Anders Rahbek
Rasmus Søndergaard Pedersen
University of Copenhagen

September 2024

Part IV
Cointegration Analysis in Vector Autoregressions

IV.1 Introduction

The literature on cointegration is by now enormous, see for example the survey by Johansen (2005), and cointegration analysis is applied everywhere in applied and theoretical time series analysis. This part gives a brief introduction to the rich theory of cointegration analysis in VAR models, with emphasis on analysis of the cointegrated VAR(1) model.

IV.1.1 Futures and no arbitrage

Let K denote the expiry date of a futures contract at time t on the index with spot price S_t, and $F(t, K)$ the price of the future contract. Considering forward pricing as equivalent to futures pricing, the condition for no arbitrage from financial theory can be written as the identity,

$$F(t, K) = S_t \exp(r_{t,K}(t - K))$$

where $r_{t,K}$ is the zero coupon rate for the period between t and expiry date K. With $f_t = \log(F(t, K)/r_{t,K}(t-K))$ and $s_t = \log(S_t)$, the identity reduces to,

$$f_t - s_t = 0.$$

While often applied, the assumption of equivalence of future and forward pricing is actually not correct, essentially due to the different types of settlements of the contracts. It is therefore of interest to see to what extend it may be correct, and to find out if there is an empirical relationship between s_t and f_t.

Italian MIB30 daily data are shown below for the period 29/11-94 - 30/6-97. Previously it was concluded that the spot price s_t seemed to be modelled well by a random walk plus drift type model, that is, as an I(1) process with

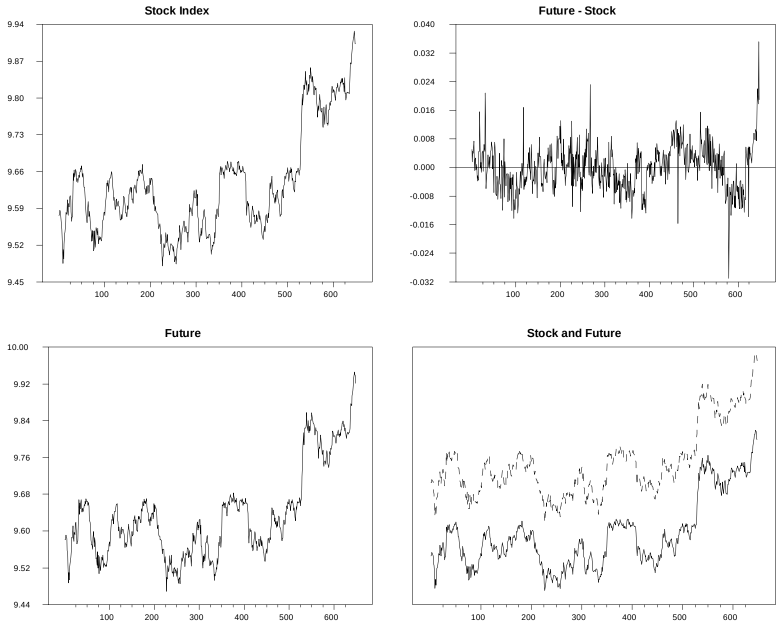

Figure 1: The time series f_t, s_t and $f_t - s_t$.

a linear drift, ignoring the ARCH type misspecification. Hence a VAR model for $X_t = (s_t, f_t)'$ must allow for unit root(s), and, at the same time, for 'identities', or 'stable relations', such as the above. In terms of cointegration, the identity or cointegration relation, is interpreted as $s_t - f_t$ being asymptotically stable or stationary, as opposed to I(1). The series f_t and s_t are shown in Figure 1 below, together with the cointegrating relation as given by the linear combination,

$$\beta' X_t = (1, -1) X_t = f_t - s_t. \tag{IV.1}$$

Clearly, both s_t and f_t resemble random walks, while the cointegrating relation, $\beta' X_t$, seems more stable.

More can be seen by considering the cross-plots of s_t and f_t in Figure 2. As expected the series are clearly correlated. Furthermore, they seem to be pushed, or moved, up and down around the straight line $s_t = f_t$. Essentially, if $|s_t - f_t| > 0$, then $|s_{t+1} - f_{t+1}| < |s_t - f_t|$ as illustrated for a part of the sample, $t = 40, ..., 50$. Thus the 'error' at time t, as measured by $s_t - f_t$, is being corrected such that at time $t + 1$, the spread, $s_{t+1} - f_{t+1}$, is 'smaller'.

Thus a VAR model for X_t must allow for I(1) type non-stationary (integrated) variables with a stable or 'cointegrated' linear combination, $\beta' X_t$. At the same time the dynamics of the model must be such that ΔX_{t+1} can adjust to $\beta' X_t$, that is, it must allow error correction.

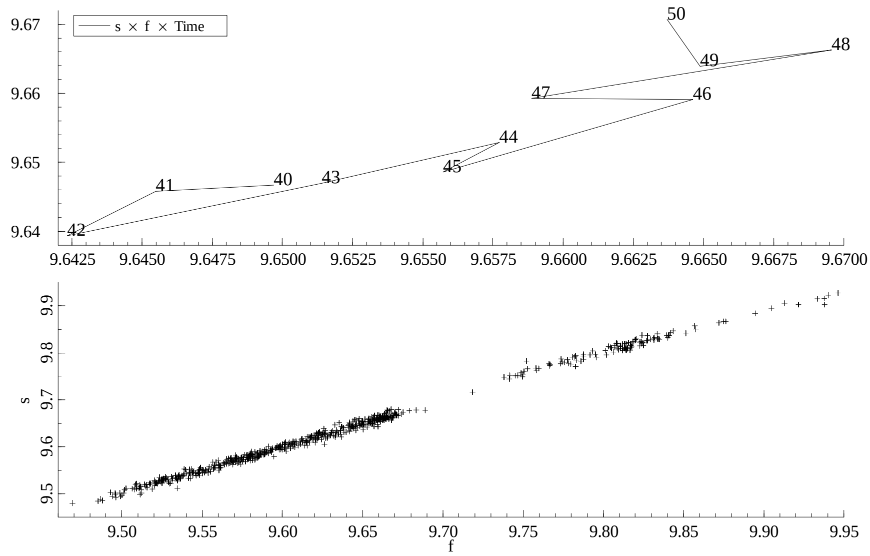

Figure 2: Cross-plots of s_t and f_t for $t = 1, ..., 648$ and $t = 40, ..., 50$.

The cointegrated VAR model allows exactly for this as will be shown below. Moreover, inference in the VAR model makes it possible to determine the number of stable or cointegrated relations β, and also to do hypothesis testing on these.

IV.2 VAR(1) model

Consider the p-dimensional VAR model of order one, VAR(1), model as given by,

$$X_t = AX_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., T \tag{IV.2}$$

with $A \in \mathbb{R}^{p \times p}$, X_0 fixed and ε_t i.i.d. $\mathrm{N}_p(0, \Omega)$ distributed, $\Omega > 0$.

Recall the condition $\rho(A) < 1$ which implies X_t in (IV.2) is geometrically ergodic. The condition may equivalently be stated in terms of the roots of the so-called characteristic polynomial, $A(z) = I - Az$, $z \in \mathbb{C}$,

$$\det(A(z)) = \det(I_p - Az) = 0 \;\Rightarrow\; |z| > 1.$$

Allowing eigenvalues of A at one implies $\rho(A) = 1$, and can be stated in terms of $A(z)$ as

$$\det(A(1)) = \det(I_p - A) = 0.$$

That is, with

$$\Pi = A - I_p,$$

$A(z)$ has one or more roots at $z = 1$ if, and only if, the $(p \times p)$ dimensional matrix Π has reduced rank $r < p$. This is central to the formulation of cointegration.

One may reparameterize the VAR(1) model in terms Π, using $\Delta X_t = X_t - X_{t-1}$,

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \quad t = 1, ..., T, \tag{IV.3}$$

where $\Pi \in \mathbb{R}^{p \times p}$, X_0 is fixed and ε_t are i.i.d.$\mathrm{N}_p(0, \Omega)$.

We denote the hypothesis that Π has rank less than or equal to r by H_r, where $0 \leq r \leq p$. It follows that under H_r the $p \times p$ matrix Π can be factorized as

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

where α, β are $p \times r$ matrices, see Lemma B.1 in the appendix. The following two examples illustrate this:

Example IV.2.1 *Consider the case of a* 2×2 *matrix* Π *of rank* $r = 1$*, given by*

$$\Pi = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

Simple calculations show,

$$\Pi = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}' = \alpha\beta'.$$

Note that β*, and hence* α*, are unique up to a normalization as*

$$\Pi = (\alpha m)(\beta m^{-1})',$$

for any $m \in \mathbb{R}$*,* $m \neq 0$*. Stated differently,* α *spans the column space of* Π, $\mathrm{sp}(\Pi)$*, while* β *spans the row space of* Π, $\mathrm{sp}(\Pi')$*.*

Example IV.2.2 *Consider next the case of a* 3×3 *matrix* Π *of rank* $r = 2$*,*

$$\Pi = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}' = \alpha\beta',$$

where as before, $\mathrm{sp}(\alpha) = \mathrm{sp}(\Pi)$*,* $\mathrm{sp}(\beta) = \mathrm{sp}(\Pi')$*. In particular,* $\Pi = (\alpha m')(\beta m^{-1})'$*, with* m *any* 2×2 *dimensional invertible matrix.*

IV.2.1 Stochastic behavior of X_t

As illustrated, the fact that the characteristic polynomial has one or more roots at $z = 1$, implies that the matrix parameter Π has reduced rank, and hence that $\Pi = \alpha\beta'$ for some $(p \times r)$-dimensional matrices α and β. For α and β of full rank r, we define their orthogonal complements $\alpha_\perp$ and $\beta_\perp$. These are $(p \times (p - r))$-dimensional matrices of full rank $(p - r)$, for which $\alpha'_\perp \alpha = \beta'_\perp \beta = 0$, $\det(\alpha, \alpha_\perp) \neq 0$, and $\det(\beta, \beta_\perp) \neq 0$.

Consider next what this implies for the stochastic behavior of X_t.

Example IV.2.3 *Consider the 2-dimensional VAR(1) process X_t as given by,*

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t, \tag{IV.4}$$

where $\alpha = (-1, 0)'$, $\beta' = (1, -1)$ and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ i.i.d.$N_2(0, \Omega)$. Note initially that, with $\alpha_\perp = (0, 1)'$,

$$\alpha'_\perp \Delta X_t = \Delta X_{2t} = \alpha'_\perp (\alpha\beta' X_{t-1} + \varepsilon_t) = \alpha'_\perp \varepsilon_t = \varepsilon_{2t}.$$

That is, $X_{2t} = \sum_{i=1}^t \varepsilon_{2i} + X_{20}$. Equivalently, X_{2t} is the sum of a random walk and the initial value, and X_{2t} is an I(1) process.

Next, note that

$$\beta' X_t = X_{1t} - X_{2t} = \beta' \varepsilon_t = \varepsilon_{1t} - \varepsilon_{2t},$$

that is, the difference, or 'spread', between X_{1t} and X_{2t} is i.i.d. Gaussian, and therefore, in particular, geometrically ergodic with a stationary solution (often simply referred to as stationary).

