his potential errors or misinterpretations.

# 1 Introduction

This paper is entirely based on Pesaran et al. (2000). As the title suggests, Pesaran et al. (2000) are primarily concerned with extending the ubiquitous Johansen test for cointegration (Johansen, 1991) to the inclusion of exogenous variables: a sub-system of I(1) variables is assumed as *structurally exogenous* as defined by Pesaran et al. (2000, p.294):

> *[...] structurally exogenous; that is, any cointegrating vectors present do not appear in the sub-system vector error correction model (VECM) for these exogenous variables and the error terms in this sub-system are uncorrelated with those in the rest of the system.*

This type of test provides a useful extension of the cointegrating rank tests proposed by Johansen (1991). Its usage in applied econometric studies, however, suffers from two minor drawbacks: First, the tests introduced by Pesaran et al. (2000) are, to our knowledge, not yet implemented in standard software packages. Second, the derivation of their test statistics is quite dense, which may have deterred readers with a less theoretic background.

Consequently, the motivation for this first year paper is twofold: Firstly, it aims at reiterating the derivation of the centrepiece in Pesaran et al. (2000), the test statistic for cointegration rank, but presenting the corresponding proofs in a consolidated fashion and thus facilitating their comprehension to readers without particular knowledge of Johansen-type tests. Secondly, this paper is concerned with implementing the major test statistics of Pesaran et al. (2000) in Matlab.

The latter objective has materialised in a routine available from the author; its implementation is described in Appendix A.1. For a quick reference on using this routine, please refer to Appendix A.2. The main body of this paper, though, is concerned with the former objective: Restating Pesaran et al. (2000) in a consolidated fashion. In line with this motivation, we consequently prefer to state even small proofs in the main text, rather than banning them to the appendix. The article by Pesaran et al. (2000) broadly follows the structure given in Johansen (1991), which is itself an extension of the earlier Johansen (1988).

Likewise, we will start with specifying a VAR($p$) structure embodying both I(1) and I(0) processes and spend section 2 on transforming its structure into a VMA($\infty$) model. Just as in Pesaran et al. (2000), the following section (3) examines the derivation of the likelihood function for the cointegrating vector(s). Section 4 slightly alters Pesaran et al. (2000)'s structure by only presenting the likelihood ratio tests for cointegrating ranks as well as Theorems 2 and 3 specifying their limiting distributions.[1] Section 5 thereafter provides the proofs for Theorems 2 and 3. Especially during the latter two sections, this paper provides the proofs in much more detail than Pesaran et al. (2000), where the corresponding proofs are presented in an extremely dense format. Finally, our paper concludes with introducing diagnostic tests for the restrictions on the exogenous variables (proofs omitted): I.e. whether the exogenous data are not cointegrated, and whether there is no impact of the endogenous on the exogenous variables.

[1] Pesaran et al. (2000) proceed differently by presenting the theorems for all limit distributions together, while their proofs are in the appendix.

## 2 The VAR model, assumptions and deterministics

We start with outlining the VAR model specification for the vector process $\{z_t\}_{t=-\infty}^{\infty}$, as in Pesaran et al. (2000, p.295). Section 2 will follow their paper's section 2 closely, which draws heavily on section 4 in Johansen (1991, pp.1558-61) - in contrast, we will elaborate more on the intermediate steps of, and retain the notation of, Pesaran et al. (2000) in order to facilitate comprehension and comparison. The basic idea is to transform the VAR model in (2), respectively its VAR($p$) expression in levels into a VMA($\infty$) expression in the differences for later use in the derivation of the test statistic. This is done via an appropriate reformulation such that Granger's representation theorem can be applied (cf. Johansen (1991, p.1559)).
Let $\{z_t\}_{t=-\infty}^{\infty}$ denote an $m\times 1$-dimensional vector random process. Moreover, let it be generated by the vector autoregressive model of order $p$ (VAR($p$)):

$$\mathbf{\Phi}(L)(z_t - \boldsymbol{\mu} - \boldsymbol{\gamma} t) = e_t \quad , \qquad t = 1, 2, \ldots \tag{1}$$

Here, $L$ is the lag or backshift operator, $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ are fixed $m\times 1$ vectors and $\mathbf{\Phi}(L)$ is the $m\times m$ lag polynomial of order $p$ defined as $\mathbf{\Phi}(L) = I_m - \sum_{i=1}^{p} \mathbf{\Phi}_i L^i$, where $\mathbf{\Phi}_i$ are unknown $m\times m$ coefficient matrices. Moreover the initial values $(z_0, \ldots, z_{0-p+1})$ are assumed to be given. The assumptions on $\{e_t\}$ are stated below:

Assumption 1 *The error process $\{e_t\}_{t=-\infty}^{\infty}$ is assumed to be identically normally distributed $IN(\mathbf{0}, \mathbf{\Omega})$ for all $t$, with $\Omega$ positive definite.*

The lag polynomial $\mathbf{\Phi}(L)$ may be expressed as follows:

$$\mathbf{\Phi}(L) = I_m - \sum_{i=1}^{p} \mathbf{\Phi}_i L^i = \underbrace{\left(\mathbf{I}_m - \sum_{i=1}^{p} \mathbf{\Phi}_i\right)}_{=\mathbf{\Phi}(1)\equiv -\mathbf{\Pi}} L + \underbrace{\left(\mathbf{I}_m - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i L^i\right)}_{\equiv \mathbf{\Gamma}(L)} (1 - L) \tag{2}$$

where $\mathbf{\Gamma}_i = -\sum_{j=i+1}^{p} \mathbf{\Phi}_j$.

Proof of (2):

$$-\mathbf{\Pi} L + \mathbf{\Gamma}(L)(1-L) = \mathbf{I}_m L \underbrace{- \sum_{i=1}^{p} \mathbf{\Phi}_i L}_{+\mathbf{\Gamma}_0 L} + \mathbf{I}_m - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i L^i - \mathbf{I}_m L + \sum_{i=1}^{p-1} \mathbf{\Gamma}_i L^{i+1} =$$

$$\mathbf{I} + \sum_{i=1}^{p} \mathbf{\Gamma}_{i-1} L^i - \sum_{i=1}^{p-1} \mathbf{\Gamma}_i L^i = \mathbf{I}_m + \sum_{i=1}^{p} (\mathbf{\Gamma}_{i-1} - \mathbf{\Gamma}_i) L^i = \mathbf{I}_m - \sum_{i=1}^{p} \mathbf{\Phi}_i L^i = \mathbf{\Phi}(L)$$

$$\text{since } \mathbf{\Gamma}_{i-1} - \mathbf{\Gamma}_i = -\sum_{j=i}^{p} \mathbf{\Phi}_j + \sum_{j=i+1}^{p} \mathbf{\Phi}_j = -\mathbf{\Phi}_i$$

thus we may re-express (1) as follows:

$$\mathbf{\Phi}(L) z_t = \mathbf{a}_0 + \mathbf{a}_1 t + e_t \tag{3}$$

where

$$\mathbf{a_0} \equiv -\boldsymbol{\Pi}\boldsymbol{\mu} + (\boldsymbol{\Gamma} + \boldsymbol{\Pi})\boldsymbol{\gamma} \quad , \qquad \mathbf{a_1} \equiv -\boldsymbol{\Pi}\boldsymbol{\gamma} \tag{4}$$

and $\boldsymbol{\Gamma} \equiv \boldsymbol{\Gamma}(1)$. The result follows since $L\boldsymbol{\mu} = \boldsymbol{\mu}$ and $L\boldsymbol{\gamma}t = \boldsymbol{\gamma}(t-1)$. Moreover note that $\boldsymbol{\Gamma}(L)\boldsymbol{\gamma} = \boldsymbol{\gamma} - \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i L^i \boldsymbol{\gamma} = \boldsymbol{\Gamma}(1)\boldsymbol{\gamma}$.[2]

Proof of (3):

$$\begin{aligned}
\boldsymbol{\Phi}(L)z_t &= \boldsymbol{\Phi}(L)\left(\boldsymbol{\mu} + \boldsymbol{\gamma}\right) + e_t = -\boldsymbol{\Pi}\boldsymbol{\mu} + \boldsymbol{\Gamma}(L)(1-L)\boldsymbol{\gamma}t - \boldsymbol{\Pi}L\boldsymbol{\gamma}t + e_t = \\
&= -\boldsymbol{\Pi}\boldsymbol{\mu} + \boldsymbol{\Gamma}(L)\boldsymbol{\gamma}t - \boldsymbol{\Gamma}(L)\boldsymbol{\gamma}(t-1) - \boldsymbol{\Pi}\boldsymbol{\gamma}(t-1) + e_t = \\
&= -\boldsymbol{\Pi}\boldsymbol{\mu} + \underbrace{\boldsymbol{\Gamma}(L)\boldsymbol{\gamma}}_{=\boldsymbol{\Gamma}\boldsymbol{\gamma}} + \boldsymbol{\Pi}\boldsymbol{\gamma} - \boldsymbol{\Pi}\boldsymbol{\gamma}t = \mathbf{a_0} + \mathbf{a_1}t + e_t
\end{aligned}$$

Furthermore, consider the identity $\boldsymbol{\Gamma} \equiv \boldsymbol{\Gamma}(1) = -\boldsymbol{\Pi} + \sum_{i=1}^{p} i\boldsymbol{\Phi}_i$:

$$\begin{aligned}
\boldsymbol{\Gamma}(1) &= \mathbf{I}_m + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} \boldsymbol{\Phi}_j = \mathbf{I}_m + (p-1)\boldsymbol{\Phi}_p + \ldots + 1\boldsymbol{\Phi}_2 \\
&= \mathbf{I}_m + \sum_{i=2}^{p} (i-1)\boldsymbol{\Phi}_i + \underbrace{\left(-\mathbf{I}_m + \sum_{i=1}^{p} \boldsymbol{\Phi}_i - \boldsymbol{\Pi}\right)}_{0} = -\boldsymbol{\Pi} + \sum_{i=1}^{p} i\boldsymbol{\Phi}_i
\end{aligned}$$

Finally, we may use (3) to express $\Delta z_t = (1-L)z_t$ as follows:

$$\Delta z_t = \boldsymbol{\Pi} z_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i \Delta z_{t-i} + \mathbf{a}_0 + \mathbf{a}_1 t + e_t \tag{5}$$

Now consider the hypothesis on the rank of $\boldsymbol{\Pi}$:

Hypothesis $H_r^m : \text{rk}(\boldsymbol{\Pi}) = r \qquad r = 0, 1, \ldots, m$

Under $H_r^m$ we may express $\boldsymbol{\Pi}$ as the composition of two full-rank $m \times r$ matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\underset{m \times r}{\boldsymbol{\Pi}} = \underset{m \times r}{\boldsymbol{\alpha}} \ \underset{r \times m}{\boldsymbol{\beta}'} \tag{6}$$

In this case one may define the $m \times (m-r)$ matrices $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ whose columns form the basis of the kernel of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$: viz. $\boldsymbol{\alpha}'_\perp \boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta}'_\perp \boldsymbol{\beta} = \mathbf{0}$. While retaining hypotheses $H_r^m$, we adopt the following two assumptions needed for the application of Theorem 1:

Assumption **2** *The roots $z$ of $|\boldsymbol{\Phi}(z)| = 0$ are one or outside the unit circle: either $|z| > 1$ or $z = 1$.*

Assumption **3** *The $(m-r) \times (m-r)$ matrix $\boldsymbol{\alpha}'_\perp \boldsymbol{\Gamma} \boldsymbol{\beta}_\perp$ has full rank $m - r$.*

[2] Actually, any constant $\zeta$ is shorthand notation for $\zeta 1$ where 1 is the $t$-th element of an infinite vector $\mathbf{1}$ with all elements equal to 1. So $L\zeta\mathbf{1}_t = \zeta\mathbf{1}_{t-1} = \zeta 1$.

Assumptions 2 and 3 enable us to invoke Granger's representation theorem as it is used in Johansen (1991, p.1559) in order to transform system (1) into a VMA($\infty$) representation. Theorem 1 below reiterates Theorem 4.1 in Johansen (1991, p.1559) adjusted for our particular needs.

THEOREM 1 *Under $H_r^m$ with $\mathbf{\Pi} = \boldsymbol{\alpha\beta}'$, $\mathrm{rk}(\boldsymbol{\alpha}) = \mathrm{rk}(\boldsymbol{\beta}) = r$ and under assumptions 2 and 3, $\Delta z_t$ and $\boldsymbol{\beta} z_t$ can be given initial distributions such that*

- $\Delta z_t - E(\Delta z_t)$ *is stationary* $\Rightarrow$ $z_t$ *is* $I(1)$
- $\boldsymbol{\beta}' z_t - E(\boldsymbol{\beta}' z_t)$ *is stationary*

*Furthermore, if the initial distributions are expressed in terms of $\{e_t\}_{t=-\infty}^{\infty}$ then $\Delta z_t$ has a representation:*

- $\Delta z_t = C(L)(a_0 + a_1 t + e_t)$

*where* $C(1) \equiv C = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}'_\perp \mathbf{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}'_\perp$
*and $C(L)$ may be expressed as follows:* $C(L) = \mathbf{I}_m + \sum\limits_{j=1}^{\infty} C_j L^j$

The proof of Theorem 1[3] is one of the few omitted here since it largely parallels the one of Theorem 4.1 in Johansen (1991, pp.1159-61), only with adjusted deterministic terms.[4] Theorem 1 implies $C(L)\mathbf{\Phi}(L) = (1-L)\mathbf{I}_m$, and, by setting $C^*(L) = \frac{C(L)-C}{(1-L)}$:

$$C(L) = \mathbf{I}_m + \sum_{j=1}^{\infty} C_j L^j = C + (1-L)C^*(L) \tag{7}$$

Equation (7) provides a decomposition into level and difference terms similar to (2). One may as well express $C^*(L)$ as an infinite sum $C^*(L) = \sum\limits_{j=0}^{\infty} C_j^* L^j$ and denote $C^* \equiv C^*(1)$.
By applying (7), we get

$$\begin{aligned}\Delta z_t &= C(L)(\mathbf{a}_0 + \mathbf{a}_1 t + e_t) = C\mathbf{a}_0 + \overbrace{C^*(L)(1-L)\mathbf{a}_0}^{=C^*(L)\mathbf{0}=\mathbf{0}} + C\mathbf{a}_1 t + \overbrace{C^*(L)\mathbf{a}_1 t - C^*(L)\mathbf{a}_1(t-1)}^{=C^*\mathbf{a}_1} + \\ &\quad + C(L)e_t = \underbrace{C\mathbf{a}_0 + C^*\mathbf{a}_1}_{\equiv b_0} + \underbrace{C\mathbf{a}_1}_{\equiv b_1} t + C(L)e_t = \mathbf{b}_0 + \mathbf{b}_1 t + C(L)e_t = \\ &\quad = \mathbf{b}_0 + \mathbf{b}_1 t + Ce_t + C^*(L)\Delta e_t \end{aligned} \tag{8}$$

Adding up (8) yields

$$\begin{aligned} z_t &= \sum_{i=1}^{t} \Delta z_i = t\mathbf{b}_0 + \mathbf{b}_1 \frac{t(t+1)}{2} + C\sum_{i=1}^{t} e_i + \sum_{i=1}^{t} (C^*(L)e_i - C^*(L)e_{i-1}) + z_0 = \\ &= t\mathbf{b}_0 + \mathbf{b}_1 \frac{t(t+1)}{2} + C\sum_{i=1}^{t} e_i + C^*(L)(e_t - e_0) + z_0 \end{aligned} \tag{9}$$

[3] The basic idea of the proof is as follows: since $\mathrm{rk}(\mathbf{\Pi}) < m$, the AR representation is not directly invertible. One proceeds by projecting on the spaces spanned by $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_\perp$, respectively, and thus obtaining two invertible subsystems in the differences and in the levels; the invertibility of the latter is guaranteed by Assumption 3. The subsequent rearrangement in terms of $\Delta z_t$ allows for $C = \boldsymbol{\beta}_\perp \left(\boldsymbol{\alpha}'_\perp \mathbf{\Gamma} \boldsymbol{\beta}_\perp\right)^{-1} \boldsymbol{\alpha}'_\perp$.

[4] The reader may furthermore want to compare the proof of Theorem 4.2 in Johansen (1995).