Collecting terms, we find

$$\begin{pmatrix} X_{1t} - X_{2t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} X_t = \begin{pmatrix} \varepsilon_{1t} - \varepsilon_{2t} \\ \sum_{i=1}^t \varepsilon_{2i} + X_{20} \end{pmatrix},$$

or, re-organizing terms,

$$\begin{aligned} X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} &= \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon_{1t} - \varepsilon_{2t} \\ \sum_{i=1}^t \varepsilon_{2i} + X_{20} \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \sum_{i=1}^t \varepsilon_{2t} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} (\varepsilon_{1t} - \varepsilon_{2t}) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} X_{02}. \end{aligned}$$

As $\beta_\perp = (1, 1)'$, this may be stated as.

$$X_t = \beta_\perp \alpha'_\perp \sum_{i=1}^t \varepsilon_t + \alpha\beta' \varepsilon_t + \beta_\perp \alpha'_\perp X_0. \tag{IV.5}$$

Thus X_t has a representation as a sum of a "common trend" (i.e. the random walk $\alpha'_\perp \sum_{i=1}^t \varepsilon_t$), a "stationary" (geometrically ergodic) process, $\beta'\varepsilon_t$, and the initial value X_0. Hence X_t is a non-stationary I(1) process.

Note that, in line with the mentioned example of spot and futures rates, where $X_t = (s_t, f_t)'$, X_t is error correcting to $\beta' X_{t-1}$ with adjustment vector α, which can be seen by rewriting (IV.4) as,

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t. \tag{IV.6}$$

Simulating a process with these parameter values will give a cross plot of the form in Figure 2. More precisely, by the first factor $\beta_\perp \alpha'_\perp \sum_{i=1}^t \varepsilon_t$ in (IV.5), the points (X_{1t}, X_{2t}) will be moved up and down by $\alpha'_\perp \sum_{i=1}^t \varepsilon_t$ on the attractor given by $\beta' X_t = 0$, or along the line spanned by $\beta_\perp = (1, -1)'$. The movements away from, or around, the attractor, as given by $\alpha\beta'\varepsilon_t$, are then error corrected by the term $\alpha\beta' X_{t-1}$ in (IV.6).

To generalize Example IV.2.3 to the case where the cointegrating relations $\beta' X_t$ are not i.i.d. the following assumption plays a key role throughout:

Assumption IV.2.1 *Let X_t be a vector autoregressive process with characteristic polynomial $A(z)$, $z \in \mathbb{C}$. Assume that $A(z)$ has exactly $(p-r)$ roots at $z = 1$ and the remaining roots are larger than one in absolute value, $|z| > 1$.*

Remark IV.2.1 *The condition may equivalently be stated for the VAR(1) process as A has $(p-r)$ eigenvalues equal to one, while the remaining eigenvalues are smaller than one in absolute value, $\rho(I_r + \beta'\alpha) < 1$; see also the proof of Proof of Theorem IV.2.1.*

One has the following theorem which is a version of the so-called 'Granger representation theorem', see also Johansen (1996).

Theorem IV.2.1 *Consider the VAR(1) process given by (IV.3) under the hypothesis*

$$H_r : \Pi = \alpha\beta', \quad \alpha, \beta \in \mathbb{R}^{p \times r}. \tag{IV.7}$$

Under Assumption IV.2.1, X_t is an I(1) process, and has the representation,

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + C_S S_t + C_0, \tag{IV.8}$$

where $C = \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp$ is a $p \times p-$dimensional matrix of rank $(p-r)$, and $C_S = \alpha(\beta'\alpha)^{-1}$. Moreover, C_0 depends on the initial values and is given by $C_0 = CX_0$. The r cointegrating relations $\beta' X_t = S_t$ are geometrically ergodic. In particular, the initial values $\beta' S_0$ can be given an initial distribution such that S_t has the stationary representation, $S_t^* = \sum_{i=0}^\infty (I + \beta'\alpha)^i \beta'\varepsilon_i$.

Remark IV.2.2 *Recall that by the geometric ergodicty, the LLN and CLT apply to functions of $\{S_t\}$ as will be used throughout.*

The theorem states that when Π has rank r, and if there are exactly $(p-r)$ unit roots, and the remaining roots correspond to the asymptotically stable case, the p dimensional process X_t is non-stationary and I(1). Moreover, it has r cointegrating relations, $\beta' X_t$, which are geometrically ergodic.

Assumption IV.2.1 is indeed vital for the theorem to hold. For example, if there are exactly $p-r$ roots at $z=1$, but all, or some of, the remaining roots are smaller than one in absolute value, $\beta' X_t$ is an explosive process, and hence not 'cointegrating'. Also, by allowing for more than $p-r$ roots at $z=1$, the process may be integrated of order two, instead of one. *In other words, the assumption of reduced rank r of Π is a necessary, but not sufficient, condition for the VAR process to be an I(1) process which is cointegrated.*

Note also that under H_p, the theorem states that X_t is asymptotically stable, provided that all roots of the characteristic polynomial are larger than one in absolute value, which is in accordance with the previous theory for geometrically ergodic VAR processes.

Proof of Theorem IV.2.1: The arguments are identical to the ones used in Example IV.2.3: Note first, that by definition, $\alpha_\perp' \Delta X_t = \alpha_\perp' \varepsilon_t$, and hence, $\alpha_\perp' X_t = \alpha_\perp' \sum_{i=1}^t \varepsilon_i + \alpha_\perp' X_0$. Next,

$$\beta' X_t = (I_r + \beta'\alpha)\, \beta' X_{t-1} + \beta' \varepsilon_t, \tag{IV.9}$$

which is geometricallty ergodic provided $\rho\,(I_r + \beta'\alpha) < 1$. This holds by Assumption IV.2.1, as

$$\det(A(z)) = 0 \Leftrightarrow \det(I_r - (I_r + \beta'\alpha)z) \det((1-z)\, I_{p-r}) = 0, \tag{IV.10}$$

which follows by pre- and post multiplying $A(z)$ by $(\beta, \beta_\perp)'$ and $(\beta, \beta_\perp)$ respectively.

The final result holds by using the (skew-projection) identity

$$I_p = \alpha\,(\beta'\alpha)^{-1} \beta' + \beta_\perp (\alpha_\perp' \beta_\perp)^{-1} \alpha_\perp' = C_S \beta' + C, \tag{IV.11}$$

such that X_t has the decomposition,

$$X_t = C X_t + C_S \beta' X_t = C X_t + C_S S_t, \tag{IV.12}$$

as claimed. That (IV.11) holds, follows by $\beta'\alpha$ having full rank r as implied by (IV.10). The full rank of $\beta'\alpha$ implies that $\alpha_\perp' \beta_\perp$ has full rank $p-r$, and that $(\beta, \alpha_\perp)$ has full rank. Multiplying in (IV.11) by $(\beta, \alpha_\perp)'$ from the left establishes the identity. $\square$

IV.2.2 Econometric analysis

Next, turn to ML estimation of the parameters in the VAR(1) model in (IV.3) under H_r, given by the equations,

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t,$$
$$H_r : \Pi = \alpha\beta'.$$

In terms of the usual regression model with $Y_t = \Delta X_t$ and $Z_t = X_{t-1}$, the hypothesis H_r of reduced rank of Π is a nonlinear restriction, and the previous linear regression results cannot be used. Instead the following result holds:

Theorem IV.2.2 *Under H_r, the ML estimators of α, β and Ω are given by,*

$$\hat{\alpha} = S_{yz}\hat{\beta}(\hat{\beta}'S_{zz}\hat{\beta})^{-1}, \tag{IV.13}$$

$$\hat{\Omega} = S_{yy\cdot\hat{\beta}} = S_{yy} - S_{yz}\hat{\beta}(\hat{\beta}'S_{zz}\hat{\beta})^{-1}\hat{\beta}'S_{zy}. \tag{IV.14}$$

Here $Y_t = \Delta X_t$ and $Z_t = X_{t-1}$ in the product moment matrices, such that e.g. $S_{yz} = T^{-1}\sum_{t=1}^{T} Y_t Z_t$. The MLE $\hat{\beta}$ is found by solving the eigenvalue problem

$$\det\left(\lambda S_{zz} - S_{zy}S_{yy}^{-1}S_{yz}\right) = 0, \tag{IV.15}$$

with eigenvalues $1 > \hat{\lambda}_1 > ... > \hat{\lambda}_r > ...\hat{\lambda}_p > 0$ and corresponding eigenvectors $\hat{V} = (\hat{v}_1, ..., \hat{v}_p)$, for which $\hat{V}'S_{zz}\hat{V} = I_p$ and $\hat{V}'S_{zy}S_{yy}^{-1}S_{yz}\hat{V} = \hat{\Lambda}$, where $\hat{\Lambda} = diag\left(\hat{\lambda}_1, ..., \hat{\lambda}_p\right)$. Then

$$\hat{\beta} = (\hat{v}_1, ..., \hat{v}_r), \tag{IV.16}$$

and the maximized likelihood function is given by,

$$L(\hat{\alpha}, \hat{\beta}, \hat{\Omega}) = (\det(S_{yy})\prod_{i=1}^{r}(1-\hat{\lambda}_i))^{-T/2}. \tag{IV.17}$$

This is an example of so-called reduced rank regression, RRR, which was developed in Anderson (1951), and the ML estimators in Theorem IV.2.2 are said to be found by RRR of ΔX_t on X_{t-1}. Note that the expressions in (IV.13) and (IV.14) show that $\hat{\alpha}$ and $\hat{\Omega}$ are found by OLS regression with β known, and the nonlinear restriction implies that in order to find $\hat{\beta}$ an eigenvalue problem has to be solved.

Proof: Under H_r the log-likelihood function is, apart from a constant, given by

$$\log L(\alpha, \beta, \Omega) = -\frac{T}{2}\log\det(\Omega) - \frac{1}{2}\sum_{t=1}^{T}(\Delta X_t - \alpha\beta' X_{t-1})'\Omega^{-1}(\Delta X_t - \alpha\beta' X_{t-1}). \tag{IV.18}$$

With $\beta \in \mathbb{R}^{p\times r}$ fixed, this is the likelihood function for the usual linear regression model, and $\hat{\alpha}(\beta)$ and $\hat{\Omega}(\beta)$ are found by OLS regression of $Y_t = \Delta X_t$ on $\beta' Z_t = \beta' X_{t-1}$,

$$\hat{\alpha}(\beta) = S_{yz}\beta(\beta' S_{zz}\beta)^{-1}, \tag{IV.19}$$

$$\hat{\Omega}(\beta) = S_{yy\cdot\beta} = S_{yy} - S_{yz}\beta(\beta' S_{zz}\beta)^{-1}\beta' S_{zy}. \tag{IV.20}$$

Inserting these in $L(\alpha, \beta, \Omega)$ gives the concentrated likelihood function,

$$\log L(\beta, \hat{\alpha}(\beta), \hat{\Omega}(\beta)) = -\frac{T}{2}\det\left(\hat{\Omega}(\beta)\right). \tag{IV.21}$$

Next, exploiting the determinant of a block-matrix gives,

$$\det\begin{pmatrix} S_{yy} & S_{yz}\beta \\ \beta' S_{zy} & \beta' S_{zz}\beta \end{pmatrix} = \det(S_{yy\cdot\beta})\det(S_{\beta\beta}) = \det(S_{yy})\det(S_{\beta\beta\cdot y}), \tag{IV.22}$$

where the index β refers to $\beta' Z_t$, such that e.g. $S_{\beta\beta\cdot y} = \beta' S_{zz}\beta - \beta' S_{zy} S_{yy}^{-1} S_{yz}\beta$. Use this to see that, as $\hat{\Omega}(\beta) = S_{yy\cdot\beta}$, $\hat{\beta}$ is found by minimization of,

$$\det\left(\hat{\Omega}(\beta)\right) = \det(S_{yy})\frac{\det(S_{\beta\beta\cdot y})}{\det(S_{\beta\beta})} = \det(S_{yy})\frac{\det(\beta' S_{zz\cdot y}\beta)}{\det(\beta' S_{zz}\beta)}. \tag{IV.23}$$

That this ratio of quadratic forms is minimized by solving the eigenvalue problem in (IV.15), and setting $\hat{\beta} = (\hat{v}_1, ..., \hat{v}_r)$, that is the eigenvectors corresponding to the r largest eigenvalues, follows by Johansen (1996, Theorem 6.1). □

IV.2.3 Hypothesis testing

Hypothesis testing for two hypotheses will be discussed here: First the hypothesis of Π having reduced rank r, and next, given that the rank r is known, linear restrictions on the cointegration vectors β. For other kinds of hypotheses, see Section IV.3 for a discussion of the asymptotic properties of $\hat{\alpha}$ and $\hat{\beta}$.

Turn first to the hypothesis of Π having reduced rank r.

IV.2.3.1 Rank test

By definition, the LR test statistic for the hypothesis H_r against the unrestricted model, H_p, is given by $\mathrm{LR}_r = -2\log Q$, where Q is the ratio of the maximized likelihood functions. This equals, by Theorem IV.2.2,

$$\mathrm{LR}_r \equiv \mathrm{LR}(H_r|H_p) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i). \tag{IV.24}$$

That is, the computation of the test statistics LR_i for $i = 0, 1, ..., p-1$ is based on solving one, and only one, eigenvalue problem. Now to address the limiting distribution of the test statistic LR_r, the properties of the process X_t are needed. By Theorem IV.2.1, X_t has a representation as a sum of a random walk $\sum_{i=1}^{t} \varepsilon_i$, and a geometrically ergodic process. Under the assumptions in Theorem IV.2.1,

$$\frac{1}{\sqrt{T}} X_{[Tu]} \xrightarrow{D} C\mathcal{B}_u, \tag{IV.25}$$

as $T \to \infty$, by application of the invariance principle as in the univariate case. Here $\mathcal{B}$ is a $(p-r)$ dimensional Brownian motion with variance Ω, while $C\mathcal{B}$ has variance,

$$\Omega_\infty = C\Omega C'. \tag{IV.26}$$

The $(p \times p)$ variance Ω_∞ is referred to as the long-run variance of X_t. By Theorem IV.2.1 $C = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'$, and hence Ω_∞ has rank $p - r < p$ and it is therefore *singular.* This corresponds to the assumption of exactly $p-r$ unit roots, and reflects that the $(p-r)$ dimensional random walk, or the $p-r$ common trends, $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$ enter X_t by the $(p \times (p-r))$ dimensional coefficient matrix, $\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}$.

Thus, in addition to the assumption H_r of rank Π less than or equal to r, the additional assumptions in Theorem IV.2.1 are vital for interpreting the (limiting) behavior of X_t. Denote therefore by H_r^0 the hypothesis H_r where the additional assumptions hold, i.e. that $A(z)$ has exactly $p-r$ roots at $z = 1$, while the remaining roots are larger than one in absolute value.

The following result holds:

Theorem IV.2.3 *Under H_r^0, it holds that*

$$LR_r \xrightarrow{D} DF_{p-r}(\mathcal{W}), \tag{IV.27}$$

where $\mathcal{W}$ is a $p-r$ dimensional standard Brownian motion, and

$$DF_{p-r}(\mathcal{W}) = \text{tr}\{\int_0^1 d\mathcal{W}\mathcal{W}'(\int_0^1 \mathcal{W}_u\mathcal{W}_u' du)^{-1} \int_0^1 \mathcal{W}d\mathcal{W}'\}. \tag{IV.28}$$

The limiting distribution $\text{DF}_1(\mathcal{W})$ is identical to the limiting distribution of the LR test for a unit root in the univariate AR model, see Theorem **??** and $\text{DF}_{p-r}(\mathcal{W})$ is the multivariate generalization.

Some quantiles of $\text{DF}_{p-r}(\mathcal{W})$ for different values of $p-r$ are reported in Section IV.5.4.

Proof: Consider here the case of $r = 0$ only, see Johansen (1996) for the general case. That $r = 0$ is equivalent to $\Pi = 0$ which is a linear restriction.

That is, as for the linear regression model with $Y_t = \Delta X_t$ and $Z_t = X_{t-1}$, the LR_0 test statistic can be written as,

$$\mathrm{LR}_0 = -T \log\det\left(I_p - \tilde{W}_T\right), \quad \tilde{W}_T = S_{yy}^{-1}\hat{\Pi} S_{zz}\hat{\Pi}' = S_{yy}^{-1}S_{yz}S_{zz}^{-1}S_{zy}. \tag{IV.29}$$

Equivalently, this also follows by noting that by definition $\mathrm{LR}_0 = -T\sum_{i=1}^{p}\log(1-\hat{\lambda}_i) = -T\log\det\left(I - \hat{\Lambda}\right)$, where $\hat{\Lambda}$ is given in Theorem IV.2.2. Next, under $H_0^0 = H_0$, $\Delta X_t = \varepsilon_t$ and $X_t = \sum_{i=1}^{t}\varepsilon_i + X_0$. Hence,

$$S_{yy} = \frac{1}{T}\sum_{t=1}^{T}\Delta X_t \Delta X_t' \xrightarrow{P} \Omega, \tag{IV.30}$$

by the LLN for i.i.d. variables. By the FCLT for i.i.d. variables, and as in Theorem **??**,

$$\frac{1}{\sqrt{T}}X_{[Tu]} \xrightarrow{D} \mathcal{B}_u, \tag{IV.31}$$

$$T^{-1}S_{zz} = T^{-2}\sum_{t=1}^{T}X_{t-1}X_{t-1}' \xrightarrow{D} \int_0^1 \mathcal{B}_u\mathcal{B}_u' du \quad \text{and} \tag{IV.32}$$

$$S_{zy} = T^{-2}\sum_{t=1}^{T}X_{t-1}\Delta X_t' \xrightarrow{D} \int_0^1 \mathcal{B}d\mathcal{B}', \tag{IV.33}$$

where $\mathcal{B}$ is a Brownian motion on $[0,1]$ with covariance Ω. Thus $\mathrm{tr}\{\tilde{W}_T\} = O_P(T^{-1})$, and a stochastic Taylor expansion of $\log\det\left(I - \tilde{W}_T\right)$ gives,

$$\mathrm{LR}_0 = -T\log\det\left(I_p - \tilde{W}_T\right) = \mathrm{tr}\{T\tilde{W}_T\} + o_P(1) \xrightarrow{D} \tag{IV.34}$$

$$\mathrm{tr}\{\Omega^{-1}\int_0^1 d\mathcal{B}\mathcal{B}'(\int_0^1 \mathcal{B}\mathcal{B}'du)^{-1}\int_0^1 \mathcal{B}d\mathcal{B}'\} \overset{D}{=} \mathrm{DF}_p(\mathcal{W}), \tag{IV.35}$$

where $\mathcal{W} = \Omega^{-1/2}\mathcal{B}$, is a p dimensional standard Brownian motion. □

When testing for the rank r we proceed as follows:

Considering the test statistic LR_r, the limiting distribution will be different from the one in Theorem IV.2.3 if H_r^0 does not hold. In particular, H_r^0 states that the rank of Π equals r, while H_r states that the rank is less than or equal to r. Thus the limiting distribution of LR_r will be different for each H_i^0 for $i = 0, 1, .., r$ which are all nested in the H_r hypothesis. This problem is avoided by using a sequential testing procedure to find the rank $\hat{r}$.

The idea is to compute LR_i for all $i = 0, 1, ..., p-1$. Then the rank r is accepted, $\hat{r} = r$, provided that for the test statistics concerning lower ranks it holds that, $\mathrm{LR}_0 > c_0$, $\mathrm{LR}_1 > c_1, ...,\mathrm{LR}_{r-1} > c_{r-1}$, while $\mathrm{LR}_r < c_r$, where c_j are the critical values corresponding to, say, the 95% quantile of $\mathrm{DF}_{p-j}(\mathcal{W})$ in Theorem IV.2.3. This way rank r is accepted, if previous smaller ranks have all been rejected. The idea is that, when evaluating LR_r, all previous cases have been rejected, and hence H_{r-1}, or rank Π less than or equal to $r-1$, does not hold. It is therefore sufficient to consider LR_r under H_r^0, where the limiting distribution is given in Theorem IV.2.3.

Example IV.2.4 *For the spot and futures data, analysis of a bivariate VAR model with $X_t = (s_t, f_t)'$ gives $LR_0 > c_0$, and hence rank $r = 0$ is rejected. Next, $LR_1 < c_1$ and H_1 is accepted. As H_0, or rank equal to zero, was rejected, accepting H_1, or rank less than or equal to 1, implies $\hat{r} = 1$.*

IV.2.3.2 Linear hypotheses on β

Next, with the rank r given, consider here linear hypotheses on β of the form,

$$H_{\mathrm{lin}} : \beta = H\varphi, \tag{IV.36}$$

where H is a $p \times s$ dimensional known matrix, while φ is a $s \times r$ dimensional matrix with freely varying parameters, with $r \leq s \leq p$. The hypothesis is equivalent to

$$H_{\mathrm{lin}} : R'\beta = 0, \tag{IV.37}$$

where $R = H_\perp$ and is a $p \times (p-s)$ dimensional matrix. For example, this kind of hypothesis allows for testing if a variable X_{it} can be excluded in all cointegrating relations, by setting $R' = (0, ..0, 1, 0, ..., 0)$ with the '1' in place i.

Example IV.2.5 *Continuing with the $X_t = (s_t, f_t)'$ example, with $r = 1$, the hypothesis that, say, f_t does not enter in the cointegrating relation can be formulated as*

$$\beta = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \varphi = H\varphi. \tag{IV.38}$$

Rewriting the VAR(1) under H_{lin},

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t = \alpha\varphi' H' X_{t-1} + \varepsilon_t, \tag{IV.39}$$

shows that α, φ and Ω can be found as in Theorem IV.2.2 by setting $Y_t = \Delta X_t$ as there, while $Z_t = H' X_{t-1}$ and $\beta = H\varphi$. That is, RRR of ΔX_t on $H' X_{t-1}$ gives the ML estimators $\hat{\alpha}$, $\hat{\varphi}$ and $\hat{\Omega}$ in this case.

Using this, it follows directly that the likelihood ratio test statistic of H_{lin} against H_r is given by,

$$\text{LR}_{\text{lin}} = \text{LR}(H_{\text{lin}}|H_r) = T\sum_{i=1}^{r}\log(\frac{1-\hat{\lambda}_i}{1-\tilde{\lambda}_i}), \tag{IV.40}$$

which is asymptotically χ^2 distributed with $r\,(p-s)$ degrees of freedom. Here the eigenvalues $\hat{\lambda}_i$ solve the eigenvalue problem in (IV.15) with $Z_t = X_{t-1}$, while $\tilde{\lambda}_i$ solve the eigenvalue problem with $Z_t = H'X_{t-1}$, corresponding to ML estimation under H_{lin}.

That LR_{lin} is asymptotically χ^2 distributed may seem somewhat surprising. It is an implication of the result that under H_r^0, $\hat{\beta}$, appropriately normalized, is super-consistent and asymptotically *mixed* Gaussian distributed, while $\hat{\alpha}$ is consistent at the standard rate and asymptotically Gaussian, see Section IV.3.