Expression (9) simplifies if we take into account the restrictions in (4): Since $C(z)\mathbf{\Phi}(z) = \mathbf{I}_m(1-z)$, we have $C(1)\mathbf{\Phi}(1) = -C\mathbf{\Pi} = \mathbf{0}$; Hence we obtain

$$\mathbf{b}_1 = C\mathbf{a}_1 = C(-\mathbf{\Pi}\boldsymbol{\gamma}) = \mathbf{0} \tag{10}$$

Moreover, noting that $C\mathbf{\Gamma} - C^*\mathbf{\Pi} = \mathbf{I}_m$,[5] we may simplify $\mathbf{b}_0$ as follows:

$$\begin{aligned}\mathbf{b}_0 &= C\mathbf{a}_0 + C^*\mathbf{a}_1 = C\left(-\mathbf{\Pi}\boldsymbol{\mu} + (\mathbf{\Gamma}+\mathbf{\Pi})\boldsymbol{\gamma}\right) + C^*(-\mathbf{\Pi}\boldsymbol{\gamma}) = \\ &= -\underbrace{C\mathbf{\Pi}\boldsymbol{\mu}}_{\mathbf{0}} + \underbrace{(C\mathbf{\Gamma} - C^*\mathbf{\Pi})}_{\mathbf{I}_m}\boldsymbol{\gamma} + C\mathbf{\Pi}\boldsymbol{\gamma} = \boldsymbol{\gamma}\end{aligned} \tag{11}$$

Finally, initialise $\boldsymbol{\mu} = z_0 - C^*(L)e_0$ to obtain from (9)

$$z_t = \boldsymbol{\mu} + \boldsymbol{\gamma}t + C\sum_{i=1}^{t} e_i + C^*(L)e_t \tag{12}$$

Thus the restrictions (4) ensure that,in its VMA($\infty$) form, the long-run deterministic trending behaviour of $\{z_t\}_{t=-\infty}^{\infty}$ is invariant to the rank of $\mathbf{\Pi}$. If, however, the restriction $\mathbf{a}_1 = -\mathbf{\Pi}\boldsymbol{\gamma}$ did not hold, $\mathbf{b}_1$ would in general be different from zero, and therefore a quadratic trend term would be present in $z_t$ (cf. Pesaran et al. (2000, p.298)); except in the case of rk($\mathbf{\Pi}$) $= m$, which implies $C = \mathbf{0}$. In particular it follows that under $H_r^m$ there would be $m-r$ independent quadratic time-trends in equation (12). Hence under differing values of cointegrating rank $r$, quite different trend patterns should be observed in the levels process $\{z_t\}_{t=-\infty}^{\infty}$ (Ibid.). For the further derivation in sections 4 and 5, this notion is of particular relevance.
In addition, note that (12) implies an inclusion of trending behaviour in the (trend-)stationary[6] term $\boldsymbol{\beta}'z_t$:

$$\boldsymbol{\beta}'z_t = \boldsymbol{\beta}'\boldsymbol{\mu} + (\boldsymbol{\beta}'\boldsymbol{\gamma})t + \boldsymbol{\beta}'C^*(L)e_t \tag{13}$$

Thus in general, $\boldsymbol{\beta}'\boldsymbol{\gamma}$, respectively $\mathbf{a}_1$ in (3) has to be estimated along with the cointegrating regression. Imposing a co-trending restriction, i.e. excluding trending behaviour from the cointegrating regression requires setting $\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$ (if and only if $\mathbf{a}_1 = \mathbf{0}$ [7] ). Imposing $\mathbf{a}_1 = \mathbf{0}$ would imply different asymptotic behaviour, which leads in turn to the "cases" known from the Johansen (1991) test.

# 3 Exogenous I(1) variables in the VECM log-likelihood

Noting that $\mathbf{\Phi}(L)$ may be expressed as in (2), we may rewrite (3) in the familiar VECM form

$$\Delta z_t = \mathbf{a}_0 + \mathbf{a}_1 t + \sum_{i=1}^{p-1}\mathbf{\Gamma}_i\Delta z_{t-i} + \mathbf{\Pi}z_{t-1} + e_t \tag{14}$$

In order to introduce exogeneity in the system, partition the $m\times 1$ vector $z_t$ into the $n \times 1$ vector $y_t$ and the $k\times 1$ vector $x_t$, and partition $e_t$ correspondingly into $e_{yt}$ and $e_{xt}$:

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} \quad , \qquad e_t = \begin{pmatrix} e_{yt} \\ e_{xt} \end{pmatrix} \tag{15}$$

[5] Proof that $C\mathbf{\Gamma} - C^*\mathbf{\Pi} = \mathbf{I}_m$: for $z \neq 1$ we know: $\mathbf{I}_m = \frac{C(z)}{(1-z)}\mathbf{\Phi}(z) = (\frac{C}{(1-z)} + C^*(z))(-\mathbf{\Pi}z + \mathbf{\Gamma}(z)(1-z)) = -\frac{C\mathbf{\Pi}z}{1-z} - C^*(z)\mathbf{\Pi}z + C\mathbf{\Gamma}(z) + C^*(z)\mathbf{\Gamma}(z)(1-z)$, where $C\mathbf{\Pi} = \mathbf{0}$ as above. Now, as $z \to 1$, we obtain: $C\mathbf{\Gamma}(1) - C^*(1)\mathbf{\Pi} = \mathbf{I}_m$

[6] "Stationary" in this case refers to $\boldsymbol{\beta}'z_t - E(\boldsymbol{\beta}'z_t)$ being weakly covariance stationary.

[7] under $H_r^m$, $\mathbf{a}_1 = \boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0} \Leftrightarrow \boldsymbol{\beta}'\boldsymbol{\gamma} = \mathbf{0}$, since $\boldsymbol{\alpha}$ has full rank $r$.

The covariance matrix of $e_t$ may be partitioned accordingly, where, by Assumption 1, $\Omega_{yy}$ and $\Omega_{xx}$ are positive definite:

$$\Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix} \tag{16}$$

Thus Assumption 1 enables re-expressing $e_t$ in terms of the independent processes $\{e_{xt}\}$ and $\{u_t\}$

$$u_t = e_{yt} - \Omega_{yx}\Omega_{xx}^{-1}e_{xt} \quad , \qquad u_t \sim \mathrm{IN}(\mathbf{0}, \Omega_{uu}) \tag{17}$$

Proof of (17): From the properties of normally distributed random variables we know the following: Let the $m \times 1$ random vector $e_t \sim \mathrm{N}(\mathbf{0}, \Omega)$ and $A$ be a fixed $n \times m$ matrix, while $B$ is a fixed $k \times m$ matrix: Then $Ae_t$ and $Be_t$ are independent normally distributed random variables if and only if $A\Omega B' = \mathbf{0}$. In our case, let

$$A = \begin{pmatrix} \mathbf{I}_n & -\Omega_{yx}\Omega_{xx}^{-1} \end{pmatrix} \quad , \qquad B = \begin{pmatrix} \mathbf{0} & \mathbf{I}_k \end{pmatrix}$$

Then we obtain $u_t \equiv Ae_t$, $e_{xt} = Be_t$ and verify $A\Omega B' = \mathbf{0}$ and $\mathrm{E}(u_t u_t') = A\mathrm{E}(e_t e_t')A' = \Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy}$.

Accordingly, partition $\mathbf{a}_0$, $\mathbf{a}_1$ into $n \times 1$ and $k \times 1$ vectors, and $\mathbf{\Gamma}$ and $\mathbf{\Pi}$ into $n \times m$ and $k \times m$ matrices:

$$\mathbf{\Pi} = \begin{pmatrix} \mathbf{\Pi}_y \\ \mathbf{\Pi}_x \end{pmatrix} \quad , \qquad \mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_y \\ \mathbf{\Gamma}_x \end{pmatrix} \quad , \qquad \mathbf{a}_0 = \begin{pmatrix} \mathbf{a}_{y0} \\ \mathbf{a}_{x0} \end{pmatrix} \quad , \qquad \mathbf{a}_1 = \begin{pmatrix} \mathbf{a}_{y1} \\ \mathbf{a}_{x1} \end{pmatrix}$$

Now substitute (17) into (14) and define $\mathbf{\Lambda} \equiv \Omega_{yx}\Omega_{xx}^{-1}$ for convenience

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \mathbf{a}_{y0} \\ \mathbf{a}_{x0} \end{pmatrix} + \begin{pmatrix} \mathbf{a}_{y1} \\ \mathbf{a}_{x1} \end{pmatrix} t + \sum_{i=1}^{p-1} \begin{pmatrix} \mathbf{\Gamma}_{yi} \\ \mathbf{\Gamma}_{xi} \end{pmatrix} \begin{pmatrix} \Delta y_{t-i} \\ \Delta x_{t-i} \end{pmatrix} + \begin{pmatrix} \mathbf{\Pi}_y \\ \mathbf{\Pi}_x \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{\Lambda} e_{xt} + u_t \\ e_{xt} \end{pmatrix} \tag{18}$$

Multiply from the left with $\begin{pmatrix} \mathbf{I}_n & -\mathbf{\Lambda} \end{pmatrix}$ in order to obtain

$$\begin{aligned} \Delta y_t - \mathbf{\Lambda}\Delta x_t = {} & \underbrace{\mathbf{a}_{y0} - \mathbf{\Lambda}\mathbf{a}_{x0}}_{\equiv \mathbf{c}_0} + \underbrace{\mathbf{a}_{y1} - \mathbf{\Lambda}\mathbf{a}_{x1}}_{\equiv \mathbf{c}_1} t + \sum_{i=1}^{p-1} \underbrace{(\mathbf{\Gamma}_{yi} - \mathbf{\Lambda}\mathbf{\Gamma}_{xi})}_{\equiv \boldsymbol{\psi}_i} \begin{pmatrix} \Delta y_{t-i} \\ \Delta x_{t-i} \end{pmatrix} + \\ & + \underbrace{(\mathbf{\Pi}_y - \mathbf{\Lambda}\mathbf{\Pi}_x)}_{\equiv \mathbf{\Pi}_{yy.x}} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \underbrace{\mathbf{\Lambda} e_{xt} + u_t - \mathbf{\Lambda} e_{xt}}_{= u_t} \end{aligned} \tag{19}$$

or, equivalently

$$\Delta y_t = \mathbf{c}_0 + \mathbf{c}_1 t + \mathbf{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \mathbf{\Pi}_{yy.x} z_{t-1} + u_t \tag{20}$$

Now we may state the following assumption:

Assumption 4 $\mathbf{\Pi}_x = \mathbf{0}$, *i.e.* $\{x_t\}_{t=1}^{\infty}$ *is weakly exogenous for* $\mathbf{\Pi}$

where we reiterate the definition of weak exogeneity with respect to our case:

Definition: A random process $\{x_t\}_{t=1}^{\infty}$ is weakly exogenous for the parameter vector $\Theta_y$ and its elements, if there exists a sequential cut (i.e. $(\Theta_y, \Theta_x) \in \Theta_y \times \Theta_x$) such that $f(y_t, x_t | \{x_i\}_{i=1}^{t-1}, \{y_i\}_{i=1}^{t-1}, \Theta_y, \Theta_x) = f(y_t | \{x_i\}_{i=1}^{t}, \{y_i\}_{i=1}^{t-1}, \Theta_y) \times f(x_t | \{x_i\}_{i=1}^{t-1}, \{y_i\}_{i=1}^{t-1}, \Theta_x)$; where $f(|)$ denotes the conditional density function.

Here, setting $\mathbf{\Pi}_x = \mathbf{0}$ is sufficient for this sequential cut, i.e. whatever the parameters for the conditional distributions of $\Delta y$ and $\Delta x$, the parameters of their joint distribution may be derived therefrom. Assumption 4 implies $\mathbf{\Pi}$ can be efficiently estimated without taking into account the law of $x_t$, since $\mathbf{\Pi}$ does not affect the long-run evolution of $x_t$. This implies firstly that $\mathbf{\Pi}_{yy.x} = \mathbf{\Pi}_y$ and, secondly, that the elements of $\{x_t\}_{t=1}^{\infty}$ are not cointegrated among themselves. Moreover, since $x_t$ does not depend on the levels of $y_t$, while vice versa this dependence exists, $x_t$ may be considered as *long-run forcing* for $y_t$ (Granger and Lin, 1995). Finally, Assumption 4 implies a reformulation of hypothesis $H_r^m$, since $r = \text{rk}(\mathbf{\Pi})$ can at most be $n < m$:

HYPOTHESIS $H_r : \text{rk}(\mathbf{\Pi}) = \text{rk}(\mathbf{\Pi}_y) = r \qquad r = 0, 1, \ldots, n$

Equivalently, Assumption 4 implies $\mathbf{a}_{x1} = -\mathbf{\Pi}_x\boldsymbol{\gamma} = \mathbf{0}$, a fact to be considered when substituting (20) into (18) [8]

$$\Delta y_t = \mathbf{c}_0 + \mathbf{c}_1 t + \mathbf{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \mathbf{\Pi}_y z_{t-1} + u_t \tag{21}$$

$$\Delta x_t = \mathbf{a}_{x0} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_{xi} \Delta z_{t-i} + e_{xt} \tag{22}$$

Moreover, since $\mathbf{a}_{x1} = \mathbf{0}$, the expression for $\mathbf{c}_1$ reduces to $\mathbf{c}_1 = \mathbf{a}_{y1} - \mathbf{\Lambda}\mathbf{a}_{x1} = \mathbf{a}_{y1}$. Hence the restrictions on $\mathbf{c}_0$ and $\mathbf{c}_1$ introduced in (3) modify to

$$\mathbf{c}_0 = \mathbf{a}_{y0} - \mathbf{\Lambda}\mathbf{a}_{x0} = -\mathbf{\Pi}_y\boldsymbol{\mu} + \left(\mathbf{\Gamma}_y - \Omega_{yx}\Omega_{xx}^{-1}\mathbf{\Gamma}_x + \mathbf{\Pi}_y\right)\boldsymbol{\gamma} \quad , \qquad \mathbf{c}_1 = \mathbf{a}_{y1} = -\mathbf{\Pi}\boldsymbol{\gamma} \tag{23}$$

Concerning whether or not the restrictions in (23) on $\mathbf{c}_0$ and $\mathbf{c}_1$ affect the further derivation of the test statistics, Pesaran et al. (2000) specify five cases, similar to the five cases of the Johansen test (Johansen, 1995, pp.211-212).[9]

## Hypotheses on deterministic terms - five cases

- *Case I: No intercepts, no trends*, i.e. $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\gamma} = \mathbf{0}$, which implies $\mathbf{c}_0 = \mathbf{0}$ and $\mathbf{c}_1 = \mathbf{0}$. The structural VECM (21) reduces to

$$\Delta y_t = \mathbf{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \mathbf{\Pi}_y z_{t-1} + u_t \tag{24}$$

- *Case II: restricted intercepts, no trends*, i.e. $\mathbf{c}_0$ as in (23) but $\boldsymbol{\gamma} = \mathbf{0}$, which implies $\mathbf{c}_1 = \mathbf{0}$. The structural VECM (21) then becomes

$$\Delta y_t = (-\mathbf{\Pi}_y\boldsymbol{\mu}) + \mathbf{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \mathbf{\Pi}_y z_{t-1} + u_t \tag{25}$$

[8] respectively multiply (18) from the left with $\begin{pmatrix} \mathbf{I}_n & -\mathbf{\Lambda} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}$

[9] Note that cases III and V differ from Johansen (1995, p.81, p.211) in that Johansen's Case III includes an additional trend term and his Case V includes a quadratic trend.

- *Case III: unrestricted intercepts, no trends*: $\mathbf{c}_0$ is to be estimated freely (i.e. it is not restricted to (23)), but $\boldsymbol{\gamma} = \mathbf{0}$ implies $\mathbf{c}_1 = \mathbf{0}$. Consequently

$$\Delta y_t = \mathbf{c}_0 + \boldsymbol{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \boldsymbol{\Pi}_y z_{t-1} + u_t \tag{26}$$

- *Case IV: unrestricted intercepts, restricted trends*: $\mathbf{c}_1 = -\boldsymbol{\Pi}\boldsymbol{\gamma}$, as in (23), but ignore the restrictions on $\mathbf{c}_0$. We obtain for (21):

$$\Delta y_t = \mathbf{c}_0 - \boldsymbol{\Pi}_y \boldsymbol{\gamma} t + \boldsymbol{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \boldsymbol{\Pi}_y z_{t-1} + u_t \tag{27}$$

- *Case V: unrestricted intercepts, unrestricted trends*: ignore the restrictions in (23) both for $\mathbf{c}_0$ and $\mathbf{c}_1$. Therefore

$$\Delta y_t = \mathbf{c}_0 + \mathbf{c}_1 t + \boldsymbol{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \boldsymbol{\Pi}_y z_{t-1} + u_t \tag{28}$$

Like Pesaran et al. (2000, pp.301-303), we proceed with concentrating on Case IV, which demonstrates both the treatment of restricted and of unrestricted deterministic terms (including a trend). First, note that under our modified Assumption $H_r$ ($\mathrm{rk}(\boldsymbol{\Pi}_y) = r$), we may express

$$\underset{n\times r}{\boldsymbol{\Pi}_y} = \underset{n\times r}{\boldsymbol{\alpha}_y} \underset{r\times m}{\boldsymbol{\beta}'} \quad , \qquad \mathrm{rk}(\boldsymbol{\alpha}_y) = \mathrm{rk}(\boldsymbol{\beta}) = r \tag{29}$$

here, $\boldsymbol{\beta}$, respectively $\boldsymbol{\alpha}_y$, is identified up to an $r \times r$ non-singular matrix.[10] Now consolidate $\boldsymbol{\Pi}_y$ by first defining

$$\underset{(m+1)\times r}{\boldsymbol{\beta}_*} \equiv \underset{(1+m)\times m}{\binom{-\boldsymbol{\gamma}'}{\mathbf{I}_m}} \underset{m\times r}{\boldsymbol{\beta}} \quad , \qquad \underset{n\times(m+1)}{\boldsymbol{\Pi}_{y*}} \equiv \boldsymbol{\alpha}_y \boldsymbol{\beta}_*' \quad , \qquad \underset{(1+m)\times 1}{z_{t-1}^*} \equiv \binom{t}{z_{t-1}} \tag{30}$$

which yields

$$\boldsymbol{\Pi}_{y*} z_{t-1}^* = -\boldsymbol{\Pi}_y \boldsymbol{\gamma} t + \boldsymbol{\Pi}_y \Delta z_{t-1} = \boldsymbol{\alpha}_y \boldsymbol{\beta}' \begin{pmatrix} -\boldsymbol{\gamma} & \mathbf{I}_m \end{pmatrix} \begin{pmatrix} t \\ z_{t-1} \end{pmatrix} = \boldsymbol{\alpha}_y \boldsymbol{\beta}_*' z_{t-1}^*$$

Consequently we may rewrite the VECM model from Case IV in (27) as

$$\Delta y_t = \mathbf{c}_0 + \boldsymbol{\Lambda}\Delta x_t + \sum_{i=1}^{p-1} \boldsymbol{\psi}_i \Delta z_{t-i} + \boldsymbol{\Pi}_{y*} z_{t-1}^* + u_t \tag{31}$$

[10] I.e. $\boldsymbol{\alpha}_y \boldsymbol{\beta}' = (\boldsymbol{\alpha}_y K^{-1})(K\boldsymbol{\beta}')$ for any non-singular matrix $K$.