In general, hypotheses of the form,

$$\beta = (\beta_1, ..., \beta_r) = (H_1\varphi_1, ..., H_r\varphi_r), \tag{IV.41}$$

that is where each cointegrating vector β_i is restricted by a linear restriction given by H_i, are of interest. Estimation, and asymptotic distributions of the LR tests, are not discussed here. However, observe that the hypothesis of a single variable X_{it} being stationary, or asymptotically stable, can be written in exactly this form by setting $H_i = (0, ..0, 1, 0, ..., 0)'$ with the '1' in place i, and leaving the other cointegrating vectors unrestricted. Moreover, in the simple case of $r = 1$, this collapses to a restriction of the form H_{lin} discussed above.

Example IV.2.6 *Continuing with the $X_t = (s_t, f_t)'$ example, with $r = 1$, the hypothesis that, $s_t - f_t$ is a cointegrating relation can be formulated as*

$$\beta = \begin{pmatrix} 1 \\ -1 \end{pmatrix}\varphi = H\varphi. \tag{IV.42}$$

IV.3 Asymptotics for the MLEs of β and α

Having derived the MLEs for α and β given in Theorem IV.2.2, and secondly being able to test for cointegration (and the order r) from the result in Theorem IV.2.3, we now discuss the asymptotic properties of $\hat{\alpha}$ and $\hat{\beta}$ under H_0^r.

Before discussing distributional theory for $\hat{\alpha}$ and $\hat{\beta}$, consider first the issue of identification. As mentioned in the previous section the sub-spaces sp (β)

and $\mathrm{sp}(\alpha)$ are identified, which follows by noting that with m any $(r \times r)$ matrix of full rank r, we have

$$\Pi = \alpha\beta' = [\alpha m'] \left[\left(\beta m^{-1}\right)' \right] = \alpha_m \beta_m',$$

with $\beta_m = \beta m^{-1}$ and $\alpha_m = \alpha m'$. Thus, while the spaces spanned by (the columns of) α and β are identified, the individual parameters in α and β are not identified. To identify these, some normalization as e.g. given by a specific known choice of m can be imposed.

For example, with $p = 2$ and $r = 1$, we have for $\beta = (\beta_1, \beta_2)'$. With $m = \beta_1$, we get

$$\beta_m = \begin{pmatrix} 1 \\ b \end{pmatrix}, \quad \alpha_m = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix},$$

such that three parameters b, a_1 and a_2 are identified. That is, while the 4 parameters in β and α are not identified, the 3 parameters in the new parametrization in terms of b, a_1 and a_s identified, where $b = \beta_2/\beta_1$ and $a_1 = \alpha_1\beta_1, a_2 = \alpha_2\beta_1$. We say that β (and hence α) are identified by the normalization with $m = \beta_1$, or simply, by $\beta_1 = 1$.

For general p and r, often m is chosen as $m = c'\beta$ with c some known $(p \times r)$ matrix. Corresponding to the simple case above, consider $c = (I_r, 0)'$, such that

$$\beta_m = \beta \left(c'\beta\right)^{-1} = \begin{pmatrix} I_r \\ b \end{pmatrix},$$

where b $((p - r) \times r)$ and α_m $(p \times r)$ are identified. Note that, as $m = c'\beta$, this requires in particular $c'\beta$ to have full rank.

IV.3.1 Preliminary considerations

Consider again the simple case of $p = 2$ and $r = 1$, with normalisation $m = c'\beta$, $c = (1, 0)'$, and hence

$$\beta_m = (1, b)' \text{ and } \alpha_m = (a_1, a_2)'.$$

The VAR model can then be stated in terms of the new identified parameters α_m and β_m (that is, b) as,

$$\Delta X_t = \alpha_m \beta_m' X_{t-1} + \varepsilon_t$$

Consider the MLE $\hat{b}$, or equivalently, $\hat{\beta}_m = (1, \hat{b})'$ with Ω and α_m fixed for simplicity. By definition, $\hat{b}$ satisfies

$$\partial \log L\left(\alpha_m, \beta_m, \Omega\right) / \partial b|_{\hat{\beta}_m} = 0,$$

where

$$\log L(\alpha_m, \beta_m, \Omega) = -\frac{1}{2}\left[T \log\det(\Omega) + \sum_{t=1}^{T}(\Delta X_t - \alpha_m\beta_m' X_{t-1})'\Omega^{-1}(\Delta X_t - \alpha_m\beta_m' X_{t-1})\right].$$

Using $\partial(\beta_m' X_{t-1})/\partial b = \partial(1 + bX_{2t-1})/\partial b = X_{2t-1}$, simple differentiation gives

$$\partial \log L(\alpha_m, \beta_m, \Omega)/\partial b|_{\hat{\beta}_m} = \sum_{t=1}^{T} X_{2t-1}\alpha_m'\Omega^{-1}(\Delta X_t - \alpha_m\hat{\beta}_m' X_{t-1})$$

Next, insert $\Delta X_t = \alpha_m\beta_m' X_{t-1} + \varepsilon_t$, such that we get

$$\begin{aligned}\partial \log L(\alpha_m, \beta_m, \Omega)/\partial b|_{\hat{\beta}_m} &= \sum_{t=1}^{T} X_{2t-1}\alpha_m'\Omega^{-1}(\varepsilon_t - \alpha_m(\hat{\beta}_m - \beta_m)' X_{t-1}) \\ &= \sum_{t=1}^{T} X_{2t-1}\alpha_m'\Omega^{-1}(\varepsilon_t - \alpha_m(\hat{b} - b)X_{2t-1})\end{aligned}$$

We conclude that $\partial \log L(\alpha_m, \beta_m, \Omega)/\partial b|_{\hat{\beta}_m} = 0$ is equivalent to,

$$\alpha_m'\Omega^{-1}\sum_{t=1}^{T}\varepsilon_t X_{2t-1} = \left(\alpha_m'\Omega^{-1}\alpha_m\right)(\hat{b} - b)\sum_{t=1}^{T} X_{2t-1}^2$$

such that solving for $\hat{b} - b$ gives,

$$\begin{aligned}\hat{b} - b &= \left(\alpha_m'\Omega^{-1}\alpha_m\right)^{-1}\alpha_m'\Omega^{-1}\sum_{t=1}^{T}\varepsilon_t X_{2t-1}\left[\sum_{t=1}^{T} X_{2t-1}^2\right]^{-1} \\ &= \left[\sum_{t=1}^{T} X_{2t-1}^2\right]^{-1}\sum_{t=1}^{T} X_{2t-1}\varepsilon_t'\Omega^{-1}\alpha_m\left(\alpha_m'\Omega^{-1}\alpha_m\right)^{-1} \qquad \text{(IV.43)}\end{aligned}$$

Observe that the stochastic behaviour of $\hat{b}$ is given by the limiting behaviour of the two key quantities,

$$\sum_{t=1}^{T}\varepsilon_t X_{2t-1}, \quad \sum_{t=1}^{T} X_{2t-1}^2.$$

For the properties of X_{2t}, note initially that by (IV.8) in Theorem IV.2.1, it follows directly that X_{2t} has the representation,

$$X_{2t} = (0,1)\,X_t = (0,1)\left[C\sum_{i=1}^{t}\varepsilon_i + C_S S_t + C_0\right],$$

where $C = \beta_{m\perp} \left(\alpha'_{m\perp}\beta_{m\perp}\right)^{-1} \alpha'_{m\perp}$. By defintion of β_m, we may choose

$$\beta_{m\perp} = \begin{pmatrix} -b \\ 1 \end{pmatrix},$$

such that $(0,1)\,\beta_{m\perp} = 1 \neq 0$ and hence $(0,1)\,C \neq 0$, implying that $X_{2t} = (0,1)\,X_t$ is I(1). Using the previous results for convergence to the Brownian motion in (IV.25), we find with $u \in (0,1)$,

$$\begin{aligned}\frac{1}{\sqrt{T}} X_{2[Tu]} &= \frac{1}{\sqrt{T}} (0,1)\, X_{[Tu]} \\ &= \left(\alpha'_{m\perp}\beta_{m\perp}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \alpha'_{m\perp}\varepsilon_i + o_p(1) \\ &\xrightarrow{D} \gamma'\mathcal{B}_u, \qquad \gamma = \alpha_{m\perp} \left(\beta'_{m\perp}\alpha_{m\perp}\right)^{-1}.\end{aligned}$$

Multiplying $(\hat{b} - b)$ by T, we get

$$T\left(\hat{b} - b\right) = \left[T^{-2}\sum_{t=1}^{T} X_{2t-1}^2\right]^{-1} T^{-1}\sum_{t=1}^{T} X_{2t-1}\varepsilon'_t\Omega^{-1}\alpha_m \left(\alpha'_m\Omega^{-1}\alpha_m\right)^{-1}$$

and hence, collecting terms,

$$T\left(\hat{b} - b\right) \xrightarrow{D} \left[\int_0^1 \gamma'\mathcal{B}_u\mathcal{B}'_u\gamma du\right]^{-1} \int_0^1 \gamma'\mathcal{B}_u d\mathcal{B}'_u\Omega^{-1}\alpha_m \left(\alpha'_m\Omega^{-1}\alpha_m\right)^{-1},$$

with $\gamma = \alpha_{m\perp}\left(\beta'_{m\perp}\alpha_{m\perp}\right)^{-1}$. We note that as $\mathrm{Cov}\left(\alpha'_m\Omega^{-1}\mathcal{B}_u, \gamma'\mathcal{B}_u\right) = \alpha'_m\Omega^{-1}\Omega\gamma = \alpha'_m\gamma = 0$, then $\gamma'\mathcal{B}_u$ and $\alpha'_m\Omega^{-1}\mathcal{B}_u$ are independent, which means $\hat{b}$ is asymptotically mixed Gaussian (MG) distributed. Moreover, note that $\hat{b}$ is super consistent due to the rate T of convergence.

A different way of stating this is, using the properties of stochastic integrals and MG,

$$T\left(\hat{b} - b\right) \xrightarrow{D} \mathrm{MG}\left(0, 1/\sigma_b^2\right), \quad \text{with } \sigma_b^2 = \left[\int_0^1 \gamma'\mathcal{B}_u\mathcal{B}'_u\gamma du\right]\left[\alpha'_m\Omega^{-1}\alpha_m\right]$$

Likewise, in terms of the t-ratio τ,

$$\tau = T\left(\hat{b} - b\right)\sqrt{\sigma_b^2} \xrightarrow{D} \mathrm{N}(0,1).$$

Finally, in terms of the vector $\hat{\beta}_m = (1, \hat{b})'$, we get, using that $m = c'\beta$ with $c = (1,0)'$, we can set $c_\perp = (0,1)'$ and hence,

$$T\left(\hat{\beta}_m - \beta_m\right) = \begin{pmatrix} 0 \\ T\left(\hat{b} - b\right) \end{pmatrix} = c_\perp T\left(\hat{b} - b\right) \xrightarrow{D} c_\perp\,\mathrm{MG}\left(0, 1/\sigma_b^2\right).$$

In other words, $c_\perp' \hat{\beta}_m = \hat{b}_m$ is asymptotically non-standard distributed with a limiting mixed Gaussian distribution, such that the t-ratio τ is standard Gaussian distributed, and, as can be shown, in general LR statistics for hypothesis testing on β are asymptotically χ^2 distributed.