Proceed by stacking $z_t$ etc. into "$T \times k$"-type matrices[11]

$$\underset{T\times n}{\Delta\mathbf{Y}} = \begin{pmatrix} \Delta y_1' \\ \vdots \\ \Delta y_T' \end{pmatrix} \qquad \underset{T\times 1}{\boldsymbol{\iota}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad \underset{T\times k}{\Delta\mathbf{X}} = \begin{pmatrix} \Delta x_1' \\ \vdots \\ \Delta x_T' \end{pmatrix} \quad \underset{T\times n}{\mathbf{U}} = \begin{pmatrix} u_1' \\ \vdots \\ u_T' \end{pmatrix}$$

$$\underset{T\times m}{\Delta\mathbf{Z}_{-}i} = \begin{pmatrix} \Delta z_{1-i}' \\ \vdots \\ \Delta z_{T-i}' \end{pmatrix} \quad \underset{T\times 1}{\boldsymbol{\tau}} = \begin{pmatrix} 1 \\ \vdots \\ T \end{pmatrix} \quad \underset{T\times m}{\mathbf{Z}_{-1}} = \begin{pmatrix} z_0' \\ \vdots \\ z_{T-1}' \end{pmatrix}$$

Moreover define the meta-structures

$$\underset{T\times(m+1)}{\mathbf{Z}_{-1}^*} = \begin{pmatrix} z_0^{*\prime} \\ \vdots \\ {z_{T-1}^*}' \end{pmatrix} = \begin{pmatrix} \boldsymbol{\tau} & \mathbf{Z}_{-1} \end{pmatrix} \qquad \breve{\boldsymbol{\Psi}} = \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{\psi}_1 & \dots & \boldsymbol{\psi}_{p-1} \end{pmatrix}$$

$$\underset{T\times(k+(p-1)m)}{\Delta\mathbf{Z}_{-}^*} = \begin{pmatrix} \Delta\mathbf{X} & \Delta\mathbf{Z}_{-1} & \dots & \Delta\mathbf{Z}_{-(p-1)} \end{pmatrix} = \begin{pmatrix} \Delta x_1' & \Delta z_0' & \dots & \Delta z_{1-p}' \\ \Delta x_2' & \Delta z_1' & \dots & \Delta z_{2-p}' \\ \vdots & \vdots & & \vdots \\ \Delta x_T' & \Delta z_{T-1}' & \dots & \Delta z_{T-p}' \end{pmatrix}$$

Note that sample size $T$ is redefined by deducting $p$ from the original $T$, such that the negative indexes above make sense. Hence

$$\Delta\mathbf{Z}_{-}\breve{\boldsymbol{\Psi}}' = \Delta\mathbf{X}\boldsymbol{\Lambda}' + \Delta\mathbf{Z}_{-1}\boldsymbol{\psi}_1' + \dots + \Delta\mathbf{Z}_{-(p-1)}\boldsymbol{\psi}_{p-1}' \tag{32}$$

whose $t$-th row is $(\boldsymbol{\Lambda}x_t + \sum_{i=1}^{p-1}\boldsymbol{\psi}_i z_{t-i})'$. With these definitions at hand, we may rewrite (31) in its stacked form:

$$\Delta\mathbf{Y} = \boldsymbol{\iota}c_0' + \Delta\mathbf{Z}_{-}\breve{\boldsymbol{\Psi}}' + \mathbf{Z}_{-1}^*\boldsymbol{\Pi}_{y*}' + \mathbf{U} \tag{33}$$

In order to proceed to maximum likelihood estimation, consider the density function of the error term in (33), which by 1 is from the multivariate normal distribution

$$f(u_t) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{|\Omega_{uu}|}} \exp\left(-\frac{1}{2}u_t'\Omega_{uu}^{-1}u_t\right) \tag{34}$$

As by Assumption 1 the individual $u_t$ are independent, we obtain $f(\mathbf{U}) = f(u_t)^T$. Hence the log-likelihood of $\boldsymbol{\Theta}$ is represented as follows, where $\boldsymbol{\Theta}$ is a vector collecting the unknown parameters in $\Omega_{uu}$, $\mathbf{c}_0$, $\breve{\boldsymbol{\Psi}}$ and $\boldsymbol{\Pi}_y$

$$\ell(\boldsymbol{\Theta}) = -\frac{nT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Omega_{uu}| - \frac{1}{2}\underbrace{\left(\sum_{t=1}^{T} u_t'\Omega_{uu}u_t\right)}_{=\mathrm{tr}(\mathbf{U}\Omega_{uu}\mathbf{U}')} \tag{35}$$

[11] Here we depart from Pesaran et al. (2000, p.302), since they stack the corresponding variables into "$k \times T$"-type matrices, i.e. the transposed versions of the ones used in our paper. The reason for this departure is to reduce the need for the numerous transpositions appearing in Pesaran et al. (2000).

Here, tr denotes the trace function. Start to "concentrate out"[12] parameters by replacing the parameters $\mathbf{c}_0$ and $\breve{\mathbf{\Psi}}$ with their maximisers obtained from the first order conditions of the likelihood with respect to these parameters. This translates into replacing $\mathbf{U}$ with the OLS residual $\hat{\mathbf{U}}$ and $\Omega_{uu}$ with its ML estimator $\hat{\Omega}_{uu}$. Note that $\hat{\Omega}_{uu} = \frac{1}{T}\hat{\mathbf{U}}'\hat{\mathbf{U}}$, hence

$$\operatorname{tr}\left(\hat{\mathbf{U}}\hat{\Omega}_{uu}^{-1}\hat{\mathbf{U}}'\right) = \operatorname{tr}\left(\hat{\mathbf{U}}'\hat{\mathbf{U}}\hat{\Omega}_{uu}^{-1}\right) = \operatorname{tr}\left(T\Omega_{uu}\Omega_{uu}^{-1}\right) = \operatorname{tr}\left(T\mathbf{I}_n\right) = Tn$$

which by substituting into (35) yields

$$\ell(\mathbf{\Theta}) = -\frac{nT}{2}\left(1 + \ln(2\pi)\right) - \frac{T}{2}\ln\left|\frac{1}{T}\hat{\mathbf{U}}'\hat{\mathbf{U}}\right| \tag{36}$$

Now suppose the "true" $\mathbf{\Pi}_{y*}$ is known and concentrate the parameters $\mathbf{c}_0$ and $\breve{\mathbf{\Psi}}$ out via their estimators. For this purpose, represent $\hat{\mathbf{U}}$ in (37)

$$\hat{\mathbf{U}} = \Delta\mathbf{Y} - \boldsymbol{\iota}\hat{\mathbf{c}}_0' + \Delta\mathbf{Z}_{-}\hat{\breve{\mathbf{\Psi}}}' + \mathbf{Z}_{-1}^{*}\mathbf{\Pi}_{y*}' \tag{37}$$

via the Frisch-Waugh theorem: Let

$$\hat{\mathbf{Z}}_{-1}^{*} = \mathbf{Z}_{-1}^{*} - \boldsymbol{\iota}\hat{\mathbf{c}}_{01}' - \Delta\mathbf{Z}_{-}\hat{\breve{\mathbf{\Psi}}}_1' = \left(\begin{pmatrix} \boldsymbol{\iota} & \Delta\mathbf{Z}_{-} \end{pmatrix}'\begin{pmatrix} \boldsymbol{\iota} & \Delta\mathbf{Z}_{-} \end{pmatrix}\right)^{-1}\begin{pmatrix} \boldsymbol{\iota} & \Delta\mathbf{Z}_{-} \end{pmatrix}'\Delta\mathbf{Z}_{-1}^{*}$$

$$\Delta\hat{\mathbf{Y}} = \Delta\mathbf{Y} - \boldsymbol{\iota}\hat{\mathbf{c}}_{02}' - \Delta\mathbf{Z}_{-}\hat{\breve{\mathbf{\Psi}}}_2' = \left(\begin{pmatrix} \boldsymbol{\iota} & \Delta\mathbf{Z}_{-} \end{pmatrix}'\begin{pmatrix} \boldsymbol{\iota} & \Delta\mathbf{Z}_{-} \end{pmatrix}\right)^{-1}\begin{pmatrix} \boldsymbol{\iota} & \Delta\mathbf{Z}_{-} \end{pmatrix}'\Delta\mathbf{Y}$$

...resulting in:

$$\hat{\mathbf{U}} = \Delta\hat{\mathbf{Y}} - \hat{\mathbf{Z}}_{-1}^{*}\mathbf{\Pi}_{y*}' \tag{38}$$

Then (36) collapses to the log-likelihood of $\mathbf{\Pi}_{y*}$ as in (39).

$$\ell(\mathbf{\Pi}_{y*}) = -\frac{nT}{2}\left(1 + \ln(2\pi)\right) - \frac{T}{2}\ln\underbrace{\left|\frac{1}{T}\left(\Delta\hat{\mathbf{Y}} - \hat{\mathbf{Z}}_{-1}^{*}\mathbf{\Pi}_{y*}'\right)'\left(\Delta\hat{\mathbf{Y}} - \hat{\mathbf{Z}}_{-1}^{*}\mathbf{\Pi}_{y*}'\right)\right|}_{=|T^{-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}|} \tag{39}$$

where maximizing (39) boils down to minimizing its latter part, the determinant $|T^{-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}|$. In order to re-express (39), introduce the notation

$$\frac{1}{T}\begin{pmatrix} \Delta\hat{\mathbf{Y}}' \\ \hat{\mathbf{Z}}_{-1}^{*\prime} \end{pmatrix}\begin{pmatrix} \Delta\hat{\mathbf{Y}} & \hat{\mathbf{Z}}_{-1}^{*} \end{pmatrix} = \begin{pmatrix} S_{yy} & S_{yz} \\ S_{zy} & S_{zz} \end{pmatrix} \tag{40}$$

thus the determinant $|T^{-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}|$ reduces to

$$\begin{aligned} \left|\frac{1}{T}\hat{\mathbf{U}}'\hat{\mathbf{U}}\right| &= \left|\frac{1}{T}\left(\Delta\hat{\mathbf{Y}}'\Delta\hat{\mathbf{Y}} - \Delta\hat{\mathbf{Y}}'\hat{\mathbf{Z}}_{-1}^{*}\mathbf{\Pi}_{y*}' - \mathbf{\Pi}_{y*}\hat{\mathbf{Z}}_{-1}^{*\prime}\Delta\hat{\mathbf{Y}} - \mathbf{\Pi}_{y*}\hat{\mathbf{Z}}_{-1}^{*\prime}\hat{\mathbf{Z}}_{-1}^{*}\mathbf{\Pi}_{y*}'\right)\right| = \\ &= \left|S_{yy} - S_{yz}\mathbf{\Pi}_{y*}' - \mathbf{\Pi}_{y*}S_{zy} + \mathbf{\Pi}_{y*}S_{zz}\mathbf{\Pi}_{y*}'\right| \end{aligned} \tag{41}$$

[12] For a brief outline of concentrated likelihood with respect to our problem, refer to Hayashi (2000, p.524) or Hamilton (1994, p.638)

Furthermore, assume that only the "true" $\boldsymbol{\beta}_*$ is known and thus concentrate $\boldsymbol{\alpha}_y$ out by its OLS estimator $\hat{\boldsymbol{\alpha}}_y$. The corresponding residual of this regression is then:

$$\hat{\mathbf{U}} = \left(\mathbf{I}_n - \hat{\mathbf{Z}}^*_{-1}\boldsymbol{\beta}_* \left(\boldsymbol{\beta}_*'\hat{\mathbf{Z}}^{*\prime}_{-1}\hat{\mathbf{Z}}^*_{-1}\boldsymbol{\beta}_*\right)^{-1}\boldsymbol{\beta}_*'\hat{\mathbf{Z}}^{*\prime}_{-1}\right)\Delta\hat{\mathbf{Y}} \tag{42}$$

By the properties of the projecting matrix in (42), inserting the determinant $|T^{-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}|$ into equation (39) results in

$$\ell\left(\boldsymbol{\beta}_*|r\right) = -\frac{nT}{2}\left(1+\ln\left(2\boldsymbol{\Pi}\right)\right) - \frac{T}{2}\ln\left|S_{yy} - S_{yz}\boldsymbol{\beta}_*\left(\boldsymbol{\beta}_*'S_{zz}\boldsymbol{\beta}_*\right)^{-1}\boldsymbol{\beta}_*'S_{zy}\right| \tag{43}$$

Here, hypothesis $H_r$, i.e. the rank of $\boldsymbol{\beta}_*$ plays an important role, hence we explicitly condition on $H_r$. Now note that since $\left|\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right| = |A||D - CA^{-1}B| = |D||A - CD^{-1}B|$,[13] we obtain for the matrix below

$$\left|\begin{pmatrix} S_{yy} & S_{yz}\boldsymbol{\beta}_* \\ \boldsymbol{\beta}_*'S_{zy} & \boldsymbol{\beta}_*'S_{zy}S_{yz}\boldsymbol{\beta}_* \end{pmatrix}\right| = |\boldsymbol{\beta}_*'S_{zz}\boldsymbol{\beta}_*|\left|S_{yy} - S_{yz}\boldsymbol{\beta}_*\left(\boldsymbol{\beta}_*'S_{zz}\boldsymbol{\beta}_*\right)^{-1}\boldsymbol{\beta}_*'S_{zy}\right| =$$
$$= |S_{yy}|\,|\boldsymbol{\beta}_*'S_{zz}\boldsymbol{\beta}_* - \boldsymbol{\beta}_*'S_{zy}S_{yy}^{-1}S_{yz}\boldsymbol{\beta}_*|$$

Consequently, minimizing the determinant $|T^{-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}|$ corresponds to minimizing

$$\left|S_{yy} - S_{yz}\boldsymbol{\beta}_*\left(\boldsymbol{\beta}_*'S_{zz}\boldsymbol{\beta}_*\right)^{-1}\boldsymbol{\beta}_*'S_{zy}\right| = \frac{|S_{yy}|\left|\boldsymbol{\beta}_*'\left(S_{zz} - S_{zy}S_{yy}^{-1}S_{yz}\right)\boldsymbol{\beta}_*\right|}{|\boldsymbol{\beta}_*'S_{zz}\boldsymbol{\beta}_*|} \tag{44}$$

with respect to $\boldsymbol{\beta}_*$. The minimization of (44) is to be achieved by applying Lemma A.8 of Johansen (1995, p.244), as it is quoted below

> *Lemma A.8 Let M be symmetric and positive semi-definite and N symmetric and positive definite. The function*
>
> $$f(x) = |x'Mx|/|x'Nx|$$
>
> *is maximised among all $p \times r$ matrices by $\hat{x} = (v_1, \dots, v_r)$, and the maximal value is $\prod_{i=1}^r \lambda_i$, where again $\lambda_i$ and $v_i$ are solutions to the eigenvalue problem*
>
> $$|\lambda N - M| = 0$$
>
> *where we assume that $\lambda_1 \geq \ldots \geq \lambda_p > 0$. We can also choose $\hat{x}$ times any non-singular $r \times r$ matrix as the maximizing argument. [...]*

The proof of this lemma is omitted here for brevity, but is given in Johansen (1995, p.224-226). Note that under the conditions above, minimizing $f(x) = |x'Mx|/|x'Nx|$ corresponds to maximizing $|x'Nx|/|x'Mx|$, i.e solving $|\lambda M - N| = 0$ and the minimised value is then $f(\hat{x}) = (\prod_i \lambda_i)^{-1}$. By redefining $\delta = \lambda^{-1}$, minimizing $f(x)$ corresponds to solving $|\delta N - M| = 0$ and the minimised value is $\prod_{i=1}^r \delta_i$.

[13] since $\left|\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right| = \left|\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} \mathbf{I} & -A^{-1}B \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\right| = |A||D - CA^{-1}B|$, similarly we obtain $= |D||A - CD^{-1}B|$

This implies that for our case the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_*$ for $\boldsymbol{\beta}_*$ is given by the eigenvectors of the generalised eigenvalues $\hat{\lambda}$ solving

$$\left|\hat{\delta} S_{zz} - \left(S_{zz} - S_{zy} S_{yy}^{-1} S_{yz}\right)\right| = |\underbrace{\hat{\lambda}}_{=1-\hat{\delta}} S_{zz} - S_{zy} S_{yy}^{-1} S_{yz}| = 0 \tag{45}$$

where we re-defined $\lambda \equiv 1 - \delta$; By the definition given in Hamilton (1994, p. 631), $\lambda$ corresponds to the canonical correlations between $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^{*\prime}_{-1}$.[14] The concentrated maximised likelihood of $\boldsymbol{\beta}_*$ is therefore

$$\begin{aligned} \ell_T^c(r) &= -\frac{nT}{2}\left(1 + \ln\left(2\boldsymbol{\Pi}\right)\right) - \frac{T}{2}\ln\left(|S_{yy}| \prod_{i=1}^{r} \hat{\delta}_i\right) \\ &= -\frac{nT}{2}\left(1 + \ln\left(2\boldsymbol{\Pi}\right)\right) - \frac{T}{2}\ln|S_{yy}| - \frac{T}{2}\sum_{i=1}^{r} \ln(1 - \hat{\lambda}_i) \end{aligned} \tag{46}$$

and the maximiser $\hat{\boldsymbol{\beta}}_*$ is given by the corresponding generalised eigenvectors. Note that the maximised log-likelihood $\ell_T^c$ is only a function of $r$, the cointegrating rank.
Up to now we have outlined the maximisation of the log-likelihood for $\boldsymbol{\beta}_*$ in Case IV; concerning the other four cases, the procedure of concentration in (37) is to be put slightly differently:

- *Case I:* $\mathbf{c}_0 = \mathbf{0}$ and $\mathbf{c}_1 = \mathbf{0}$: define $\mathbf{Z}^*_{-1} = \mathbf{Z}_{-1}$ and let $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ be the residuals of regressing $\Delta\mathbf{Y}$ and $\mathbf{Z}^*_{-1}$ on $\Delta\mathbf{Z}_{-}$.

- *Case II:* $\mathbf{c}_0 = -\boldsymbol{\Pi}_y\boldsymbol{\gamma}$ and $\mathbf{c}_1 = \mathbf{0}$: since the intercept $\mathbf{c}_0$ is only to be found in the cointegrating regression, define $\mathbf{Z}^*_{-1} = (\boldsymbol{\iota}, \quad \mathbf{Z}_{-1})$ and let $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ be the residuals of regressing $\Delta\mathbf{Y}$ and $\mathbf{Z}^*_{-1}$ on $\Delta\mathbf{Z}_{-}$.

- *Case III:* $\mathbf{c}_0$ unrestricted and $\mathbf{c}_1 = \mathbf{0}$: since the intercept affects the entire VECM, define $\mathbf{Z}^*_{-1} = \mathbf{Z}_{-1}$ and let $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ be the residuals of regressing $\Delta\mathbf{Y}$ and $\mathbf{Z}^*_{-1}$ on $(\boldsymbol{\iota}, \Delta\mathbf{Z}_{-})$.

- *Case IV:* $\mathbf{c}_0$ unrestricted and $\mathbf{c}_1 = \boldsymbol{\Pi}_y\boldsymbol{\gamma}$: since the trend term is in the cointegrating regression, define $\mathbf{Z}^*_{-1} = (\tau, \mathbf{Z}_{-1})$, and concentrate out by constructing $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ as in (37).

- *Case V:* $\mathbf{c}_0$ and $\mathbf{c}_1$ unrestricted: since both deterministic terms are present in the VECM model directly, concentrate out by constructing $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ as the residuals of regressing $\Delta\mathbf{Y}$ and $\mathbf{Z}_{-1}$ on $(\boldsymbol{\iota}, \boldsymbol{\tau}, \Delta\mathbf{Z}_{-})$.

[14] Solving for the generalised eigenvalue problem in (45), respectively solving for the eigenvalues $\zeta$ of $S_{zz}^{-1}S_{zy}S_{yy}^{-1}S_{yz}$ may be less straightforward than solving for the eigenvalues of a symmetric matrix. This obstacle is usually eluded by the following procedure: Let $A$, $B$ be two positive definite symmetric matrices, with the eigenvalue problem $ABv = \zeta v$, where $\zeta$ denotes the eigenvalues, and $v$ the eigenvectors of $AB$. We then may factorise $B = LL'$. (This decomposition may be e.g. easily obtained from the eigenvectors and eigenvalues of $B$.) Hence $ABv = ALL'v = \zeta v$. Multiply with $L'$ from the right and denote $w \equiv L'v$ to obtain $L'ALw = \zeta w$. Hence the eigenvalues of $AB$ are identical with those of $L'AL$. The eigenvectors $v$ of $AB$ may be recovered by $v = L'^{-1}w$.

# 4 The test for cointegrating rank

Pesaran et al. (2000) proceed by constructing likelihood ratio test-statistics out of (46). Subsequently they weaken the normal distribution assumption on $e_t$ and hence $u_t$[15] by stating the appropriate assumptions in order to ensure that a multivariate invariance principle (as in Phillips and Solo (1992)) can be applied, i.e. the scaled cumulated sum of $e_t$ converges to a multivariate Brownian motion (conditional on past $z_{t-i}$). This allows for conclusions on the asymptotic behaviour of likelihood-ratio statistics constructed out of (46).
Out of the likelihood defined in equation (46), we may test the null $H_r$ against two alternatives in Tests 1 and 2. Test 1 examines the alternative hypothesis whether the cointegrating rank might be $r+1$ versus the null of cointegrating rank $r$:

TEST 1 *$H_r$ against $H_{r+1}$*
*The log-likelihood ratio statistic for testing $H_r : \mathrm{rk}(\mathbf{\Pi}_y) = r$ against $H_{r+1} : \mathrm{rk}(\mathbf{\Pi}_y) = r+1$ is given by*

$$\mathcal{LR}\left(H_r|H_{r+1}\right) = -T \ln\left(1 - \hat{\lambda}_{r+1}\right) \tag{47}$$

*where $\hat{\lambda}_{r+1}$ is the $(r+1)$-th largest eigenvalue from equation* (45)*, applying the appropriate definitions for $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ (as on p. 7).*

In contrast, Test 2 is tests $H_r$ against the alternative of full cointegrating rank, that is, zero unit roots, respectively stationarity of $y_t$.

TEST 2 *$H_r$ against $H_n$*
*The log-likelihood ratio statistic for testing $H_r : \mathrm{rk}(\mathbf{\Pi}_y) = r$ against $H_n : \mathrm{rk}(\mathbf{\Pi}_y) = n$ (stationarity) is given by*

$$\mathcal{LR}\left(H_r|H_n\right) = -T \sum_{i=r+1}^{n} \ln\left(1 - \hat{\lambda}_i\right) \tag{48}$$

*where $\hat{\lambda}_i$ is the $i$-th largest eigenvalue from equation* (45)*, applying the appropriate definitions for $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$.*

In order to derive the asymptotic behaviour of test statistics 1 and 2 under more general conditions than just normal distribution of $e_t$, we invoke Assumption 5 below (cf. Assumption 4.1, Pesaran et al. (2000, p.304)).

ASSUMPTION **5**

*a. The error process $\{e_t\}_{t=-\infty}^{\infty}$ is such that $E(e_t|\{z_{t-i}\}_{i=-\infty}^{t-1}, z_0) = \mathbf{0}$ and $Var(e_t|\{z_{t-i}\}_{i=-\infty}^{t-1}, z_0) = \Omega_{ee}$, with $\Omega_{ee}$ being positive definite.*

*b. For the conditional term $u_t = e_{yt} - \Omega_{yx}\Omega_{yx}^{-1}e_{xt}$, assume $E(u_t|x_t, \{z_{t-i}\}_{i=-\infty}^{t-1}, z_0) = \mathbf{0}$ and $Var(u_t|x_t, \{z_{t-i}\}_{i=-\infty}^{t-1}, z_0) = \Omega_{uu}$*

*c. $\sup_t E(||e_t||^s) < \infty$ for some $s > 2$*

[15] Note that in order to carry out conditional inference on $u_t$, Pesaran et al. (2000, p.304) introduce a conditional linearity assumption as in Assumption 5(b) (cf. p.15).

Assumption 5 weakens the normality assumption 1 and, together with assumptions 2, 3 and 4, allows to infer the asymptotic behaviour for the two likelihood ratio statistics (under $H_r$) as given in Theorems 2 and 3.

First, denote the m-dimensional standard Brownian motion $W_m(a), a \in [0,1]$ with variance matrix $\mathbf{I}_m$ and partition it into the $(n–r)$- and $k$-dimensional sub-vector independent Brownian motions $W_{n-r}(a)$ and $W_k(a)$: $W_m(a) = (W_{n-r}(a)', W_m(a)')'$. Moreover denote the demeaned $(m–r)$ Brownian motion $\tilde{W}_{m-r}(a)$ and the demeaned and de-trended Brownian motion $\hat{W}_{m-r}(a)$ as follows (Pesaran et al., 2000, p.306)[16]

$$\tilde{W}_{m-r}(a) \equiv W_{m-r}(a) - \int_0^1 W_{m-r}(a)da \tag{49}$$

$$\hat{W}_{m-r}(a) \equiv \tilde{W}_{m-r}(a) - 12\left(a - \frac{1}{2}\right)\int_0^1 \left(a - \frac{1}{2}\right)\tilde{W}_{m-r}(a)da \tag{50}$$

Under this conditions, Pesaran et al. (2000, pp.306-307) derive the following statements on the asymptotic behaviour of the test statistics $\mathcal{LR}\left(H_r|H_{r+1}\right)$ and $\mathcal{LR}\left(H_r|H_n\right)$:

Theorem 2 *Under $H_r$ and Assumptions 2, 3, 4 and 5, the limit distribution of the test statistic $\mathcal{LR}\left(H_r|H_{r+1}\right)$ is given by the distribution of the maximum eigenvalue of*

$$\int_0^1 dW_{n-r}(a)F_{m-r}(a)'\left(\int_0^1 F_{m-r}(a)F_{m-r}(a)'da\right)^{-1}\int_0^1 F_{m-r}(a)dW_{n-r}(a)' \tag{51}$$

*where*

$$F_{m-r}(a) = \left\{\begin{array}{ll} W_{m-r}(a) & \textit{Case I} \\ (W_{m-r}(a)', 1)' & \textit{Case II} \\ \tilde{W}_{m-r}(a) & \textit{Case III} \\ (\tilde{W}_{m-r}(a)', a - 1/2)' & \textit{Case IV} \\ \hat{W}_{m-r}(a) & \textit{Case V} \end{array}\right\} \quad a \in [0,1] \tag{52}$$

Theorem 3 *Under $H_r$ and Assumptions 2, 3, 4 and 5, the limit distribution of the test statistic $\mathcal{LR}\left(H_r|H_n\right)$ is given by the distribution of*

$$\mathrm{tr}\left[\int_0^1 dW_{n-r}(a)F_{m-r}(a)'\left(\int_0^1 F_{m-r}(a)F_{m-r}(a)'da\right)^{-1}\int_0^1 F_{m-r}(a)dW_{n-r}(a)'\right] \tag{53}$$

*where $F_{m-r}(a)$ is specified according to cases as in* (52).

Note that Theorems 2 and 3 only apply to the case where Assumption 4 holds and the process of exogenous variables $\{x_t\}$ is driven by $k$ I(1) processes (i.e. the elements of $\{x_t\}$ are not cointegrated among themselves).

[16] To visualise the composition of the demeaned and detrended expression $\hat{W}_{m-r}$ in (49), imagine a finite data matrix $\mathbf{W}_T$ with observations $w_t$. By the Frisch-Waugh-Theorem, $\mathbf{W}_T$ can be demeaned and detrended by first constructing $\tilde{\mathbf{W}}_T$ (a demeaned $\mathbf{W}_T$) and $\tilde{\boldsymbol{\tau}}$ (a demeaned vector $\boldsymbol{\tau}$ as on p.9, i.e. $\tilde{\boldsymbol{\tau}} = \boldsymbol{\tau} - \boldsymbol{\iota}(T+1)/2$). Then the demeaned and detrended vector $\hat{W}_T$ may be constructed as $\hat{w}_t = \tilde{w}_t - (t - (T+1)/2)(\tilde{\boldsymbol{\tau}}'\tilde{\boldsymbol{\tau}})^{-1}\tilde{\boldsymbol{\tau}}'\tilde{\mathbf{W}}_T$. In the limit, this expression corresponds to the Wiener functional $\hat{W}(a) = \tilde{W}(a) - (a - 1/2)(\int_0^1 (a-1/2)^2 da)^{-1}\int_0^1 (a-1/2)W(a)da$, where $\int_0^1 (a-1/2)^2 da = 1/12$.

# 5 Asymptotic distribution for cointegrating rank test

This section will provide the proof for Theorems 2 and 3.[17] For this purpose it will combine material of Section 4 and Appendix A in Pesaran et al. (2000) and the corresponding sections in Johansen (1991, section 2, appendix A) and Johansen (1995, section 10, section 11). The section starts with lining out the foundations needed for an appropriate convergence of the partial sums involved, and then proceeds with reiterating several lemmata from Johansen (1991), or the corresponding versions provided by Pesaran et al. (2000) when needed. As in the latter article, the focus is on the asymptotics for the test statistic in Case IV - the other cases may be covered by slightly altering the proofs given in this section. The section closes with a proof that under $H_r$, the $n{-}r$ smallest roots of (46) converge to an expression that does not depend on nuisance terms and thus allows for asymptotic distributions not affected by these.

## Preliminary results

Remember from equation (8) that we have $C(L) = \sum_0^\infty C_j L^j$, which can be decomposed into $C(L) = C - (1-L)C^*(L)$, where $C^*(L) = \sum_0^\infty C_j^* L^j$, $C_j^* = \sum_{k=j+1}^\infty C_k$. From this we may infer that $\sum_1^\infty j|C_j| < \infty$ implies $\sum_1^\infty |C_j^*| < \infty$ and $C(1) = C < \infty$ (Phillips and Solo, 1992, p.972). Assumptions 2 and 3 together with the VAR(p) specification (1) ensure that this requirement is satisfied, i.e. $\sum_1^\infty j|C_j| < \infty$ (Pesaran et al. (2000, p.305)).
Furthermore, Phillips and Solo (1992, Theorem 3.15(b) p.983) state that if (a) we have a martingale difference sequence[18] $\{e_t\}$ with constant conditional variance (as in Assumption 5(a)) and (b) restrictions on the moments as in Assumption 5(c), and (c) $\sum_1^\infty j|C_j| < \infty$, then an *invariance principle* applies such that (cf. Pesaran et al. (2000, p.305)):

$$\mathcal{S}_T^e(a) = \frac{1}{\sqrt{T}} \sum_{s=1}^{[aT]} e_s \Rightarrow B_m(a) \qquad a \in [0,1] \tag{54}$$

where $[aT]$ denotes the integer fraction of $aT$, and $B_m(a)$ is an m-dimensional Brownian motion with variance matrix $\Omega_{ee}$; i.e. $B_m(a) = \Omega_{ee}^{\frac{1}{2}} W(a)$ where $W(a)$ is a vector-Brownian motion with independent elements.
Furthermore, note that Assumption 5(b) ensures that even when relaxing the normality assumption on $\{e_t\}$, $\mathrm{E}(u_t)$ is still linear in $e_{xt}$, respectively the condititonal expected value is $\mathrm{E}(e_{yt}|x_t, \{z_{t-i}\}_{i=-\infty}^{t-1}, z_0) = -\Omega_{yx}\Omega_{yx}^{-1} e_{xt}$. Together with the assumption on the variance of $u_t$, this puts the constant conditional variance to $\mathrm{Var}(e_{yt}|x_t, \{z_{t-i}\}_{i=-\infty}^{t-1}, z_0) = \Omega_{uu}$. Hence we may retain the notion of (31) being a conditional model for $\Delta y_t$ given $\Delta x_t$, $\{\Delta z_{t-i}\}_{i=-\infty}^{t-1}$ and the levels $\{z_t\}$ - upon which we may conduct conditional inference (Pesaran et al., 2000, pp. 304-5).
The *continuous mapping theorem* (as in Johansen (1995, Theorem B.5, p.243)) states that if $\mathcal{S}_T^e(a) \Rightarrow B_m(a)$ on $C[0,1]$[19] , and $\mathcal{F}()$ is any continuous functional on $C[0,1]$ with values in $\mathbb{R}^m$,

[17] The purpose of this paper to provide the proofs leading to Theorems 2 and 3 in the main body of the article, and in a consolidated fashion.

[18] Moreover, a technical requirement by Phillips and Solo (1992) is that $\{e_t\}$ is strongly uniformly integrable, viz. there exists a dominating random variable $Z$, for which $\mathrm{E}|Z| < \infty$, such that $\mathrm{P}(|e_t| \geq x) \leq c\mathrm{P}(|Z| \geq x) \forall x \geq 0, \quad \forall t \geq 1$ and some constant $c$.

[19] where $C[0,1]$ is the space of continuous functions on the unit interval.

then $\mathcal{F}(S_T^e(a)) \Rightarrow \mathcal{F}(B_m(a))$.

Referring to the model in (14), recall that $e_t$ was partitioned into $n$- and $k$-dimensional subvectors $e_{yt}$ and $e_{xt}$; with covariance matrix as in (16)

$$\mathcal{S}_T^e(a) = \frac{1}{\sqrt{T}} \sum_{s=1}^{[aT]} \begin{pmatrix} e_{yt} \\ e_{xt} \end{pmatrix} \Rightarrow B_m(a) = \begin{pmatrix} B_n(a) \\ B_k(a) \end{pmatrix} \quad \mathrm{E}(B_m B_m') = \begin{pmatrix} \Omega_{yy} & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{pmatrix} \equiv \Omega_{ee} \tag{55}$$

Moreover, remember from (17) that $u_t = e_{yt} - \Omega_{yx}\Omega_{xx}^{-1} e_{xt} = e_{yt} - \mathbf{\Lambda} e_{xt}$; via the continuous mapping theorem we therefore may express the convergence of the partial sums of $u_t$ as follows:

$$\frac{1}{\sqrt{T}} \sum_{s=1}^{[aT]} \begin{pmatrix} u_t \\ e_{xt} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n & -\mathbf{\Lambda} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \frac{1}{\sqrt{T}} \sum_{s=1}^{[aT]} \begin{pmatrix} e_{yt} \\ e_{xt} \end{pmatrix} \Rightarrow$$

$$\Rightarrow \begin{pmatrix} \mathbf{I}_n & -\mathbf{\Lambda} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \begin{pmatrix} B_n(a) \\ B_k(a) \end{pmatrix} = \begin{pmatrix} B_n(a) - \mathbf{\Lambda} B_k(a) \\ B_k(a) \end{pmatrix} \equiv \begin{pmatrix} B_n^*(a) \\ B_k(a) \end{pmatrix} = B_m^*(a) \tag{56}$$

where

$$\mathrm{E}(B_n^* B_k') = \mathrm{E}(B_n B_k') - \mathbf{\Lambda}\mathrm{E}(B_k B_k') = \Omega_{yx} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xx} = \mathbf{0},$$
$$\mathrm{E}(B_n^* B_n^{*\prime}) = \mathrm{E}(B_n B_n' - \mathbf{\Lambda} B_k B_n' + B_n B_k' \mathbf{\Lambda}' - \mathbf{\Lambda} B_k B_k' \mathbf{\Lambda}') = \Omega_{yy} - \Omega_{yx}\Omega_{xx}^{-1}\Omega_{xy} \equiv \Omega_{uu}$$

hence

$$\mathrm{E}(B_m^* B_m^{*}{}') = \begin{pmatrix} E(B_n^* B_n^{*\prime}) & E(B_n^* B_k') \\ E(B_k B_n^{*\prime}) & E(B_k B_k') \end{pmatrix} = \begin{pmatrix} \Omega_{uu} & \mathbf{0} \\ \mathbf{0} & \Omega_{xx} \end{pmatrix}$$

Proceed by defining

$$\underset{m\times(m-r)}{\boldsymbol{\alpha}_\perp} \equiv \begin{pmatrix} \underset{n\times(n-r)}{\boldsymbol{\alpha}_y^\perp} & \underset{n\times k}{\mathbf{0}} \\ \underset{k\times(n-r)}{\mathbf{0}} & \underset{k\times k}{\boldsymbol{\alpha}_x^\perp} \end{pmatrix} \qquad \boldsymbol{\alpha}_\perp' \boldsymbol{\alpha} = \mathbf{0}$$

where $\boldsymbol{\alpha}_x^\perp$ is any $k \times k$-dimensional non-singular matrix.