IV.3.2 Asymptotics for MLE of β

The previous considerations can be extended to the general case of dimension p and rank r, with $\beta_m = \beta m^{-1}$:

Theorem IV.3.1 *Under the I(1) conditions H_0^r, then with $m = c'\beta$ of full rank r,*

$$T\left(\hat{\beta}_m - \beta_m\right) \xrightarrow{D} c_\perp \left[\int_0^1 \gamma' \mathcal{B}_u \mathcal{B}_u' \gamma du\right]^{-1} \int_0^1 \gamma' \mathcal{B}_u d\mathcal{B}_u' \Omega^{-1} \alpha_m \left(\alpha_m' \Omega^{-1} \alpha_m\right)^{-1},$$

where $\gamma = \alpha_{m\perp} \left(\beta_{m\perp}' \alpha_{m\perp}\right)^{-1}$.

Note that while the considerations in the previous section were made under the assumption of fixed α_m and Ω, Theorem IV.3.1 holds for α_m and Ω estimated. A proof, similar to the proof for the case of $p = 2$ and $r = 1$ is given in Section IV.3.4 below (see also Johansen, 1996, proof of Theorem 13.3).

Note also that Theorem IV.3.1 states that $\hat{\beta}_m$ is asymptotically mixed Gaussian,

$$T\left(\hat{\beta}_m - \beta_m\right) \xrightarrow{D} c_\perp \,\mathrm{MG}\left(0, \Sigma^{-1}\right)$$

where $\Sigma = \left[\int_0^1 \gamma' \mathcal{B}_u \mathcal{B}_u' \gamma du\right]^{-1} \otimes [\alpha_m' \Omega^{-1} \alpha_m]$. Also note that as $c' c_\perp \,\mathrm{MG}\left(0, \Sigma^{-1}\right) = 0$, the limiting distribution of $\hat{\beta}_m$ is singular, reflecting that $\beta_m = (1, b)' = c_\perp b$, and hence that it is the distribution of $\hat{b}$ in $\hat{\beta}_m$ which is non-singular.

Note Recall that when estimating β by solving the eigenvalue problem in (IV.15), by definition we have

$$\hat{\beta}' S_{zz} \hat{\beta} = I_r \quad \text{and} \quad \hat{\beta}' S_{zy} S_{yy}^{-1} S_{yz} \hat{\beta} = \hat{\Lambda}_r,$$

where $\hat{\Lambda}_r = \mathrm{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_r)$. These normalizations – which are more complicated than using m^{-1} for β_m – ensure identification as well. However, the limiting distribution of $\hat{\beta}$ normalized this way we do not consider as it is not needed for our considerations.

IV.3.3 Asymptotics for MLE of α

With β normalized by m, $\beta_m = \beta m^{-1}$, recall that $\alpha_m = \alpha m'$, and hence by (IV.13),

$$\hat{\alpha}_m - \alpha_m = S_{\varepsilon z}\hat{\beta}_m(\hat{\beta}_m' S_{zz}\hat{\beta}_m)^{-1}$$

Thus with β_m fixed, it follows that $\hat{\alpha}_m(\beta_m)$ satisfies,

$$\begin{aligned}\sqrt{T}(\hat{\alpha}_m(\beta_m) - \alpha_m) &= \sqrt{T} S_{\varepsilon z}\beta_m(\beta_m' S_{zz}\beta_m)^{-1} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t X_{t-1}'\beta_m(\frac{1}{T}\sum_{t=1}^{T}\beta_m' X_{t-1}X_{t-1}'\beta_m)^{-1}.\end{aligned}$$

And, using standard arguments for the CLT and LLN,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t X_{t-1}'\beta_m \to_d \mathrm{N}\left(0, \Omega\otimes\Sigma_{\beta\beta}\right), \qquad \frac{1}{T}\sum_{t=1}^{T}\beta_m' X_{t-1}X_{t-1}'\beta_m \to_p \Sigma_{\beta\beta},$$

where $\Sigma_{\beta\beta} = \mathbb{V}\left(\beta_m' X_t^*\right)$, we find

$$\sqrt{T}(\hat{\alpha}_m(\beta_m) - \alpha_m) \to_d \mathrm{N}\left(0, \Omega\otimes\Sigma_{\beta\beta}^{-1}\right).$$

For $\hat{\alpha}_m = \hat{\alpha}_m(\hat{\beta}_m)$, that is with $\hat{\beta}_m$ inserted, we find the equivalent result:

Theorem IV.3.2 *Under the I(1) conditions, then with $m = c'\beta$ of full rank r,*

$$\sqrt{T}(\hat{\alpha}_m - \alpha_m) \to_d \mathrm{N}\left(0, \Omega\otimes\Sigma_{\beta\beta}^{-1}\right),$$

with $\Sigma_{\beta\beta} = \mathbb{V}\left(\beta_m' X_t^*\right)$.

A proof can be found in Johansen (1996, proof of Theorem 13.3) which is based on arguments as for the case with β_m known ($\hat{\alpha}_m(\beta_m)$) treated above, using that $\hat{\beta}_m$ is super consistent, and hence, $\hat{\alpha}_m - \alpha_m = \hat{\alpha}_m(\hat{\beta}_m) - \alpha_m = \hat{\alpha}_m(\beta_m) - \alpha_m + o_p(1)$.

In the next section the distribution is derived for $\hat{\beta}_m$ for general p and r, with α_m and Ω fixed.

IV.3.4 General proof for $\hat{\beta}_m$

Consider the VAR model is given by

$$\Delta X_t = \alpha_m\beta_m' X_{t-1} + \varepsilon_t,$$

where $\beta_m = \beta m^{-1} = (I_r, b')$, such that $m = c'\beta$ and $c = (I_r, 0)'$.

Consider the MLE $\hat{b}$, or equivalently, $\hat{\beta}_m = (I_r, \hat{b}')'$ with Ω and α_m fixed. By definition, $\hat{b}$ satisfies (with d denoting differential),

$$d \log L\left(\alpha_m, \beta_m, \Omega; db\right)|_{\hat{\beta}_m} = 0, \tag{IV.44}$$

where db is $(p-r) \times r$ dimensional, and

$$\log L\left(\alpha_m, \beta_m, \Omega\right) = -\frac{T}{2} \log \det\left(\Omega\right) - \frac{1}{2} \sum_{t=1}^{T} (\Delta X_t - \alpha_m \beta_m' X_{t-1})' \Omega^{-1} (\Delta X_t - \alpha_m \beta_m' X_{t-1}).$$

Set, $c_\perp = (0, I_{p-r})'$, and note that as $\beta_m = c_\perp b$, then $d\beta_m = c_\perp db$, and therefore $d\left(\beta' X_{t-1}\right) = db' c_\perp' X_t$. We find that (IV.44) is given by

$$\begin{aligned} d \log L\left(\alpha_m, \beta_m, \Omega : db\right)|_{\hat{\beta}_m} &= \sum_{t=1}^{T} (\Delta X_t - \alpha_m \hat{\beta}_m' X_{t-1})' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1} \\ &= \sum_{t=1}^{T} (\varepsilon_t - \alpha_m (\hat{\beta}_m - \beta_m)' X_{t-1})' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1} \\ &= \sum_{t=1}^{T} (\varepsilon_t - \alpha_m (\hat{b} - b)' c_\perp' X_{t-1})' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1} = 0 \end{aligned}$$

Hence,

$$\sum_{t=1}^{T} \varepsilon_t' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1} = \sum_{t=1}^{T} X_{t-1}' c_\perp \left(\hat{b} - b\right) \alpha_m' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1} \tag{IV.45}$$

Using that $\operatorname{tr}(a) = a$ with a scalar, and $\operatorname{tr}(AB) = \operatorname{tr}(BA)$,

$$\operatorname{tr}\Big(\sum_{t=1}^{T} \varepsilon_t' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1}\Big) = \operatorname{tr}\Big(\sum_{t=1}^{T} c_\perp' X_{t-1} \varepsilon_t' \Omega^{-1} \alpha_m db'\Big)$$

$$\operatorname{tr}\Big(\sum_{t=1}^{T} X_{t-1}' c_\perp \left(\hat{b} - b\right) \alpha_m' \Omega^{-1} \alpha_m db' c_\perp' X_{t-1}\Big) = \operatorname{tr}\Big(\sum_{t=1}^{T} c_\perp' X_{t-1} X_{t-1}' c_\perp \left(\hat{b} - b\right) \alpha_m' \Omega^{-1} \alpha_m db'\Big)$$

and as (IV.45) holds for all db,

$$T^{-1} \sum_{t=1}^{T} c_\perp' X_{t-1} \varepsilon_t' \Omega^{-1} \alpha_m = T^{-1} \sum_{t=1}^{T} c_\perp' X_{t-1} X_{t-1}' c_\perp \left(\hat{b} - b\right) \alpha_m' \Omega^{-1} \alpha_m,$$

or

$$c_\perp' S_{z\varepsilon} \Omega^{-1} \alpha_m = c_\perp' S_{zz} c_\perp \left(\hat{b} - b\right) \alpha_m' \Omega^{-1} \alpha_m,$$

That is, analogous to the univariate case,

$$\hat{b} - b = (c_\perp' S_{zz} c_\perp)^{-1} c_\perp' S_{z\varepsilon} \Omega^{-1} \alpha_m \left(\alpha_m' \Omega^{-1} \alpha_m\right)^{-1}$$

Again by (IV.8),

$$c_\perp' X_t = c_\perp' \left[C \sum_{i=1}^{t} \varepsilon_i + C_S S_t + C_0 \right],$$

where $C_\Sigma = \beta_{m\perp} \left(\alpha_{m\perp}' \beta_{m\perp}\right)^{-1} \alpha_{m\perp}'$, with

$$\beta_{m\perp} = \begin{pmatrix} -b' \\ I_{p-r} \end{pmatrix}, \quad \text{where } \beta_m' \beta_{m\perp} = -b' + b' = 0.$$

In particular, $c_\perp' \beta_{m\perp} = I_{p-r} \neq 0$, and hence, as above,

$$\begin{aligned} \frac{1}{\sqrt{T}} c_\perp' X_{[Tu]} &= \left(\alpha_{m\perp}' \beta_{m\perp}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tu]} \alpha_{m\perp}' \varepsilon_i + o_p(1) \\ &\xrightarrow{D} \left(\alpha_{m\perp}' \beta_{m\perp}\right)^{-1} \alpha_{m\perp}' \mathcal{B}_u = \gamma' \mathcal{B}_u. \end{aligned}$$

Collecting terms,

$$T\left(\hat{b} - b\right) \xrightarrow{D} \mathrm{MG} := \left[\int_0^1 \gamma' \mathcal{B}_u \mathcal{B}_u' \gamma du \right]^{-1} \int_0^1 \gamma' \mathcal{B}_u d\mathcal{B}_u' \Omega^{-1} \alpha_m \left(\alpha_m' \Omega^{-1} \alpha_m\right)^{-1}$$

That is, as $\mathrm{Cov}\left(\alpha_m' \Omega^{-1} \mathcal{B}_u, \gamma' \mathcal{B}_u\right) = \alpha_m' \Omega^{-1} \Omega \gamma = \alpha_m' \gamma = 0$, $\hat{b}$ is super-consistent and asymptotically mixed Gaussian distributed. This can also be stated as in Theorem IV.3.1, using $\beta_m = c_\perp b$,

$$T\left(\hat{\beta}_m - \beta_m\right) \xrightarrow{D} c_\perp \mathrm{MG} = \left[\int_0^1 \gamma' \mathcal{B}_u \mathcal{B}_u' \gamma du \right]^{-1} \int_0^1 \gamma' \mathcal{B}_u d\mathcal{B}_u' \Omega^{-1} \alpha_m \left(\alpha_m' \Omega^{-1} \alpha_m\right)^{-1}$$

IV.4 Orthogonal complements

Having obtained the asymptotic distribution for the normalized $\hat{\alpha}_m$ and $\hat{\beta}_m$, we now consider their orthogonal complements.

Recall that $\alpha_\perp$ and $\beta_\perp$ are $(p \times (p - r))$-dimensional matrices of full rank $p - r$ and such that

$$\beta_\perp' \beta = 0 \text{ and } \alpha_\perp' \alpha = 0,$$

with $\det(\alpha, \alpha_\perp) \neq 0$ and $\det(\beta, \beta_\perp) \neq 0$.

A main challenge is that, as for α and β, the orthogonal complements $\alpha_\perp$ and $\beta_\perp$ are not identified in the sense that the parameters are not unique, and a normalization is needed; importantly, this holds even if α and β are identified.

IV.4.1 Considering $\beta_\perp$

To overcome the identification issue observe that with $\beta_m = \beta m^{-1}$ where $m = c'\beta$, and $c = (I_r, 0)'$ then the analog

$$\beta_{m\perp} = \beta_\perp \left(c_\perp' \beta_\perp\right)^{-1}, \quad c_\perp = (0, I_{p-r})',$$

is identified. Next, use the simple identity, or skew-projection,

$$\begin{aligned} I_p &= \beta_\perp \left(c_\perp' \beta_\perp\right)^{-1} c_\perp' + c\left(\beta' c\right)^{-1} \beta' \\ &= \beta_{m\perp} c_\perp' + c\beta_m' \end{aligned}$$

to solve for $\beta_{m\perp}$,

$$\beta_{m\perp} = (I - c\beta_m^{'})c_\perp.$$

In other words, we can find a unique $\beta_{m\perp}$ by using the identified β_m, and by this specific choice

$$\begin{aligned} T(\hat{\beta}_{m\perp} - \beta_{m\perp}) &= T((I - c\hat{\beta}_m^{'})\bar{c}_\perp - (I - c\beta_m^{'})\bar{c}_\perp) \\ &= -Tc(\hat{\beta}_m - \beta_m)'\bar{c}_\perp \end{aligned}$$

From above, we have $T\left(\hat{\beta}_m - \beta_m\right) \xrightarrow{D} c_\perp \,\mathrm{MG}$, and hence, as for $\hat{\beta}_m$, we find

$$T(\hat{\beta}_{m\perp} - \beta_{m\perp}) \xrightarrow{D} -c\,\mathrm{MG}' = -c\left(\alpha_m' \Omega^{-1} \alpha_m\right)^{-1} \alpha_m' \Omega^{-1} \int_0^1 d\mathcal{B}_u \mathcal{B}_u' \gamma \left[\int_0^1 \gamma' \mathcal{B}_u \mathcal{B}_u' \gamma du\right]^{-1}$$

Recall that $c = (I_r, 0)'$, and hence with $c_\perp = (0, I_{p-r})'$ we can state the above as "dual" results:

$$\begin{aligned} T(\hat{\beta}_{m\perp} - \beta_{m\perp}) &= -Tc(\hat{b} - b)' \xrightarrow{D} -c\,\mathrm{MG}' \\ T(\hat{\beta}_m - \beta) &= Tc_\perp(\hat{b} - b) \xrightarrow{D} c_\perp \,\mathrm{MG} \end{aligned}$$

IV.4.2 Asymptotics for $\hat{\alpha}_\perp$

As for $\beta_\perp$, we have $\alpha_{\perp n} = \alpha_\perp n^{-1}$ is identified, where n is a known $(p-r) \times (p-r)$ dimensional matrix of full rank. Which kind of normalization is of interest depends on the application. Thus if for example estimation of the C_Σ in (IV.8) is of interest, we note by definition,

$$C_\Sigma = \beta_\perp(\alpha_\perp' \beta_\perp)^{-1} \alpha_\perp = \beta_{m\perp}(\alpha_{n\perp}' \beta_{m\perp})^{-1} \alpha_{n\perp}.$$

That is, any normalization for $\beta_\perp$ and $\alpha_\perp$ respectively will work; i.e. C_Σ is invariant to the choices. A different situation is if $\alpha_\perp$ itself is of interest, in which case the concrete application implies which normalization may be of interest.

IV.4.2.1 Case of $p = 2, r = 1$

Consider initially a simple choice of $\alpha_\perp$ for the $p = 2, r = 1$ case previously initially considered.

Here $\beta_m = \beta m^{-1} = (1, b)'$, with $m = c'\beta$ and $c = (1, 0)'$. Hence β_m and $\alpha_m = (a_1, a_2)'$ are identified, and we can choose,

$$\alpha_{m\perp} = \begin{pmatrix} \alpha_{m\perp,1} \\ \alpha_{m\perp,2} \end{pmatrix} = \begin{pmatrix} -a_2 \\ a_1 \end{pmatrix} = A\alpha, \quad \text{where } A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

For this particular choice it follows immediately that with $\sigma^2_{\beta\beta} = \mathbb{V}(\beta'_m X^*_t)$,

$$\sqrt{T}(\hat{\alpha}_{m\perp} - \alpha_{m\perp}) = A\sqrt{T}(\hat{\alpha}_m - \alpha_m) \xrightarrow{D} A\,\mathrm{N}(0, \Omega/\sigma^2_{\beta\beta}) = \mathrm{N}(0, A\Omega A'/\sigma^2_{\beta\beta}).$$

Alternatively, normalizing as $\alpha_{n\perp} = \alpha_\perp n^{-1}$, with $n = c'_\perp \alpha_\perp = a_1$, assuming $a_1 \neq 0$, we get.

$$\alpha_{\perp n} = \begin{pmatrix} a_\perp \\ 1 \end{pmatrix} = \begin{pmatrix} -a_2/a_1 \\ 1 \end{pmatrix}.$$

Hence,

$$\hat{\alpha}_{n\perp} - \alpha_{n\perp} = \begin{pmatrix} \hat{a}_\perp - a_\perp \\ 0 \end{pmatrix} = c\left(\hat{a}_\perp - a_\perp\right),$$

where, by definition

$$\hat{a}_\perp - a_\perp = -\left(\frac{\hat{a}_2}{\hat{a}_1} - \frac{a_2}{a_1}\right)$$

which does not have a standard limiting Gaussian distribution it seems at a first sight. However, note that as

$$\frac{\hat{a}_2}{\hat{a}_1} - \frac{a_2}{a_1} = \frac{\hat{a}_2 a_1 - \hat{a}_1 a_2}{a_1 \hat{a}_1} = \frac{\left(\hat{a}_2 - a_2\right) a_1 - \left(\hat{a}_1 - a_1\right) a_2}{a_1 \hat{a}_1}$$

we get,

$$\begin{aligned} -\sqrt{T}\left(\frac{\hat{a}_2}{\hat{a}_1} - \frac{a_2}{a_1}\right) &= \frac{\sqrt{T}\left(\hat{a}_1 - a_1\right) a_2 - \sqrt{T}\left(\hat{a}_2 - a_2\right) a_1}{a_1 \hat{a}_1} = \frac{(a_2, -a_1)}{a_1 \hat{a}_1}\sqrt{T}\left(\hat{\alpha}_m - \alpha_m\right) \\ &\to_D -\frac{(-a_2/a_1, 1)}{a_1}\mathrm{N}\left(0, \Omega/\sigma^2_{\beta\beta}\right) = -\frac{1}{a_1}\alpha'_{n\perp}\mathrm{N}\left(0, \Omega/\sigma^2_{\beta\beta}\right). \end{aligned}$$

That is, a one dimensional Gaussian limiting distribution.

Yet another normalization, as sometimes applied in the so-called price discovery litterature, is given by

$$\alpha_{p\perp} = \begin{pmatrix} \alpha_{m\perp,1} \\ \alpha_{m\perp,2} \end{pmatrix} = \alpha_{m\perp}\left(p'\alpha_{m\perp}\right)^{-1} = \begin{pmatrix} \alpha_{m\perp,1} \\ \alpha_{m\perp,2} \end{pmatrix} / \left(\alpha_{m\perp,1} + \alpha_{m\perp,2}\right), \;\; p = (1, 1)'.$$

This way, $p'\alpha_{p\perp} = \alpha_{p\perp,1} + \alpha_{p\perp,2} = 1$. As above we get,

$$\sqrt{T}(\hat{\alpha}_{p\perp} - \alpha_{p\perp}) = \sqrt{T}\left(\hat{\alpha}_{m\perp}\left(p'\hat{\alpha}_{m\perp}\right)^{-1} - \alpha_{m\perp}\left(p'\alpha_{m\perp}\right)^{-1}\right)$$

to be asymptotically Gaussian as well.

Next we consider different normalizations of $\alpha_\perp$ by some $(p-r)\times(p-r)$ dimensional full rank n to obtain identication, $\alpha_{\perp n} = \alpha_\perp n^{-1}$.

IV.4.3 A normalization of theoretical interest

A choice of $\hat{\alpha}_\perp$, which may be of theoretical interest, would be

$$\hat{\alpha}_{m\perp} = (I_p - \alpha_m(\hat{\alpha}_m'\alpha_m)^{-1}\hat{\alpha}_m')\alpha_{m\perp} = \alpha_\perp - \alpha_m(\hat{\alpha}_m'\alpha_m)^{-1}\hat{\alpha}_m'\alpha_{m\perp},$$

where $\alpha_{m\perp}'\alpha_m = 0$ and $\det(\alpha_m, \alpha_{m\perp}) \neq 0$.

We note that by definition, $\hat{\alpha}_m'\hat{\alpha}_{m\perp} = 0$, and hence

$$\hat{\alpha}_{m\perp} - \alpha_\perp = -\alpha_m(\hat{\alpha}_m'\alpha_m)^{-1}(\hat{\alpha}_m - \alpha_m)'\alpha_{m\perp}$$

A Taylor expansion of $f(a) = (a'\alpha_m)^{-1}$ gives, with $\bar{\alpha}_m = \alpha_m\left(\alpha_m'\alpha_m\right)^{-1}$,

$$\sqrt{T}(\hat{\alpha}_{m\perp} - \alpha_\perp) = -\bar{\alpha}_m\sqrt{T}(\hat{\alpha}_m - \alpha_m)'\alpha_{m\perp} + o_p(1),$$

where the last term (by a Taylor expansion) is $o_p(1)$ as $\bar{\alpha}_m(\hat{\alpha}_m - \alpha_m)'\bar{\alpha}_m(\hat{\alpha}_m - \alpha_m)'\alpha_{m\perp}$ is $O_p(T^{-1})$.