Next, define the following

$$\underset{(m+1)\times(m-r)}{\boldsymbol{\delta}} \equiv \begin{pmatrix} -\boldsymbol{\gamma}' \\ \mathbf{I}_m \end{pmatrix} \boldsymbol{\beta}^\perp \qquad \underset{(m+1)\times 1}{\boldsymbol{\xi}} \equiv \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \underset{(m+1)\times(m+1-r)}{\mathbf{B}_T} \equiv \begin{pmatrix} \boldsymbol{\delta} & 1/\sqrt{T}\boldsymbol{\xi} \end{pmatrix}$$

According to these definitions, $(\boldsymbol{\beta}, \boldsymbol{\beta}_\perp)$ provide a basis for $\mathbb{R}^m$, and the matrix $(\boldsymbol{\beta}_*, \boldsymbol{\xi}, \boldsymbol{\delta})$ provides a basis for $\mathbb{R}^{m+1}$. Investigate the interactions of these terms with $z_t$ by recalling equation (12):

$$z_t = \boldsymbol{\mu} + \boldsymbol{\gamma} t + C \sum_{i=1}^{t} e_i + C^*(L) e_t$$

Hence we obtain for $T^{-1/2}\boldsymbol{\delta}' z_t^*$:

$$\frac{1}{\sqrt{T}} \boldsymbol{\delta}' z_t^* = -\boldsymbol{\beta}_\perp' \boldsymbol{\gamma} t + \boldsymbol{\beta}_\perp' z_t = \frac{1}{\sqrt{T}} \boldsymbol{\beta}_\perp' \boldsymbol{\mu} + \frac{1}{\sqrt{T}} \boldsymbol{\beta}_\perp' C \sum_s e_s + \frac{1}{\sqrt{T}} \boldsymbol{\beta}_\perp' C^*(L) e_t$$
$$\Rightarrow \quad \boldsymbol{\beta}_\perp' C B_m(a) \tag{57}$$

The convergence to $\boldsymbol{\beta}'_{\perp} C B_m(a)$ as $T \to \infty$ obtains since $\boldsymbol{\beta}'_{\perp}\boldsymbol{\mu}$ is fixed, and $\boldsymbol{\beta}'_{\perp} C^*(L) e_t$ is a stationary term, hence by multiplication with $1/\sqrt{T}$ it asymptotically vanishes. Similarly, we get by the continuous mapping theorem

$$\frac{1}{\sqrt{T}} \boldsymbol{\delta}' \left( z_t^* - \bar{z}_t^* \right) \quad \Rightarrow \quad \boldsymbol{\beta}'_{\perp} C \tilde{B}_m(a) \tag{58}$$

where $\tilde{B}_m(a)$ denotes the demeaned Brownian motion $B_m(a) - \int_0^1 B_m(a)da$ (cf. Johansen (1995, p.145)). The differing deterministic terms affect the asymptotic behaviour, while the stationary terms again vanish. In contrast to above, consider $\boldsymbol{\beta}'_* z_t^*$:

$$\boldsymbol{\beta}'_* z_t^* = -\boldsymbol{\beta}'\boldsymbol{\gamma} t + \boldsymbol{\beta}' z_t = \boldsymbol{\beta}'\boldsymbol{\mu} + \underbrace{\boldsymbol{\beta}' C}_{=\mathbf{0}} \sum_s e_s + \boldsymbol{\beta}' C^*(L) e_t = O_p(1) \tag{59}$$

where $O_p(1)$ is shorthand notation for a term such that $\operatorname{plim} O_p(1) = \mathbf{0}$ as $T \to \infty$. The fact $\boldsymbol{\beta}' C = \mathbf{0}$ stems from Theorem 1, where $C = \boldsymbol{\beta}_{\perp} \left( \boldsymbol{\alpha}'_{\perp} \boldsymbol{\Gamma} \boldsymbol{\beta}_{\perp} \right)^{-1} \boldsymbol{\alpha}'_{\perp}$. Equations (58) and (59) illustrate that $\boldsymbol{\beta}'_* z_t^*$ represents the $r$-dimensional stationary processes in the system, while $\boldsymbol{\delta}' z_t^*$ represents the $(m-r)$ independent I(1) processes.

Note that in addition we have

$$\frac{1}{T} \boldsymbol{\xi}' z_t^* = \frac{t}{T} \quad \Rightarrow \quad a \tag{60}$$

We may use the results for $z_t^*$ for approximating the corresponding ones for $\hat{\mathbf{Z}}^*_{-1}$: This variable is equivalent to $\mathbf{Z}_{-1}$, but demeaned, and has the stationary terms roughly eliminated. The restrictions on $C^*(L)$ from p.15 ensure that the corresponding terms do not matter in the limit. We therefore may write

$$\boldsymbol{\beta}'_* \frac{1}{T} \hat{\mathbf{Z}}^{*\prime}_{-1} \hat{\mathbf{U}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\beta}'_* \hat{z}^*_{t-1} \hat{u}'_t \approx \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\beta}'_* (z^*_{t-1} - \bar{z}^*) \underbrace{(u_t - \bar{u}_t)'}_{\tilde{u}'_t} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\beta}' C^*(L) e_{t-1} \tilde{u}'_t =$$

$$= \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\beta}' C^*(L) e_{t-1} \tilde{e}'_t \begin{pmatrix} \mathbf{I}_n \\ -\boldsymbol{\Lambda}' \end{pmatrix} \xrightarrow{p} \mathbf{0} \tag{61}$$

where the equivalence $\tilde{u}_t = (I_n, -\boldsymbol{\Lambda}')\, \tilde{e}_t$ follows from equation (17). The convergence to $\mathbf{0}$ follows since the process $\{e_t\}$ is assumed to be not autocorrelated.

For $\boldsymbol{\delta}' \frac{1}{T} \hat{\mathbf{Z}}^{*\prime}_{-1} \hat{\mathbf{U}}$, the result is entirely different: First, note that $T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{t-1} e_s e'_t$ converges to $\int_0^1 B_m(a) dB_m(a)'$; Therefore, we obtain:

$$\boldsymbol{\delta}' \frac{1}{T} \hat{\mathbf{Z}}^{*\prime}_{-1} \hat{\mathbf{U}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\delta}' \hat{z}^*_{t-1} \hat{u}'_t \approx \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\delta}' (z^*_{t-1} - \bar{z}^*) \tilde{u}'_t =$$

$$= \frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{\beta}'_{\perp} C \sum_{s=1}^{t-1} e_s + \boldsymbol{\beta}'_{\perp} C^*(L) e_{t-1}) \tilde{u}'_t \tag{62}$$

For the same reasons as in (61), we may omit the latter term $\boldsymbol{\beta}'_{\perp} C^*(L) e_{t-1}$; thus by applying the continuous mapping theorem, we obtain for (62):

$$\boldsymbol{\delta}' \frac{1}{T} \hat{\mathbf{Z}}^{*\prime}_{-1} \hat{\mathbf{U}} \approx = \boldsymbol{\beta}'_{\perp} C \frac{1}{T} \sum_{t=1}^{T} \left( \sum_{s=1}^{t-1} e_s \right) \tilde{e}'_t \begin{pmatrix} \mathbf{I}_n \\ -\boldsymbol{\Lambda}' \end{pmatrix} \Rightarrow \boldsymbol{\beta}'_{\perp} C \int_0^1 B_m(a) dB_m(a)' \begin{pmatrix} \mathbf{I}_n \\ -\boldsymbol{\Lambda}' \end{pmatrix} \tag{63}$$

Concerning limiting covariance matrices, Phillips and Solo (1992, Theorem 3.16, p.983) prove that under assumptions[20] 2, 3, 4 and 5,[21]

$$S_{zz} = \frac{1}{T}\hat{\mathbf{Z}}^{*\prime}_{-1}\hat{\mathbf{Z}}^{*}_{-1} \quad \xrightarrow{p} \quad \Sigma_{zz} \tag{64}$$

These results enable the derivation of Lemma 1 (compare the corresponding lemmas in Johansen (1991, Lemma A.1, p.1567) and Johansen (1995, Lemma 10.3, p.146), as well as several results in Pesaran et al. (2000, p.331))

Lemma 1 *Some results on covariance matrices*

- $$\boldsymbol{\beta}_*^{\prime} S_{zz}\boldsymbol{\beta}_* \quad \xrightarrow{p} \quad \boldsymbol{\beta}_*^{\prime}\Sigma_{zz}\boldsymbol{\beta}_* \tag{65}$$
- $$\boldsymbol{\beta}_*^{\prime} S_{zy} = \boldsymbol{\beta}_*^{\prime}\frac{1}{T}\hat{\mathbf{Z}}^{*\prime}_{-1}\Delta\hat{\mathbf{Y}} \quad \xrightarrow{p} \quad \boldsymbol{\beta}_*^{\prime}\Sigma_{zy} = \boldsymbol{\beta}_*^{\prime}\Sigma_{zz}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime} \tag{66}$$
- $$S_{yy} = \frac{1}{T}\Delta\hat{\mathbf{Y}}^{\prime}\Delta\hat{\mathbf{Y}} \quad \xrightarrow{p} \quad \Sigma_{yy} = \Omega_{uu} + \boldsymbol{\alpha}_y\boldsymbol{\beta}_*^{\prime}\Sigma_{zz}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime} \tag{67}$$
- $$\left(\boldsymbol{\alpha}_y^{\prime}\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y^{\prime}\Sigma_{yy}^{-1} = \left(\boldsymbol{\alpha}_y^{\prime}\Omega_{uu}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y^{\prime}\Omega_{uu}^{-1} \tag{68}$$
- $$\Sigma_{yy}^{-1} - \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\left(\boldsymbol{\alpha}_y^{\prime}\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y^{\prime}\Sigma_{yy}^{-1} = \boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime} =$$
  $$\boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime} = \Omega_{uu}^{-1} - \Omega_{uu}^{-1}\boldsymbol{\alpha}_y\left(\boldsymbol{\alpha}_y^{\prime}\Omega_{uu}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y^{\prime}\Omega_{uu}^{-1} \tag{69}$$

Proof of Lemma 1:
Noting (64), (65) follows via the continuous mapping theorem. For (66) we obtain

$$\boldsymbol{\beta}_*^{\prime} S_{zy} = \boldsymbol{\beta}_*\frac{1}{T}\hat{\mathbf{Z}}^{*\prime}_{-1}\left(\hat{\mathbf{Z}}^{*}_{-1}\boldsymbol{\Pi}_{y*}^{\prime} + \mathbf{U}\right) = \boldsymbol{\beta}_*^{\prime} S_{zz}\boldsymbol{\Pi}_{y*}^{\prime} + \frac{1}{T}\boldsymbol{\beta}_*^{\prime}\hat{\mathbf{Z}}^{*\prime}_{-1}\mathbf{U} \quad \xrightarrow{p} \quad \boldsymbol{\beta}_*^{\prime}\Sigma_{zz}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime}$$

since $\frac{1}{T}\boldsymbol{\beta}_*^{\prime}\hat{\mathbf{Z}}^{*\prime}_{-1}\mathbf{U} \to \mathbf{0}$ as in (61). This term must be equivalent to the (prob) limit $\boldsymbol{\beta}_*^{\prime}\Sigma_{zy}$, which exists and is constant by the same conditions as in (64). Relation (67) is shown as follows:

$$S_{yy} = \frac{1}{T}\Delta\hat{\mathbf{Y}}^{\prime}\Delta\hat{\mathbf{Y}} = \frac{1}{T}\left(\hat{\mathbf{Z}}^{*}_{-1}\boldsymbol{\Pi}_{y*}^{\prime} + \hat{\mathbf{U}}\right)^{\prime}\left(\hat{\mathbf{Z}}^{*}_{-1}\boldsymbol{\Pi}_{y*}^{\prime} + \hat{\mathbf{U}}\right) =$$
$$\frac{1}{T}\boldsymbol{\alpha}_y\boldsymbol{\beta}_*^{\prime}\hat{\mathbf{Z}}^{*\prime}_{-1}\hat{\mathbf{Z}}^{*}_{-1}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime} + \frac{1}{T}\boldsymbol{\alpha}_y\boldsymbol{\beta}_*^{\prime}\hat{\mathbf{Z}}^{*\prime}_{-1}\hat{\mathbf{U}} + \frac{1}{T}\hat{\mathbf{U}}^{\prime}\hat{\mathbf{Z}}^{*}_{-1}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime} + \frac{1}{T}\hat{\mathbf{U}}^{\prime}\hat{\mathbf{U}} \to$$
$$\to \boldsymbol{\alpha}_y\boldsymbol{\beta}_*^{\prime}\Sigma_{zz}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime} + \Omega_{uu}$$

where (61) implies that the mixed terms converge to $\mathbf{0}$, while the limit of $\boldsymbol{\Pi}_{y*}\hat{\mathbf{Z}}^{*\prime}_{-1}\hat{\mathbf{Z}}^{*}_{-1}\boldsymbol{\Pi}_{y*}^{\prime}$ is given as in (65); and $\Omega_{uu}$ is found from standard central limit theorems.
In order to prove (68), note first that (cf. Johansen (1991, p.1568))

$$\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp} = \Omega_{uu}\boldsymbol{\alpha}_y^{\perp} + \underbrace{\boldsymbol{\alpha}_y\boldsymbol{\beta}_*^{\prime}\Sigma_{zz}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y^{\prime}\boldsymbol{\alpha}_y^{\perp}}_{\mathbf{0}} \tag{70}$$

[20] Actually Phillips and Solo (1992, p.983) require for this result a slightly stronger assumption, namely a dominating random variable $Z$ such that $\mathrm{E}(Z^4) < \infty$ (compare footnote 18). However this requirement appears to be satisfied by Assumption 5.

[21] Note that $\hat{\mathbf{Z}}^{*\prime}_{-1}$ has mean zero in Cases II-V (and in Case I if it is correctly specified), thus the first moment does not impede the convergence to a *covariance* matrix.

Now prove (68) by multiplying it from the right with the $n\times n$ matrix $(\boldsymbol{\alpha}_y,\ \Omega_{uu}\boldsymbol{\alpha}_y^{\perp}) = (\boldsymbol{\alpha}_y,\ \Sigma_{yy}\boldsymbol{\alpha}_y^{\perp})$

$$\begin{aligned}\left(\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\begin{pmatrix}\boldsymbol{\alpha}_y & \Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\end{pmatrix} &= \left(\boldsymbol{\alpha}_y'\Omega_{uu}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y'\Omega_{uu}^{-1}\begin{pmatrix}\boldsymbol{\alpha}_y & \Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\end{pmatrix}\\ \begin{pmatrix}\mathbf{I}_r & \mathbf{0}\end{pmatrix} &= \begin{pmatrix}\mathbf{I}_r & \mathbf{0}\end{pmatrix}\end{aligned}$$

Remember that $(\boldsymbol{\alpha}_y,\ \Omega_{uu}\boldsymbol{\alpha}_y^{\perp})$ has full rank by Assumption 5, Hypothesis $H_r$ and the definition of $\boldsymbol{\alpha}_{\perp}$, which proves relation (68).
The equivalences in (69) are proven similarly to (cf. Johansen (1991, p.1568)) by multiplying the first relation again with $(\boldsymbol{\alpha}_y,\ \Sigma_{yy}\boldsymbol{\alpha}_y^{\perp})$ from the right

$$\underbrace{\Sigma_{yy}^{-1} - \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\left(\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}}_{\Xi_1}\begin{pmatrix}\boldsymbol{\alpha}_y & \Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\end{pmatrix} = \underbrace{\boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime}}_{\Xi_2}\begin{pmatrix}\boldsymbol{\alpha}_y & \Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\end{pmatrix}$$

which yields

$$\begin{aligned}\begin{pmatrix}\Xi_1\boldsymbol{\alpha}_y & \Xi_1\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\end{pmatrix} &= \begin{pmatrix}\Xi_2\boldsymbol{\alpha}_y & \Xi_2\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\end{pmatrix}\\ \begin{pmatrix}\mathbf{0} & \boldsymbol{\alpha}_y^{\perp}\end{pmatrix} &= \begin{pmatrix}\mathbf{0} & \boldsymbol{\alpha}_y^{\perp}\end{pmatrix}\end{aligned}$$

By examination of the components we obtain:

$$\begin{aligned}\Xi_1\boldsymbol{\alpha}_y &= \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y - \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\left(\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y = \mathbf{0}\\ \Xi_2\boldsymbol{\alpha}_y &= \boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime}\boldsymbol{\alpha}_y = \mathbf{0}\\ \Xi_1\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp} &= \Sigma_{yy}^{-1}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp} - \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\left(\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp} = \boldsymbol{\alpha}_y^{\perp}\\ \Xi_2\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp} &= \boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp} = \boldsymbol{\alpha}_y^{\perp}\end{aligned}$$

which proves the first relation in (69). The relation $\boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Sigma_{yy}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime} = \boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime}$ is easily proven by invoking (70).
Finally, prove $\boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime} = \Omega_{uu}^{-1} - \Omega_{uu}^{-1}\boldsymbol{\alpha}_y\left(\boldsymbol{\alpha}_y'\Omega_{uu}^{-1}\boldsymbol{\alpha}_y\right)^{-1}\boldsymbol{\alpha}_y'\Omega_{uu}^{-1}$ once again by multiplying with $(\boldsymbol{\alpha}_y,\ \Omega_{uu}\boldsymbol{\alpha}_y^{\perp})$ from the right.

We now have provided the foundations for Lemma 2, the equivalent of Lemma A.1 in Pesaran et al. (2000, p.331) and of Lemma A.4 in Johansen (1991, p.1569):

Lemma 2 *Define the* $(m-r+1)\times 1$*-dimensional* $G(a)$ *as follows:*

$$G(a) = \begin{pmatrix}G_1(a)\\ G_2(a)\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_{\perp}'C\tilde{B}_m(a)\\ a-1/2\end{pmatrix} \tag{71}$$

*where* $\tilde{B}_m(a) = B_m(a) - \int_0^1 B_m(a)da$*. Moreover, let* $\mathbf{B}_T \equiv \left(\boldsymbol{\delta},\ T^{-1/2}\boldsymbol{\xi}\right)$*, and let* $S_{zz}$*,* $S_{zy}$ *and* $S_{yy}$ *be defined as in* (40)*. Then*

$$\frac{1}{T}\mathbf{B}_T'S_{zz}\mathbf{B}_T \ \Rightarrow\ \int_0^1 G(a)G(a)'da \tag{72}$$

$$\mathbf{B}_T'\left(S_{zy} - S_{zz}\boldsymbol{\Pi}_{y*}'\right) \ \Rightarrow\ \int_0^1 G(a)d\tilde{B}_n^*(a)' \tag{73}$$

*where* $\tilde{B}_n^*(a) = \tilde{B}_n(a) - \boldsymbol{\Lambda}\tilde{B}_k(a)$*.*

PROOF of Lemma 2:
Let $\mathbf{B}_T$ be defined as above. Then

$$\frac{1}{\sqrt{T}}\mathbf{B}_T'\hat{z}_t^* \approx \frac{1}{\sqrt{T}}\mathbf{B}_T'\tilde{z}_t^* = \begin{pmatrix} \frac{1}{\sqrt{T}}\boldsymbol{\delta}'\tilde{z}_t^* \\ \frac{1}{T}\boldsymbol{\xi}'\tilde{z}_t^* \end{pmatrix} \Rightarrow \begin{pmatrix} \boldsymbol{\beta}_\perp' C\tilde{B}_m(a) \\ a-\frac{1}{2} \end{pmatrix} = G(a) \tag{74}$$

This follows straight from equation (58) and (60). (Note that both elements of $G(a)$ are demeaned.) By the continuous mapping theorem we obtain therefore

$$\frac{1}{T}\mathbf{B}_T'S_{zz}\mathbf{B}_T = \frac{1}{T^2}\mathbf{B}_T'\hat{\mathbf{Z}}_{-1}^{*\prime}\hat{\mathbf{Z}}_{-1}^*\mathbf{B}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\sqrt{T}}\left(\mathbf{B}_T'\hat{z}_{t-1}^*\right)\left(\hat{z}_{t-1}^{*\prime}\mathbf{B}_T\right)\frac{1}{\sqrt{T}}$$
$$\Rightarrow \int_0^1 G(a)G(a)'da \tag{75}$$

Equation (73) is proven by using (63):

$$\mathbf{B}_T'\left(S_{zy}-S_{zz}\mathbf{\Pi}_{y*}'\right) = \mathbf{B}_T'\frac{1}{T}\hat{\mathbf{Z}}_{-1}^{*\prime}\underbrace{\left(\Delta\hat{\mathbf{Y}}-\hat{\mathbf{Z}}_{-1}^*\mathbf{\Pi}_{y*}'\right)}_{\hat{\mathbf{U}}} \Rightarrow \boldsymbol{\beta}_\perp' C\int_0^1 \tilde{B}_m(a)d\tilde{B}_n^*(a)' \tag{76}$$

where we apply again the continuous mapping theorem using (56): $d\tilde{B}_n^*(a) = (I_n,\ -\mathbf{\Lambda})\,d\tilde{B}_m(a)$

## Proofs of theorems 2 and 3

The lemmas provided on the previous pages may now serve to determine the asymptotic behaviour of the test statistics (47) and (48)

$$\mathcal{LR}\left(H_r|H_{r+1}\right) = -T\ln\left(1-\hat{\lambda}_{r+1}\right) \quad \text{and} \quad \mathcal{LR}\left(H_r|H_n\right) = -T\sum_{i=r+1}^{n}\ln\left(1-\hat{\lambda}_i\right)$$

In that respect, we are interested in the $(n-r)$ smallest solutions to

$$|S(\lambda)| \equiv \left|\lambda S_{zz} - S_{zy}S_{yy}^{-1}S_{yz}\right| = 0 \tag{77}$$

Proceed by defining the $(m+1)$-dimensional matrix $\mathbf{A}_T \equiv \left(\boldsymbol{\beta}_*,\ \mathbf{B}_T T^{-1/2}\right) = \left(\boldsymbol{\beta}_*,\ T^{-1/2}\boldsymbol{\delta},\ T^{-1}\boldsymbol{\xi}\right)$ where $\mathrm{rk}(\mathbf{A}_T) = m+1$ by definition. The full rank property of $\mathbf{A}_T$ implies that $\lambda$ solves $|S(\lambda)| = 0$ if and only if it solves $|\mathbf{A}_T'S(\lambda)\mathbf{A}_T| = 0$. Now consider $|\mathbf{A}_T'S(\lambda)\mathbf{A}_T|$ more closely (cf. Johansen (1991, p.1570)):

$$|\mathbf{A}_T'S(\lambda)\mathbf{A}_T| = \left|\begin{pmatrix} \boldsymbol{\beta}_*'S(\lambda)\boldsymbol{\beta}_* & \frac{1}{\sqrt{T}}\boldsymbol{\beta}_*'S(\lambda)\mathbf{B}_T \\ \frac{1}{\sqrt{T}}\mathbf{B}_T'S(\lambda)\boldsymbol{\beta}_* & \frac{1}{T}\mathbf{B}_T'S(\lambda)\mathbf{B}_T \end{pmatrix}\right| \Rightarrow$$
$$\Rightarrow \left|\lambda\begin{pmatrix} \boldsymbol{\beta}_*'\Sigma_{zz}\boldsymbol{\beta}_* & \mathbf{0} \\ \mathbf{0} & \int_0^1 G(a)G(a)'da \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}_*\Sigma_{zy}\Sigma_{yy}^{-1}\Sigma_{yz}\boldsymbol{\beta}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right| \tag{78}$$

The off-diagonal terms of (78) converge to $\mathbf{0}$ due to the following fact: Note that $T^{-1/2}\mathbf{B}_T'S_{zz}\boldsymbol{\beta}_* \approx T^{-1}\sum_t T^{-1/2}\mathbf{B}_T'\tilde{z}_{t-1}^*\tilde{z}_{t-1}^{*\prime}\boldsymbol{\beta}_*$. But $\boldsymbol{\beta}_*'z_t^* = O_p(1)$ from (59) and the convergence $T^{-1/2}\mathbf{B}_T'z_t^* \Rightarrow \boldsymbol{\beta}_\perp' CB_m(a)$ from (74) only requires a factor of $T^{-1/2}$, hence the factor $T^{-3/2}$ incites $T^{-1/2}\mathbf{B}_T'S_{zz}\boldsymbol{\beta}_*$ to approach the limit $\mathbf{0}$ (compare Johansen (1995, p.146-148)).

Similarly, $\mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz} \boldsymbol{\beta}_*$ converges to a constant, therefore pre-multiplication with $T^{-1/2}$ leads to a zero limit. Noting that $T^{-1/2}\mathbf{B}_T' S_{zz}\boldsymbol{\beta}_* \to \mathbf{0}$, we find from close inspection of equation (73) in Lemma 2 that $T^{-1/2}\mathbf{B}_T' S_{zy} \to \mathbf{0}$ [22]; and since $S_{yy}^{-1}$ converges to a constant, this implies that $T^{-1}\mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\mathbf{B}_T \to \mathbf{0}$. Hence we infer that the limit of $T^{-1}\mathbf{B}_T' S(\lambda)\mathbf{B}_T$ is the same as that of $T^{-1}\mathbf{B}_T' S(zz)\mathbf{B}_T$.

Thus equation (78) shows that in the limit $(m+1-r)$ roots of $|S(\lambda)| = 0$ must be zero, while only $r$ remain positive:[23] viz. for $\lambda_i$ such that $i > r$ we have $\lambda_i \to 0$ as $T \to 0$. Now let $\lambda$ in (78) converge such that $T\lambda = \rho$ remains fixed as $\lambda \to 0$, $T \to \infty$ (cf. Johansen (1991, p.1570)).

Next, note that, quite as before, $\lambda$ solves $|S(\lambda)| = 0$ if and only if it solves $|(\boldsymbol{\beta}_*,\ \mathbf{B}_T)' S(\lambda) (\boldsymbol{\beta}_*,\ \mathbf{B}_T)| = 0$, where

$$\left|(\boldsymbol{\beta}_*,\ \mathbf{B}_T)' S(\lambda)(\boldsymbol{\beta}_*,\ \mathbf{B}_T)\right| = \left|\begin{pmatrix} \boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_* & \boldsymbol{\beta}_*' S(\lambda)\mathbf{B}_T \\ \mathbf{B}_T' S(\lambda)\boldsymbol{\beta}_* & \mathbf{B}_T' S(\lambda)\mathbf{B}_T \end{pmatrix}\right| =$$
$$= |\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*| \left|\mathbf{B}_T' S(\lambda)\mathbf{B}_T - \mathbf{B}_T' S(\lambda)\boldsymbol{\beta}_* \left(\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*\right)^{-1} \boldsymbol{\beta}_*' S(\lambda)\mathbf{B}_T\right| \tag{79}$$

since $\left|\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right| = |A||D - CA^{-1}B|$. Note that the former term $|\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*|$ has dimension $r \times r$ while the latter term has dimension $(m+1-r) \times (m+1-r)$. Now consider $|\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*|$:

$$|\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*| = \left|\frac{\rho}{T}\boldsymbol{\beta}_*' S_{zz}\boldsymbol{\beta}_* - \boldsymbol{\beta}_*' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_*\right| \ \to\ \left|-\boldsymbol{\beta}_*'\Sigma_{zy}\Sigma_{yy}^{-1}\Sigma_{yz}\boldsymbol{\beta}_*\right|$$

hence $|\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*|$ converges to a constant as $T \to \infty$; consequently it has no roots $\rho$ that solve $|\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*| = 0$. Next, examine

$$\mathbf{B}_T' S(\lambda)\boldsymbol{\beta}_* = \frac{\rho}{T}\mathbf{B}_T' S_{zz}\boldsymbol{\beta}_* - \mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_* = -\mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_* + O_p(1)$$

Therefore we obtain for the latter term of (79):

$$\left|\mathbf{B}_T' S(\lambda)\mathbf{B}_T - \mathbf{B}_T' S(\lambda)\boldsymbol{\beta}_* \left(\boldsymbol{\beta}_*' S(\lambda)\boldsymbol{\beta}_*\right)^{-1} \boldsymbol{\beta}_*' S(\lambda)\mathbf{B}_T\right| = \tag{80}$$
$$\Big|\frac{\rho}{T}\mathbf{B}_T' S_{zz}\mathbf{B}_T - \mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\mathbf{B}_T - \left(-\mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_* + O_p(1)\right) \times$$
$$\times \left(\frac{\rho}{T}\boldsymbol{\beta}_*' S_{zz}\boldsymbol{\beta}_* - \boldsymbol{\beta}_*' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_*\right)^{-1} \left(-\mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_* + O_p(1)\right)' \Big| =$$
$$= \Big|\frac{\rho}{T}\mathbf{B}_T' S_{zz}\mathbf{B}_T - \mathbf{B}_T' S_{zy} \underbrace{\left(S_{yy}^{-1} - S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_* \left(\boldsymbol{\beta}_*' S_{zy} S_{yy}^{-1} S_{yz}\boldsymbol{\beta}_*\right)^{-1} \boldsymbol{\beta}_*' S_{zy} S_{yy}^{-1}\right)}_{\equiv N} S_{yz}\mathbf{B}_T + O_p(1)\Big|$$

Consider the limit of $N$:

$$N \to\ \Sigma_{yy}^{-1} - \Sigma_{yy}^{-1}\Sigma_{yz}\boldsymbol{\beta}_* \left(\boldsymbol{\beta}_*'\Sigma_{zy}\Sigma_{yy}^{-1}\Sigma_{yz}\boldsymbol{\beta}_*\right)^{-1} \boldsymbol{\beta}_*'\Sigma_{zy}\Sigma_{yy}^{-1} = \tag{81}$$
$$= \Sigma_{yy}^{-1} - \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\boldsymbol{\beta}_*'\Sigma_{zz}\boldsymbol{\beta}_* \left(\boldsymbol{\beta}_*'\Sigma_{zy}\Sigma_{yy}^{-1}\Sigma_{yz}\boldsymbol{\beta}_*\right)^{-1} \boldsymbol{\beta}_*'\Sigma_{zz}\boldsymbol{\beta}_*\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1} = \tag{82}$$
$$= \Sigma_{yy}^{-1} - \Sigma_{yy}^{-1}\boldsymbol{\alpha}_y \left(\boldsymbol{\alpha}_y'\Sigma_{yy}^{-1}\boldsymbol{\alpha}_y\right)^{-1} \boldsymbol{\alpha}_y'\Sigma_{yy}^{-1} = \tag{83}$$
$$= \boldsymbol{\alpha}_y^{\perp} \left(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\right)^{-1} \boldsymbol{\alpha}_y^{\perp\prime} \tag{84}$$

[22] As $\mathbf{B}_T' S_{zz}\boldsymbol{\beta}_*$ vanishes asymptotically, the expression in (73) becomes dominated by $\mathbf{B}_T' S_{zy}$ which converges to $\int G(a)d\tilde{B}_n^{*\prime}(a)$. But then the latter term of $T^{-1}\mathbf{B}_T' S(\lambda)\mathbf{B}_T$, namely $T^{-1/2}\mathbf{B}_T' S_{zy} S_{yy}^{-1} S_{yz}\mathbf{B}_T T^{-1/2}$ converges to $\mathbf{0}$.