We conclude,

$$\sqrt{T}(\hat{\alpha}_{m\perp} - \alpha_\perp) \to_D -\bar{\alpha}_m \,\mathrm{N}(0, \Sigma_{\beta\beta}\otimes\Omega)\,\alpha_{m\perp} = \mathrm{N}(0, \bar{\alpha}_m\Sigma_{\beta\beta}\bar{\alpha}_m' \otimes \alpha_{m\perp}'\Omega\alpha_{m\perp}).$$

IV.4.4 Normalization with $c_\perp$

With $\alpha_{n\perp} = \alpha_\perp n^{-1}$, consider here $n = c_\perp'\alpha_\perp$, with $c_\perp = (0, I_{p-r})'$ as used for β and $\beta_\perp$, assuming n of full rank. In line with skew-projections repeatedly used, we have

$$\begin{aligned} I_p &= \alpha_\perp(c_\perp'\alpha_\perp)^{-1}c_\perp' + c\left(\alpha'c\right)^{-1}\alpha' \\ &= \alpha_{\perp n}c_\perp' + c\left(\alpha_m'c\right)^{-1}\alpha_m' \end{aligned}$$

using $\alpha_m = \alpha\beta'c$. It thus follows that, a candidate with is

$$\alpha_{\perp n} = (I_p - c\left(\alpha_m'c\right)^{-1}\alpha_m')c_\perp$$

Hence with $\hat{\alpha}_{\perp n} = (I_p - c\left(\hat{\alpha}_m'c\right)^{-1}\hat{\alpha}_m')c_\perp$, we find

$$\sqrt{T}(\hat{\alpha}_{\perp n} - \alpha_{\perp n}) = -c\sqrt{T}[\left(\hat{\alpha}_m'c\right)^{-1}\hat{\alpha}_m' - (\alpha_m'c)^{-1}\alpha_m']c_\perp$$

Rewriting, we get

$$\sqrt{T}(\hat{\alpha}_{\perp n} - \alpha_{\perp n}) = -c(\alpha'_m c)^{-1}\sqrt{T}[\hat{\alpha}_m - \alpha_m]' c_\perp + V_T$$

where

$$\begin{aligned} V_T &= -c\sqrt{T}[(\hat{\alpha}'_m c)^{-1} - (\alpha'_m c)^{-1}]\hat{\alpha}'_m c_\perp \\ &= c(\alpha'_m c)^{-1}\sqrt{T}(\hat{\alpha}_m - \alpha_m)' c(\alpha'_m c)^{-1}\alpha'_m c_\perp + o_p(1) \end{aligned}$$

with the $o_p(1)$ term as above. Collecting terms,

$$\sqrt{T}(\hat{\alpha}_{\perp n} - \alpha_{\perp n}) = -c(\alpha'_m c)^{-1}\sqrt{T}[\hat{\alpha}_m - \alpha_m]'(I - c(\alpha'_m c)^{-1}\alpha'_m)c_\perp$$

and, asymptotic normailty holds by that of $\hat{\alpha}_m$,

$$\sqrt{T}(\hat{\alpha}_{\perp n} - \alpha_{\perp n}) \to_D -c(\alpha'_m c)^{-1}\,\mathrm{N}\left(0, \Omega \otimes \Sigma_{\beta\beta}^{-1}\right)(I - c(\alpha'_m c)^{-1}\alpha'_m)c_\perp$$

Remark IV.4.1 *Another choice matching the expression for* $C = \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha_\perp = \beta_{m\perp}\left(\alpha'_\perp \beta_{m\perp}\right)^{-1}\alpha'_\perp$*, would be to use* $n = \beta'_{m\perp}\alpha_\perp$*, and hence*

$$\alpha_{n\perp} = \alpha_\perp(\beta'_{m\perp}\alpha_\perp)^{-1}.$$

A detailed discussion of this choice, and more general normalizations can be found in Paruolo (1997).

IV.5 Deterministic terms and VAR(k)

IV.5.1 Constant level

Similar to the univariate case, consider initially the VAR(1) model with a constant regressor,

$$\Delta X_t = \Pi X_{t-1} + \mu + \varepsilon_t, \quad t = 1, 2, ..., T \tag{IV.46}$$

$\Pi \in \mathbb{R}^{p\times p}$, $\mu \in \mathbb{R}^p$, X_0 fixed and ε_t i.i.d.$\mathrm{N}_p(0, \Omega)$. Under the hypothesis $H_r : \Pi = \alpha\beta'$, and Assumption IV.2.1, X_t has the representation,

$$X_t = C\sum_{i=1}^{t}(\varepsilon_i + \mu) + C_S S_{t,c} + C_0, \tag{IV.47}$$

where $S_{t,c} = \beta' X_t$ is the asymptotically stable process given by $S_{t,c} = (I_r + \beta'\alpha)S_{t-1,c} + \beta'\mu$. In other words, the reduced rank assumption on Π implies that X_t has a linear trend $C\mu t$.

The linear trend vanishes provided $C\mu = 0$, or $\alpha'_{\perp}\mu = 0$. This leads to the hypothesis of interest in the case of an included $c$onstant regressor to be given by,

$$H_{r,c} : \Pi = \alpha\beta',\ \mu = \alpha\mu_c^{'}, \tag{IV.48}$$

where μ_c' is an r dimensional vector. Denote by $H_{r,c}^0$, $H_{r,c}^0 \subseteq H_{r,c}$, the hypothesis that $H_{r,c}$ and Assumption IV.2.1 holds. Then under $H_{r,c}^0$, by Theorem IV.2.1,

$$H_{r,c}^0 : X_t = C_\Sigma \sum_{i=1}^{t} \varepsilon_i + C_S S_{t,c} + C_0. \tag{IV.49}$$

In particular, the mean of the stationary version of the cointegrating relations $\beta' S_t$ equals,

$$\mathbb{E}\left[S_{t,c}^*\right] = \sum_{i=0}^{\infty}(I + \beta'\alpha)^i \beta'\mu = -(\beta'\alpha)^{-1}\beta'\mu = -\mu_c^{'}, \tag{IV.50}$$

and X_t has the representation as an I(1) process with a constant level given by $C_0 - C_S\mu_c'$.

Under $H_{r,c}$ the VAR(1) model is given by,

$$\begin{aligned}
\Delta X_t &= \alpha\beta' X_{t-1} + \alpha\mu_c^{'} + \varepsilon_t & \text{(IV.51)}\\
&= \alpha \begin{pmatrix} \beta \\ \mu_c \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix}' + \varepsilon_t & \text{(IV.52)}\\
&= \alpha\beta_c^{'} X_{t-1,c} + \varepsilon_t, & \text{(IV.53)}
\end{aligned}$$

and hence the MLE of α, $\beta_c = (\beta', \mu_c')$ and Ω are found by RRR of ΔX_t on $X_{t-1,c}$. The unrestricted model $H_{p,c}$ is given by (IV.46), and limiting distribution of the LR test of $H_{r,c}$ against $H_{p,c}$ converge in distribution under $H_{r,c}^0$,

$$\mathrm{LR}_{r,c} \equiv \mathrm{LR}(H_{r,c}|H_{p,c}) \xrightarrow{D} DF_{p-r}^c(\mathcal{W}), \tag{IV.54}$$

where

$$DF_{p-r}^c(\mathcal{W}) = \mathrm{tr}\{\int_0^1 d\mathcal{W}\mathcal{W}^{c\prime}(\int_0^1 \mathcal{W}_u^c \mathcal{W}_u^{c\prime} du)^{-1} \int_0^1 \mathcal{W}^c d\mathcal{W}'\}, \tag{IV.55}$$

with $\mathcal{W}$ a $p-r$ dimensional standard Brownian motion, and $\mathcal{W}^c = (\mathcal{W}', 1)'$. Some quantiles of DF^c are reported in Section IV.5.4.

IV.5.2 Linear trend

Similar considerations for a model which allows for a linear trend $\tau \in \mathbb{R}^p$,

$$\Delta X_t = \Pi X_{t-1} + \tau t + \mu + \varepsilon_t, \tag{IV.56}$$

leads to consider the model $H_{r,l}$ as given by,

$$\Delta X_t = \alpha\beta' X_{t-1} + \alpha\tau_l' t + \mu + \varepsilon_t, \tag{IV.57}$$

where $\tau_l' \in \mathbb{R}^r$. Denote by $H_{r,l}^0$, $H_{r,l}^0 \subseteq H_{r,l}$, the case where $H_{r,l}$ and Assumption IV.2.1 hold. Then under $H_{r,l}^0$, X_t has the representation,

$$X_t = C_\Sigma \sum_{i=1}^{t} \varepsilon_t + C_\Sigma t + C_0 + C_S S_{t,l}, \tag{IV.58}$$

where the cointegrating relations $S_{t,l} = \beta' X_t$ are asymptotically stable around a linear trend,

$$S_{t,l} = (I + \beta'\alpha) S_{t-1,l} + \beta'\tau t + \beta'\mu + \varepsilon_t. \tag{IV.59}$$

That is, X_t is trending-I(1) and $\beta' X_t$ also have a linear trend.

Under $H_{r,l}$ the VAR(1) model is given by,

$$\begin{aligned}
\Delta X_t &= \alpha\beta' X_{t-1} + \alpha\tau_l' t + \mu + \varepsilon_t && \text{(IV.60)}\\
&= \alpha \begin{pmatrix} \beta \\ \tau_l \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ t \end{pmatrix}' + \mu + \varepsilon_t && \text{(IV.61)}\\
&= \alpha\beta_l' X_{t-1,l} + \mu + \varepsilon_t, && \text{(IV.62)}
\end{aligned}$$

and hence the MLE of α, $\beta_l = (\beta', \tau_l')$, μ and Ω are found by RRR of ΔX_t on $X_{t-1,l}$, both corrected by OLS regression on the constant. The unrestricted model $H_{p,l}$ is given by (IV.56), and limiting distribution of the LR test of $H_{r,l}$ against $H_{p,l}$ converge in distribution under $H_{r,l}^0$,

$$\mathrm{LR}_{r,l} \equiv \mathrm{LR}(H_{r,l}|H_{p,l}) \xrightarrow{D} DF_{p-r}^l(\mathcal{W}), \tag{IV.63}$$

where

$$DF_{p-r}^l(\mathcal{W}) = \mathrm{tr}\{\int_0^1 d\mathcal{W}\mathcal{W}^{l\prime}(\int_0^1 \mathcal{W}_u^l \mathcal{W}_u^{l\prime} du)^{-1} \int_0^1 \mathcal{W}^l d\mathcal{W}'\}, \tag{IV.64}$$

with $\mathcal{W}$ a $p-r$ dimensional standard Brownian motion, and $\mathcal{W}_u^l = (\mathcal{W}_u - \int_0^1 \mathcal{W}_s ds, u - 1/2)'$. Some quantiles of DF^l are reported in Section IV.5.4.

IV.5.3 The VAR(k) model

Consider the p-dimensional VAR(k) model given by,

$$X_t = A_1 X_{t-1} + ... + A_k X_{t-k} + \varepsilon_t, \quad t = 1, ..., T \tag{IV.65}$$

where $A_i \in \mathbb{R}^{p \times p}$, $X_0, ...X_{-k+1}$ are fixed and ε_t are i.i.d.$\mathrm{N}_p(0, \Omega)$, $\Omega > 0$. As before, to allow for roots at $z = 1$ in the characteristic polynomial, reparametrize the model as,

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t, \tag{IV.66}$$

where $\Pi, \Gamma_i \in \mathbb{R}^{p \times p}$ with $\Pi = \sum_{i=1}^k A_i - I_p \in$ and $\Gamma_i = -\sum_{j=i+1}^k A_j \in \mathbb{R}$ for $i = 1, ..., k-1$. The characteristic polynomial is given by

$$A(z) = (1 - z) I_p - \Pi z - \Gamma_1 (1 - z) z - ... - \Gamma_{k-1} (1 - z) z^k, \tag{IV.67}$$

and it immediately follows that $\det(A(1)) = 0$ if, and only if, Π has reduced rank.