[23] The $r$ largest roots remain positive since $S(\lambda)$ is positive semi-definite and symmetric.

where (82) follows from (66), relation (83) is a consequence of applying the matrix inversion to the expression in brackets, and (84) follows from (69).
This implies that the limit of equation (80) is equivalent to that of:

$$\left|\rho\frac{1}{T}\mathbf{B}_T'S_{zz}\mathbf{B}_T - \mathbf{B}_T'S_{zy}\boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime}S_{yz}\mathbf{B}_T + O_p(1)\right| \Rightarrow$$
$$\Rightarrow \left|\rho\int_0^1 G(a)G(a)'da - \int_0^1 G(a)dB_n^{*\prime}(a)\boldsymbol{\alpha}_y^{\perp}\left(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp}\right)^{-1}\boldsymbol{\alpha}_y^{\perp\prime}\int_0^1 dB_n^*(a)G(a)'\right| \tag{85}$$

Note that in the latter term the expression $\mathbf{B}_T'S_{zy}\boldsymbol{\alpha}_y^{\perp} = \mathbf{B}_T'\left(S_{zy} - S_{zz}\mathbf{\Pi}_{y*}'\right)\boldsymbol{\alpha}_y^{\perp}$, hence the limit of this term follows from Lemma 2; as does the limit of the term $\frac{1}{T}\mathbf{B}_T'S_{zz}\mathbf{B}_T$. This demonstrates that for the $(m+1-r)$ smallest solutions $\hat{\lambda}_i$ of $|S(\lambda)| = 0$, the term $T\hat{\lambda}_i$ converges to the corresponding root $\rho_i$ of equation (85).
In order to convert (85) into an expression consisting of independent Brownian motions, consider that the $(m-r)$-dimensional transformation $\boldsymbol{\alpha}_\perp' B_m(a)$ has covariance matrix $\mathrm{E}(\boldsymbol{\alpha}_\perp' B_m(a)B_m(a)'\boldsymbol{\alpha}_\perp) = \boldsymbol{\alpha}_\perp'\Omega_{ee}\boldsymbol{\alpha}_\perp$ where $\Omega_{ee}$ was defined in equation (55).
Hence we may construct a Brownian motion $W_{m-r}(a) = \left(\boldsymbol{\alpha}_\perp'\Omega_{ee}\boldsymbol{\alpha}_\perp\right)^{-1/2}\boldsymbol{\alpha}_\perp' B_m(a)$. (Correspondingly we obtain for the corresponding demeaned Brownian motion $\left(\boldsymbol{\alpha}_\perp'\Omega_{ee}\boldsymbol{\alpha}_\perp\right)^{1/2}\tilde{W}_{m-r}(a) = \boldsymbol{\alpha}_\perp'\tilde{B}_m(a)$.)
Next recall the composition of $G(a)$ from Lemma 2 and that of $C$ from Theorem 1:

$$G(a) = \begin{pmatrix}\boldsymbol{\beta}_\perp' C\tilde{B}_m(a)\\ a-\frac{1}{2}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp(\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp)^{-1}\boldsymbol{\alpha}_\perp'\tilde{B}_m(a)\\ a-\frac{1}{2}\end{pmatrix} =$$
$$= \begin{pmatrix}\boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp(\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp)^{-1}\left(\boldsymbol{\alpha}_\perp'\Omega_{ee}\boldsymbol{\alpha}_\perp\right)^{1/2}\tilde{W}_{m-r}(a)\\ a-\frac{1}{2}\end{pmatrix} =$$
$$\begin{pmatrix}\mathbf{L}_{11} & \mathbf{0}\\ \mathbf{0} & 1\end{pmatrix}F(a) = \mathbf{L}F(a) \qquad \text{where} \quad F(a) \equiv \begin{pmatrix}\tilde{W}_{m-r}\\ a-\frac{1}{2}\end{pmatrix} \tag{86}$$

and the left upper entry of the $(m-r+1)\times(m-r+1)$ matrix $\mathbf{L}$ is given as $\mathbf{L}_{11} = \boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp(\boldsymbol{\alpha}_\perp'\mathbf{\Gamma}\boldsymbol{\beta}_\perp)^{-1}\left(\boldsymbol{\alpha}_\perp'\Omega_{ee}\boldsymbol{\alpha}_\perp\right)^{1/2}$.
Moreover recall from (56) that $B_m^*(a)' = (B_n^*(a)',\ B_k(a)')$ and that for the $m-r$-dimensional transformation $\boldsymbol{\alpha}_\perp' B_m^*(a)$ we have

$$\mathrm{E}(\boldsymbol{\alpha}_\perp' B_m^*(a)B_m^*(a)'\boldsymbol{\alpha}_\perp) = \boldsymbol{\alpha}_\perp'\begin{pmatrix}\Omega_{uu} & \mathbf{0}\\ \mathbf{0} & \Omega_{xx}\end{pmatrix}\boldsymbol{\alpha}_\perp = \begin{pmatrix}\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\alpha}_x^{\perp\prime}\Omega_{xx}\boldsymbol{\alpha}_x^{\perp}\end{pmatrix}$$

This block-diagonality allows for constructing an $m$-dimensional independent Brownian motion as

$$\begin{pmatrix}W_{n-r}(a)\\ W_k(a)\end{pmatrix} = \begin{pmatrix}(\boldsymbol{\alpha}_y^{\perp\prime}\Omega_{uu}\boldsymbol{\alpha}_y^{\perp})^{-1/2}\boldsymbol{\alpha}_y^{\perp\prime}B_n^*(a)\\ (\boldsymbol{\alpha}_x^{\perp\prime}\Omega_{xx}\boldsymbol{\alpha}_x^{\perp})^{-1/2}\boldsymbol{\alpha}_x^{\perp\prime}B_k(a)\end{pmatrix} \tag{87}$$

Proceed likewise for the deterministics-adjusted versions $\tilde{B}_m^*(a)$ and $\hat{B}_m^*(a)$. Thus using (87) and (86), we may rewrite (85) as

$$\left|\mathbf{L}\Big(\underbrace{\rho\int_0^1 F(a)F(a)'da}_{\Theta_1} - \underbrace{\int_0^1 F(a)d\tilde{W}_{n-r}(a)'}_{\Theta_2}\underbrace{\int_0^1 d\tilde{W}_{n-r}(a)F(a)'}_{\Theta_2'}\Big)\mathbf{L}'\right| = 0 \tag{88}$$

Assumptions 3 and 5, and the definitions of $\boldsymbol{\alpha}_\perp$ and $\boldsymbol{\beta}_\perp$ ensure that $\mathbf{L}$ has full rank.[24] Hence $\rho$ solves (88) if and only if it solves $|\rho\Theta_1 - \Theta_2\Theta_2'| = 0$. Now multiply with $\Theta_2'\Theta_1^{-1}$ from the left, and with $\Theta_2(\Theta_2'\Theta_2)^{-1}$ from the right to obtain

$$\left|\rho\mathbf{I}_{n-r} - \int_0^1 d\tilde{W}_{n-r}(a)F(a)'\left(\int_0^1 F(a)F(a)'da\right)^{-1}\int_0^1 F(a)d\tilde{W}_{n-r}(a)'\right| = 0 \tag{89}$$

Therefore, $T$ times the $(n-r)$ smallest roots of $|S(\lambda)| = 0$ converge to the eigenvalues of $\Theta_2'\Theta_1^{-1}\Theta_2$ Finally, reconsider the test statistics:

$$\mathcal{LR}\left(H_r|H_n\right) = -T\sum_{i=r+1}^{n}\ln\left(1-\hat{\lambda}_i\right) = -T\sum_{i=r+1}^{n}\hat{\lambda}_i + O_p(1) \Rightarrow$$

$$\Rightarrow \sum_{i=r+1}^{n}\hat{\rho}_i = \operatorname{tr}\left(\int_0^1 d\tilde{W}_{n-r}(a)F(a)'\left(\int_0^1 F(a)F(a)'da\right)^{-1}\int_0^1 F(a)d\tilde{W}_{n-r}(a)'\right) \tag{90}$$

while

$$\mathcal{LR}\left(H_r|H_{r+1}\right) = -T\ln\left(1-\hat{\lambda}_{r+1}\right) = -T\hat{\lambda}_{r+1} + O_p(1) \Rightarrow \hat{\rho}_{r+1} \tag{91}$$

and $\hat{\rho}_{r+1}$ equals the largest eigenvalue of

$$\int_0^1 d\tilde{W}_{n-r}(a)F(a)'\left(\int_0^1 F(a)F(a)'da\right)^{-1}\int_0^1 F(a)d\tilde{W}_{n-r}(a)' \tag{92}$$

This completes the proofs of Theorems 2 and 3 for the Case IV. The corresponding proofs for the other cases differ only slightly from this proof.

# 6 Asymptotic tests on coefficient restrictions

In addition to Theorems 2 and 3, Pesaran et al. (2000) provide several test statistics on whether the restrictions on $\mathbf{c}_0$ and $\mathbf{c}_1$ from Cases I-V are correctly specified (compare Johansen (1995, pp.161-162)). Moreover, the authors elaborate on methods and tests regarding restrictions on the short run dynamics of a VECM as in (20). For our purpose however, we regard the test statistics concerned with the weak exogeneity restriction in Assumption 4 to be the most important contributions of this article: Similar to Pesaran et al. (2000, pp.309-311), we will concentrate henceforth on case IV, where the other cases follow completely analogously.
Assumption 4 implies that $\{x_t\}$ is integrated of order one and long-run forcing for $\{y_t\}$. These statements imply that $\{x_t\}$ is not cointegrated on its own; and that the differenced process $\{\Delta x_t\}$ does not depend on the lagged cointegration equation $\boldsymbol{\beta}_* z_{t-1}^*$. In order to investigate the former issue, augment the sub-system model from (22) by keeping part of the $\mathbf{\Pi}_x$ matrix:

$$\Delta x_t = \mathbf{a}_{x0} + \sum_{i=1}^{p-1}\mathbf{\Gamma}_{xi}\Delta z_{t-i} + \mathbf{\Pi}_{xx*}x_{t-1}^* + e_{xt} \tag{93}$$

where we define:

$$\boldsymbol{\gamma} = \begin{pmatrix}\boldsymbol{\gamma}_y \\ \boldsymbol{\gamma}_x\end{pmatrix} \qquad \mathbf{\Pi}_x = \begin{pmatrix}\mathbf{\Pi}_{xy} & \mathbf{\Pi}_{xx}\end{pmatrix} \qquad \mathbf{\Pi}_{xx*} = \mathbf{\Pi}_{xx}\begin{pmatrix}\boldsymbol{\gamma}_x & \mathbf{I}_k\end{pmatrix} \qquad x_{t-1}^* = \begin{pmatrix} t \\ x_{t-1}\end{pmatrix}$$

[24]Moreover the matrix $\Theta_1$ is non-singular a.s. and $\Theta_2$ has full row rank a.s.

Note that while we now allow $\mathbf{\Pi}_{xx} \neq \mathbf{0}$, we still require $\mathbf{\Pi}_{xy} = \mathbf{0}$, hence the process $\{x_t\}$ may be cointegrated but still long-run forcing for $\{y_t\}$. Now introduce the following hypothesis about $\mathbf{\Pi}_{xx}$:

HYPOTHESIS $H_r^x : \mathrm{rk}(\mathbf{\Pi}_{xx}) = r \qquad r = 0, 1, \ldots, k$

Similar to section 3, we may stack and concentrate out the parameters $\mathbf{a}_{x0}$ and $\mathbf{\Gamma}_i$:

$$\Delta\hat{\mathbf{X}} = \Delta\mathbf{X} - \boldsymbol{\iota}\hat{\mathbf{a}}_{x0}' - \sum_{i=1}^{p-1} \Delta\mathbf{X}_{-i}\hat{\mathbf{\Gamma}}_{xi}'$$

$$\hat{\mathbf{X}}_{-1}^{*}\boldsymbol{\beta}_x^{*}\boldsymbol{\alpha}_{xx}' = \left(\mathbf{X}_{-1}^{*} - \boldsymbol{\iota}\hat{\mathbf{a}}_{x0}' - \sum_{i=1}^{p-1} \Delta\mathbf{X}_{-i}\hat{\mathbf{\Gamma}}_{xi}'\right)\mathbf{\Pi}_{xx*}'$$

Next, denote $S_{11} = T^{-1}\hat{\mathbf{X}}_{-1}^{*\prime}\hat{\mathbf{X}}_{-1}^{*}$, $S_{01} = T^{-1}\Delta\hat{\mathbf{X}}'\hat{\mathbf{X}}_{-1}^{*}$ and $S_{00} = T^{-1}\Delta\hat{\mathbf{X}}'\Delta\hat{\mathbf{X}}$ and further concentrate out $\boldsymbol{\alpha}_{xx}$ to obtain

$$\frac{1}{T}\hat{\mathbf{E}}_x'\hat{\mathbf{E}}_x = S_{00} - S_{01}\boldsymbol{\beta}_{xx*}\left(\boldsymbol{\beta}_{xx*}'S_{11}\boldsymbol{\beta}_{xx*}\right)^{-1}\boldsymbol{\beta}_{xx*}'S_{10}$$

The likelihood $\ell(\boldsymbol{\beta}_{xx*}|r)$ is derived from the distribution function of $\{e_t\}$, and after concentrating out, collapses to a constant plus $|T^{-1}\hat{\mathbf{E}}_x'\hat{\mathbf{E}}_x|$, just as in (43). Similar to the further development in section 3, the maximised likelihood is given by the solutions $\lambda$ to

$$|S_{11}\lambda - S_{10}S_{00}^{-1}S_{01}| = 0 \tag{94}$$

Thus the likelihood ratio statistic for testing $H_0^x$ against $H_1^x$ is therefore

TEST 3 $H_0^x$ *vs.* $H_1^x$
*The log-likelihood ratio statistic for testing* $H_0^x : \mathbf{\Pi}_{xx} = 0$ *against* $H_1^x : \mathrm{rk}(\mathbf{\Pi}_{xx}) = 1$ *is given by*

$$\mathcal{LR}\left(H_0^x|H_1^x\right) = -T\ln\left(1 - \hat{\lambda}_1\right) \tag{95}$$

*where* $\hat{\lambda}_1$ *is the maximum solution to equation* (94).

The test statistic for the alternative $H_k^x$ is similar:

TEST 4 $H_0^x$ *vs.* $H_k^x$
*The log-likelihood ratio statistic for testing* $H_0^x : \mathbf{\Pi}_{xx} = 0$ *against* $H_k^x : \mathrm{rk}(\mathbf{\Pi}_{xx}) = k$ *is given by*

$$\mathcal{LR}\left(H_0^x|H_k^x\right) = -T\sum_{r=1}^{k}\ln\left(1 - \hat{\lambda}_r\right) \tag{96}$$

*where* $\hat{\lambda}_r$, $r = 0, 1, \ldots, k$ *are the solutions to equation* (94).

Apart from the difference in filtering on $\Delta z_{t-i}$, the test statistics are thus the same as the well-known Johansen (1991) test statistic, respectively the Pesaran et al. (2000) test statistic (cf. Tests 1 and 2) for the degenerate case $k = 0$. Unsurprisingly, the corresponding statistics converge to expressions as in Theorems 4 and 5 (Pesaran et al., 2000, Theorems 4.7 and 4.8, p.310):

Theorem 4 *Limit distribution of Test 3*
*Under $H_r$ as well as Assumptions 2, 3, 4 and 5, the limit distribution of Test 3 is the distribution of the maximum eigenvalue of*

$$\int_0^1 dW_k(a)F_k(a)' \left(\int_0^1 F_k(a)F_k(a)'da\right)^{-1} \int_0^1 F_k(a)dW_k(a)' \tag{97}$$

*where*

$$F_k(a) = \left\{ \begin{array}{ll} W_k(a) & \textit{Case I} \\ (W_k(a)', 1)' & \textit{Case II} \\ \tilde{W}_k(a) & \textit{Case III} \\ (\tilde{W}_k(a)', a - 1/2)' & \textit{Case IV} \\ \hat{W}_k(a) & \textit{Case V} \end{array} \right\} \quad a \in [0, 1] \tag{98}$$

Theorem 5 *Limit distribution of Test 4*
*Under $H_r$ as well as Assumptions 2, 3, 4 and 5, the limit distribution of Test 3 is given by the distribution of*

$$\operatorname{tr}\left(\int_0^1 dW_k(a)F_k(a)' \left(\int_0^1 F_k(a)F_k(a)'da\right)^{-1} \int_0^1 F_k(a)dW_k(a)'\right) \tag{99}$$

*where $F_k(a)$ is defined as in* (98).

The proofs of Theorems 4 and 5 are just a degenerate case of the proofs of Theorems 2 and 3.
Next, consider a specification test for the second statement (that the lagged cointegration relationship does not affect $\{\Delta x_t\}$) with respect to the sub-system model

$$\Delta x_t = \mathbf{a}_{x0} + \sum_{i=1}^{p-1} \mathbf{\Gamma}_{xi}\Delta z_{t-i} + \boldsymbol{\alpha}_{xy}\hat{\boldsymbol{\beta}}_*' x^*_{t-1} + e_{xt} \tag{100}$$

where $\hat{\boldsymbol{\beta}}_*$ denotes the estimator of the cointegrating vector under $H_r$ in Case IV from (46). The likelihood ratio statistic for $\boldsymbol{\alpha}_{xy} = \mathbf{0}$ is given by $f(\hat{E}|\hat{\boldsymbol{\alpha}}_x y) - f(\hat{E}|\boldsymbol{\alpha}_x y = \mathbf{0})$. Concentrating out $\Delta\hat{\mathbf{X}}$ as above and $\hat{\mathbf{Z}}^*_{-1,x}\hat{\boldsymbol{\beta}}_*$[25] as in

$$\hat{\mathbf{Z}}^*_{-1,x}\hat{\boldsymbol{\beta}}_* = \left(\mathbf{Z}^*_{-1} - \boldsymbol{\iota}\hat{\mathbf{a}}_{x0}' - \sum_{i=1}^{p-1} \Delta\mathbf{X}_{-i}\hat{\mathbf{\Gamma}}_{xi}'\right)\hat{\boldsymbol{\beta}}_*$$

yields

$$\hat{\boldsymbol{\alpha}}_{xy} = \Delta\hat{X}'\hat{\mathbf{Z}}^*_{-1,x}\hat{\boldsymbol{\beta}}_* \left(\hat{\boldsymbol{\beta}}_*'\hat{\mathbf{Z}}^{*\prime}_{-1,x}\hat{\mathbf{Z}}^*_{-1,x}\hat{\boldsymbol{\beta}}_*\right)^{-1}$$

and by straightforward likelihood ratio construction we arrive at Test 5:

Test 5 $\boldsymbol{\alpha}_{xy} = \mathbf{0}$ *vs.* $\boldsymbol{\alpha}_{xy} \neq \mathbf{0}$
*Under $H_r$, the log-likelihood ratio statistic for testing $H_0 : \boldsymbol{\alpha}_{xy} = \mathbf{0}$ against $H_A : \boldsymbol{\alpha}_{xy} \neq \mathbf{0}$ is given by*

$$\mathcal{LR}(\boldsymbol{\alpha}_{xy}) = T\left(\log\left|T^{-1}(\Delta\hat{\mathbf{X}} - \hat{\mathbf{Z}}^*_{-1,x}\hat{\boldsymbol{\beta}}_*\hat{\boldsymbol{\alpha}}_{xy}')'((\Delta\hat{\mathbf{X}} - \hat{\mathbf{Z}}^*_{-1,x}\hat{\boldsymbol{\beta}}_*\hat{\boldsymbol{\alpha}}_{xy}'))\right| - \log\left|T^{-1}\Delta\hat{\mathbf{X}}'\Delta\hat{\mathbf{X}}\right|\right) \tag{101}$$

*where $\hat{\mathbf{Z}}^*_{-1,x}$ and $\Delta\hat{\mathbf{X}}$ are residuals according to cases as on page 23, and $\hat{\boldsymbol{\beta}}_*$ stems from* (46).

[25] Note that $\hat{\mathbf{Z}}^*_{-1,x}$ differs from $\hat{\mathbf{Z}}^*_{-1}$ given in (38) as it is not adjusted for $\mathbf{\Lambda}\Delta\mathbf{X}$.

Pesaran et al. (2000, Theorem 4.9, p.335) infer that the limit distribution of test statistic 5 is given by a $\chi^2$-distribution:

THEOREM 6 *Limit distribution of Test 5*
*Under $H_r$ and Assumptions 2, 3, 4 & 5, the limit distribution of Test 5 has a limiting $\chi^2$-distribution with $kr$ degrees of freedom for Cases I-V, $r = 1, \ldots, n$.*

The proof of 6 is omitted here (as it is done in Pesaran et al. (2000, p.335) for its major part). The result follows straightforward from Pesaran et al. (2000, Lemma A.3, p.335) where the authors prove that $\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_* = O_p(1)$, even in the case where Assumption 4 does not hold. Hence in the limit, the resulting term $\boldsymbol{\beta}'_* \hat{z}^*_{t-1,x}$ is stationary. Since the remaining terms in (101) are stationary as well, the result follows as for the standard likelihood ratio case.

# 7 Concluding Remarks

We have introduced Test statistics 1 and 2, originally contributed by Pesaran et al. (2000), and have provided a consolidated derivation of these statistics as well as their limiting distributions (cf. Theorems 2 and 3). In particular we hope to have presented the corresponding proofs in a manner accessible to readers who are not familiar with the ubiquitous Johansen (1991) cointegration test - and thus provide a future reference for users of the cointegration tests by Pesaran et al. (2000). The mentioned test statistics serve to extend the analysis of cointegrated systems to the inclusion of exogenous I(1) variables which are assumed to be long-run forcing for the "endogenous" variables. Moreover, we have introduced the diagnostic tests 3, 4 and 5, which help to investigate whether the restrictions on the exogenous terms are correctly specified.

These types of tests may be a useful tool for empirical studies on small open economies, peripheral financial markets and in regional economics. In order to enable empirical analysis, we implemented the test statistics mentioned above in MATLAB.[26]

Further promising directions on the applied side would include implementing the additional tests and methods suggested by Pesaran et al. (2000): Firstly, diagnostic tests on case specification (Pesaran et al., 2000, p.308-310) and, secondly, the possibility for imposing restrictions on short-run parameters $\boldsymbol{\Psi}$ (Ibid., pp.311-313) of the VECM in (20). Moreover, implementing polynomial approximations for the critical values in Pesaran et al. (2000) as developed by MacKinnon et al. (1999), would provide a useful extension.

[26]The corresponding MATLAB routine may be obtained from the author.

# References


Granger, C. W. and Lin, J.-L. (1995). Causality in the long run. *Econometric Theory*, 11(3):530–536.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press, Princeton, NJ.

Harbo, I., Johansen, S., Nielsen, B., and Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics*, 16(4):388–99.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton, NJ.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–80.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Advanced Texts in Econometrics. Oxford University Press, Oxford.

MacKinnon, J. G., Haug, A. A., and Michelis, L. (1999). Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics*, 14(5):563–577.

Pesaran, M. H., Shin, Y., and Smith, R. J. (2000). Structural analysis of vector error correction models with exogenous I(1) variables. *Journal of Econometrics*, 97(2):293–343.

Phillips, P. C. and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, 20(2):971–1001.

# Appendix A.1: Notes on implementation in MATLAB

The central requirement for this first year paper was to implement the crucial parts of Pesaran et al. (2000) in Matlab. While Matlab is a versatile application for solving a wide range of mathematical problems, it does not offer an environment tailored to the manipulation of symmetric, positive semi-definitive matrices. Moreover Matlab is a procedural language that does not try to "outguess' the programmer's procedure. This fosters Matlab's usability, but puts a drag on computational efficiency, since many of Matlab's algorithms try to solve more general problems than the issues in this paper. Therefore our implementation of the Pesaran et al. (2000) test focuses on data manipulation in a way that optimises execution speed.
In particular, this implies avoiding matrix inversion wherever feasible (which may be accomplished by proper factorisation), and minimising the need for $(k \times T)$ times $(T \times k)$ - style matrix multiplication. Crucially, `for-next` loops should be kept to an absolute minimum, since the the execution of iterative statements is considerably faster at the binary level.
The routine `uPSS` is split in four main parts: It starts with (a) user input checks and ends with (d) code concerned with displaying the test results properly. Moreover, there is a subroutine (c) that returns critical values as provided in Pesaran et al. (2000, pp.337-341). These three parts are lengthy but straightforward and will not be discussed here. The central part (b) is to be found in the subroutine `fnc_psscompute`, which comprises the preliminary regressions as on page 10, passes the generalised eigenvalue problem to the subroutine `fnc_geneig` and performs the computations for the statistics outlined in Tests 3, 4 and 5. The projection matrix for the residuals $\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$ are obtained as follows: Factorise the matrix $\Delta\mathbf{Z}_-$ (augmented for deterministic terms according to cases a as on page 12) by singular value decomposition; This decomposition allows to express any $T \times k$ matrix $\mathbf{X}$ of rank $k$ as $\mathbf{USV}'$ (where $\mathbf{S}$ corresponds to the first $k$ columns of the square roots of $\mathbf{XX}'$'s eigenvalues, $\mathbf{U}$ are the eigenvectors of $\mathbf{XX}'$ and $\mathbf{V}$ the eigenvectors of $\mathbf{X}'\mathbf{X}$). Denoting $\mathbf{S}'\mathbf{S} = \mathbf{\Lambda}$, we obtain $\mathbf{US} = \mathbf{U}_{1:k}\mathbf{\Lambda}^{1/2}$, where $\mathbf{U}_{1:k}$ denotes the first $k$ columns of $\mathbf{X}$. Hence the projector $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ may be expressed as follows:

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{USV}'(\mathbf{VS}'\mathbf{SV}')^{-1}\mathbf{VS}'\mathbf{U}' = \mathbf{U}_{1:k}\mathbf{U}'_{1:k}$$

In this manner we construct the projectors for the preliminary regressions yielding $\Delta\hat{\mathbf{Y}}$ and $\hat{\mathbf{Z}}^*_{-1}$, as well as $\Delta\hat{\mathbf{X}}$, $\hat{\mathbf{X}}^*_{-1}$ and $\hat{\mathbf{Z}}^*_{-1,x}$.
Out of this "filtered" data, we construct the matrices $S_{yy}$, $S_{yz}$ and $S_{zz}$. Then, we obtain the corresponding eigenvalues of the function $S(\lambda)$ from (77) by using Cholesky factorisation to convert the generalised eigenvalue problem into one of symmetric matrices: The motivation is that Matlab's eigenvalue routine detects symmetric matrices and consequently uses an abridged version for solving the problem. For that purpose, apply the Cholesky decomposition to

$$\begin{pmatrix} S_{yy} & S_{yz} \\ S_{zy} & S_{zz} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{L}'_{11} & \mathbf{L}'_{21} \\ \mathbf{0} & \mathbf{L}'_{22} \end{pmatrix}$$

Then we may express $S_{zy}S_{yy}^{-1}S_{yz} = \mathbf{L}_{21}\mathbf{L}'_{21}$. Note that the solutions to $S(\lambda)$ are equivalent to the eigenvectors and eigenvalues of $S_{zz}^{-1}S_{zy}S_{yy}^{-1}S_{yz} = S_{zz}^{-1}\mathbf{L}_{21}\mathbf{L}'_{21}$. Now apply the Cholesky factorization once again to: $S_{zz} = \mathbf{L}_z\mathbf{L}'_z$ which implies $S_{zz}^{-1} = \mathbf{L}_z^{-1\prime}\mathbf{L}_z^{-1}$. Thus we are interested in the solutions to $|\lambda\mathbf{I} - \mathbf{L}_z^{-1\prime}\mathbf{L}_z^{-1}\mathbf{L}_{21}\mathbf{L}'_{21}| = 0$, with corresponding eigenvectors $\mathbf{v}$ satisfying $\mathbf{L}_z^{-1\prime}\mathbf{L}_z^{-1}\mathbf{L}_{21}\mathbf{L}'_{21}\mathbf{v} = \lambda\mathbf{v}$. By multiplying from the left with $\mathbf{L}_z'$, and with $\mathbf{L}_z^{-1\prime}$ from the right, we obtain that the solutions $\hat{\lambda}$ are the the same as those to $|\mathbf{I}\lambda - \mathbf{L}_z^{-1}\mathbf{L}_{21}\mathbf{L}_{21}'\mathbf{L}_z^{-1\prime}| = 0$. The corresponding eigenvectors $\mathbf{w}$ consequently satisfy $\mathbf{L}_z^{-1}\mathbf{L}_{21}\mathbf{L}_{21}'\mathbf{L}_z^{-1\prime}\mathbf{w} = \lambda\mathbf{w}$. Comparison illustrates that $\mathbf{v} = \mathbf{L}_z^{-1}\mathbf{w}$. (Note that this implies $\mathbf{v}'S_{zz}\mathbf{v} = \mathbf{I}$.) The MATLAB subroutine `fnc_geneig` solves the generalised eigenvalue problem exactly in the manner described above.[27]

Under the cointegrating rank hypothesis $H_r$ the vector $\boldsymbol{\beta}_*$ is then given by the first $r$ eigenvectors $\mathbf{v}$. Accordingly, the adjustment coefficients $\boldsymbol{\alpha}_y$ are then obtained by plain OLS estimation $\hat{\boldsymbol{\alpha}}_y = S_{yz}\hat{\boldsymbol{\beta}}_*(\hat{\boldsymbol{\beta}}_*'S_{zz}\hat{\boldsymbol{\beta}}_*)^{-1} = S_{yz}\boldsymbol{\beta}_*$.

Finally, note that $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are identified only up to an $r \times r$ non-singular matrix $\mathbf{K}$, i.e. $\boldsymbol{\alpha}_y\boldsymbol{\beta}' = (\boldsymbol{\alpha}_y\mathbf{K}^{-1})(\mathbf{K}\boldsymbol{\beta}')$ for any non-singular matrix $\mathbf{K}$ : Using the normalisation $\hat{\boldsymbol{\beta}}_*'S_{zz}\hat{\boldsymbol{\beta}}_* = \mathbf{I}$ in that respect may hinder interpretation. Rather we prefer to present $\hat{\boldsymbol{\beta}}_*$ in its reduced row echelon form, which is equivalent to choosing $\mathbf{K} = \boldsymbol{\beta}_{*r\times r}^{-1}$ where $\boldsymbol{\beta}_{*r\times r}$ denotes the $r\times r$ upper left sub-matrix of $\boldsymbol{\beta}_*$. The routine `uPSS` provides this normalisation as an additional option.

For further information on implementation, please refer to the comments in the code of `uPSS`.

[27]Note however that by the way we defined $S_{yy}$, $S_{yz}$ and $S_{zz}$, the $m\times m$ matrix $S_{zz}^{-1}S_{zy}S_{yy}^{-1}S_{yz}$ has $k$ zero roots. This causes problems, as the numerical solution by MATLAB may set these eigenvalues to complex numbers close to zero. This fact requires special treatment in the subroutine.

# Appendix A.2: User manual for the MATLAB routine uPSS

## Overview

The MATLAB routine `uPSS` implements the cointegrating rank tests by Pesaran et al. (2000) for a system of $n$ endogenous variables $\mathbf{Y}$, with $k$ exogenous variables $\mathbf{X}$ appearing in the cointegrating relationship. The exogenous variables are assumed to be I(1), not mutually cointegrated and independent of the lagged cointegration relationship with $\mathbf{Y}$. However, the dynamics of $\mathbf{Y}$ may affect $\mathbf{X}$ in the short run.
The routine provides three central statistics

- **Cointegrating rank tests:** Similar to Johansen (1991) the function returns the test statistics for the hypothesis of *cointegrating rank* $r$ of $\mathbf{\Pi}$ in equation (21), $r = 0, 1, \ldots, n$: The trace test statistic (cf. Test 2, p.13) is the likelihood ratio test of the null $H_r$ (cointegrating rank $r$) against $H_n$, while the maximum eigenvalue statistic (cf. Test 1, p.13) is the likelihood ratio test of the null $H_r$ against $H_{r+1}$. The procedure provides the corresponding critical values for the 5% and 10% significance levels and determines the cointegrating rank. (Optionally, the latter feature may be turned off)

- **Cointegrating vectors:** Moreover, `uPSS` determines the *cointegrating vectors* $\boldsymbol{\beta}$ and *adjustment coefficients* $\boldsymbol{\alpha}$ for each hypothesised rank $r$. The user may choose between two normalisations of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$: Either such that $\boldsymbol{\beta}$ is in reduced row echelon form under $H_r$ (this is the default case); or the form where $\boldsymbol{\beta}$ conforms to the generalized eigenvector normalisation (cf. Appendix A.1).

- **Exogeneity Diagnostics:** Finally, `uPSS` applies a separate cointegration rank test on $\mathbf{X}$ (cf. model (22)) to test whether the exogenous series are I(1) and not mutually cointegrated, when adjusting for the short-run impact of $\mathbf{Y}$ (cf. Tests 3 and 4, p.24). Moreover it performs a test on whether the adjustment coefficients for $\mathbf{X}$ with respect to the cointegrating relationship in $\boldsymbol{\beta}$ are different from zero (cf. Test 5, p.25). All of these tests are delivered with corresponding critical values. Optionally, these diagnostics test for $\mathbf{X}$ may be deactivated.

## Inputs

Apart from providing the data matrices $\mathbf{Y}$, and (optionally) $\mathbf{X}$, the routine requires to specify a lag order $p$ as in models (21) and (22). This parameter $p$ is the lag order in levels, hence the VECM model in (21) estimates $p - 1$ autoregressive lagged difference terms in addition to the cointegrating and deterministic parameters.
Moreover, the routine requires to specify one out of five assumptions on the deterministic terms in the cointegrating equation as well as the VAR in the differences[28] (cf. p.7):

- *Case I:* Neither intercepts nor trends in cointegration equation(s) and VAR

- *Case II:* intercept in cointegration equation(s), no deterministic terms in VAR

- *Case III:* intercept in cointegration equation(s) and VAR, no trends

[28] Note that cases III and V slightly differ from the cases by Johansen (1991) provided in most software packages.

- *Case IV:* intercept and trend in cointegration(s) equation, intercept in VAR
- *Case V:* intercept and trend in cointegration(s) equation as well as VAR

In total, there are nine input parameters for `uPSS`, three of which have to be specified in any case:

Table 1: Input parameters for uPSS

| **Parameter** | **Type** | **Description** |
|---|---|---|
| `vYseries` (required) | $T \times n$ matrix or structure | Alternatively a structure `Y` with `Y.data` the data matrix, and `Y.name` a cell string vector containing the series names. |
| `vXseries` (required) | $T \times k$ matrix or structure | type any scalar or string to omit `vXseries`. |
| `lLagOrder` (required) | Positive integer | Specifies the lag order in the levels. |
| `lCase` (optional) default: 5 | Positive integer 1:5 | Integer corresponding to case numbers I-V (cf. p.7); if none is provided or `lCase=nan`, Case V is selected |
| `lSigniLevel` (optional) default: 1 | Integer 0:2 | Significance level, type `1` for 5% critical values (resp. type `95`, `5`); `2` for 10% significance level (resp. type `90`, `10`). Type `0` if `uPSS` should refrain from retrieving critical values (does not affect critical values for Exogeneity Diagnostics) |
| `bBetaAlpha`(optional) default: `true` | Boolean | Type `0` resp. `false` if no $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ should be computed |
| `bExoDiagnostics` (optional) default: `true` | Boolean | If `bExoDiagnostics=0`, then `uPSS` refrains from computing diagnostic tests on $\mathbf{X}$. Not including `vXseries` triggers `bExoDiagnostics=false`. (cf. section 6) |
| `bDisplay` (optional) default: `true` | Boolean | `bDisplay=true` incites `uPSS` to provide a clearly arranged presentation of results. `bDisplay=false` raises speed considerably. |
| `bREF_beta_alpha` (optional) default: `true` | Boolean | `bREF_beta_alpha=true` induces a reduced row echelon form representation of $\boldsymbol{\beta}$, with $\boldsymbol{\alpha}$ adjusted accordingly (cf. Appendix A.1). `bREF_beta_alpha=false` leads to a normalisation $\boldsymbol{\beta}' S_{zz} \boldsymbol{\beta} = \mathbf{I}$ |

## Hints

- The computational parts of `uPSS` have been optimised for maximum speed. However, the final presentation of results has been trimmed to aesthetics rather than speed. If one's aim is speed (as maybe in a sequence of tests), one might consider to turn the visual display off (i.e. set `bDisplay` to `false`).
- For more information on output parameters, type `help uPSS`.
- Note that if you do not provide the exogenous data matrix $\mathbf{X}$, the routine `uPSS` returns nothing else but the Johansen (1991) test.
- If you have not worked with Matlab before, note that the file `uPSS.m` should be saved in your current working directory. You may experiment with the function as follows: Save the additional files `indprod_sa.xls`, `uXl2ml.m` and `structs2m.m` into your working directory, and the execute the following commands: First `uXl2ml('indprod_sa.xls');`, then `uPSS(structs2m(BE,NL),DE,2,5);`. With the sample data on industrial production in `indprod_sa.xls`, this will perform a test on whether Belgium and the Netherlands are cointegrated with Germany as an exogenous series, under lag order $p = 2$, and Case V.