IV.5.3.1 Representation of VAR(k) processes

The generalization of Theorem IV.2.2 is given by:

Theorem IV.5.1 *Consider the VAR(k) process given by (IV.66) under the hypothesis*

$$H_r : \Pi = \alpha\beta', \quad \alpha, \beta \in \mathbb{R}^{p \times r}, \; r < p. \tag{IV.68}$$

Then if Assumption IV.2.1 holds, X_t is an I(1) process. Moreover, it has the representation,

$$X_t = C_\Sigma \sum_{i=1}^t \varepsilon_i + C_S S_t + C_0, \tag{IV.69}$$

where $C_\Sigma = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$ is a $p \times p$ dimensional matrix of rank $(p - r)$ and $\Gamma = (I - \sum_{j=1}^{k-1} \Gamma_j)$. The $(r + p(k-1))$ dimensional process $S_t = (X'_t\beta, \Delta X'_{t-1}, ..., \Delta X'_{t-k+1})'$ is asymptotically stable. In particular, S_0 can be given an initial distribution such that S_t and the cointegrating relations $\beta' X_t$ have a stationary representation. Moreover, C_S is a $p \times (r + p(k-1))$ matrix, and C_0 depends on the initial values $X_0, \Delta X_0, ..., \Delta X_{-k+2}$, and satisfies $\beta' C_0 = 0$.

Remark IV.5.1 *With*

$$A = \begin{pmatrix} A_1 & \cdots & A_{k-1} & A_k \\ I_p & & & \\ & \ddots & & \\ & & I_p & 0 \end{pmatrix}$$

the condition is equivalent to A having $(p-r)$ eigenvalues equal to one, and the remaining smaller than one in absolute value.

Proof of Theorem IV.5.1: The proof follows by mimicking the proof of Theorem IV.2.1 by writing the VAR(k) process as a $pk-$dimensional VAR(1) process. Specifically, for $k=2$, define $X_t^* = (X_t', X_{t-1}')'$, then $\Pi^* = A - I$, with A given by (). Hence,

$$\Delta X_t^* = \Pi^* X_{t-1}^* + \varepsilon_t, \quad \Pi^* = \alpha^* (\beta^*)', \text{ and} \tag{IV.70}$$

$$\alpha^* = \begin{pmatrix} \alpha & \Gamma_1 \\ 0 & I_p \end{pmatrix}, \beta^* = \begin{pmatrix} \beta & I_p \\ 0 & -I_p \end{pmatrix}, \tag{IV.71}$$

with $\varepsilon_t^* = (\varepsilon_t', 0)'$. By the proof of Theorem IV.2.1, X_t^* has the representation,

$$X_t^* = C^* \sum_{i=1}^{t} \varepsilon_t^* + C_S^* S_t^* + C_0^*, \tag{IV.72}$$

with $C^* = \beta_\perp^* (\alpha_\perp^{*\prime} \beta_\perp^*)^{-1} \alpha_\perp^{*\prime}$, $C_S^* = \alpha^* (\beta^{*\prime} \alpha^*)^{-1}$ and $C_0^* = C^* X_0^*$. Note that $\alpha_\perp^*$ and $\beta_\perp^*$ are given by,

$$\alpha_\perp^* = \begin{pmatrix} \alpha_\perp \\ -\Gamma_1' \alpha_\perp \end{pmatrix}, \beta_\perp^* = \begin{pmatrix} \beta_\perp \\ \beta_\perp \end{pmatrix} \tag{IV.73}$$

Finally, use that $X_t = (I, 0) X_t^*$ to derive the result for X_t as desired. $\square$

IV.5.3.2 Estimation in the VAR(k) model

Rewrite the VAR(k) model in (IV.66) under H_r as,

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma^* \Delta X_{t-1}^* + \varepsilon_t, \tag{IV.74}$$

with $\Gamma^* = (\Gamma_1, ..., \Gamma_k)$, and similarly, $\Delta X_{t-1}^* = (\Delta X_{t-1}', ..., \Delta X_{t-k+1}')'$. With α, β known, the MLE $\hat{\Gamma}^*(\alpha, \beta)$ is found by OLS regression of $\Delta X_t - \alpha \beta' X_{t-1}$ on ΔX_{t-1}^*. Introduce therefore the corresponding residuals,

$$Y_t = R(\Delta X_t | \Delta X_{t-1}^*) \text{ and } Z_t = R(X_{t-1} | \Delta X_{t-1}^*), \tag{IV.75}$$

from the OLS regressions. Then $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\Omega}$ are found by RRR of Y_t on Z_t, see Theorem IV.2.2.

IV.5.4 Examples and quantiles for rank testing

Summarizing the discussion above, three different models were of interest which with $\Delta X_{t-1}^{*} = (\Delta X_{t-1}', .., \Delta X_{t-k+1}')'$ and $\Gamma^{*} = (\Gamma_1, .., \Gamma_{k-1})$ can be rewritten as,

$$\begin{aligned} H_p: \ & \Delta X_t = \Pi X_{t-1} + \Gamma^{*} \Delta X_{t-1}^{*} + \varepsilon_t && \text{(IV.76)} \\ H_{p,c}: \ & \Delta X_t = \Pi X_{t-1} + \mu + \Gamma^{*} \Delta X_{t-1}^{*} + \varepsilon_t && \text{(IV.77)} \\ H_{p,l}: \ & \Delta X_t = \Pi X_{t-1} + \tau t + \mu + \Gamma^{*} \Delta X_{t-1}^{*} + \varepsilon_t && \text{(IV.78)} \end{aligned}$$

The hypotheses of interest are for each model given by,

$$\begin{aligned} H_r &: \Pi = \alpha\beta' && \text{(IV.79)} \\ H_{r,c} &: \Pi = \alpha\beta', \ \mu = \alpha\mu_c' && \text{(IV.80)} \\ H_{r,l} &: \Pi = \alpha\beta', \ \tau = \alpha\tau_l' && \text{(IV.81)} \end{aligned}$$

The limit distributions of the likelihood ratio tests of H_r against H_p, $H_{r,c}$ against $H_{p,c}$ and $H_{r,l}$ against $H_{p,l}$, under the assumption of Theorem IV.5.1 are given by,

$$\mathrm{DF}_{p-r}(\mathcal{W}), \ \mathrm{DF}_{p-r}^{c}(\mathcal{W}) \ \text{ and } \mathrm{DF}_{p-r}^{l}(\mathcal{W}) \qquad \text{(IV.82)}$$

respectively, where $\mathcal{W}$ is a $(p-r)$ dimensional standard Brownian motion. The quantiles of (IV.82) given below are from Johansen (1996).

IV.5.4.1 The spot and futures data

Returning to the MIB30 data, consider the analysis of $X_t = (s_t, f_t)'$ by a VAR(6) model with a linear trend:

$$\Delta X_t = \Pi X_{t-1} + \tau t + \mu + \Gamma_1 \Delta X_{t-1} + ... + \Gamma_5 \Delta X_{t-5} + \varepsilon_t. \qquad \text{(IV.83)}$$

In order to find the number of possible cointegrating relations, the following rank statistics were computed:

$$\mathrm{LR}_{0,l} = 21.5, \quad \mathrm{LR}_{1,l} = 5.1. \qquad \text{(IV.84)}$$

As $\mathrm{LR}_{0,l} = 21.5 > ...$, but $\mathrm{LR}_{1,l} = 5.1 < 9.1$, the hypothesis of one cointegrating vector is accepted.

The cointegrating vector, and its linear trend coefficient, are given by

$$(\hat{\beta}', \hat{\tau}_l) = \left(1, -0.99, 6 \cdot 10^{-6}\right). \qquad \text{(IV.85)}$$

Consider the LR test for the linear hypothesis given by,

$$H_{\mathrm{lin}} : (\hat{\beta}', \hat{\tau}_l) = \varphi(1, -1, 0), \qquad \text{(IV.86)}$$

that is, the linear trend can be omitted in the cointegrating relation, and the spread $s_t - f_t$ is indeed asymptotically stable, or stationary. Computation gives, $\mathrm{LR}_{\mathrm{lin}} = 1.9$, and as $1.9 < 6$, the 95% quantile of the χ^2_2 distribution, the hypothesis is clearly accepted.

Asymptotic Quantiles of LR Test for Rank (IV.87)

$\mathrm{LR}_r \equiv \mathrm{LR}(H_r \vert H_p)$			
$p-r$	95% quantile	97.5 % quantile	99 % quantile
1	4.2	5.3	7.0
2	12.2	13.9	16.1
3	24.0	26.4	29.1
4	39.7	42.5	46.0
5	59.2	62.6	66.7

(IV.88)

$\mathrm{LR}_{r,c} \equiv \mathrm{LR}(H_{r,c} \vert H_{p,c})$			
$p-r$	95% quantile	97.5 % quantile	99 % quantile
1	9.1	10.7	12.7
2	20.0	22.0	24.7
3	34.8	37.5	40.8
4	53.4	56.5	60.4
5	75.7	79.6	83.9

(IV.89)

$\mathrm{LR}_{r,l} \equiv \mathrm{LR}(H_{r,l} \vert H_{p,l})$			
$p-r$	95% quantile	97.5 % quantile	99 % quantile
1	12.3	14.1	16.3
2	25.4	27.8	30.6
3	42.2	45.0	48.5
4	62.6	66.0	70.2
5	86.9	90.8	95.3

(IV.90)

IV.5.4.2 Money demand

Return to the first page of Part I, where it was emphasized that it is of interest to see whether relations such as the money demand type relations exist.

With money stock M, prices P, real income Y and interest rate i, it was postulated that often one finds 'stable' relations between observed variables of the form such as,

$$m_t - p_t - b_1 y_t - b_2 i_t,$$

where the coefficients b_i are estimated from data, and $m_t = \log M_t$, $p_t = \log P_t$

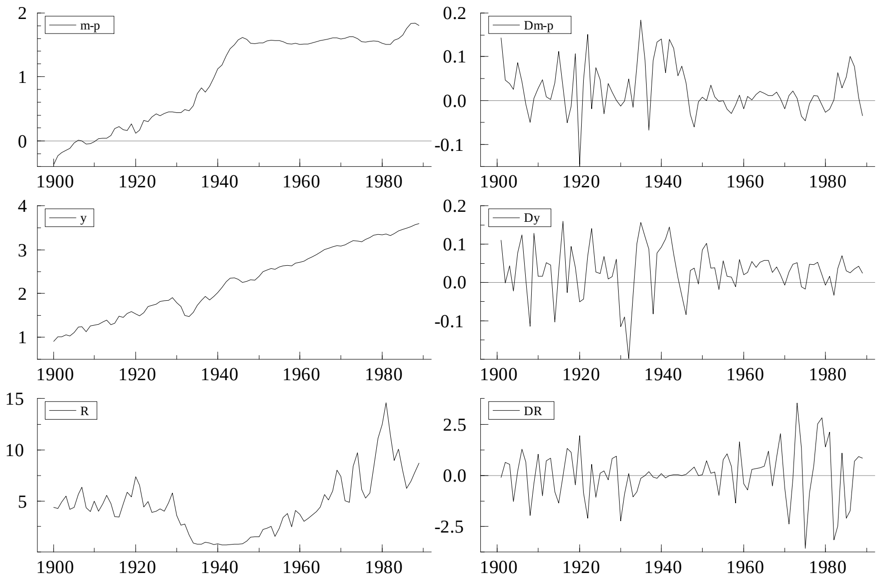

Figure 3: Data from Hayashi (2000). In the graphs $m - p = \log(M_1/P)$, R is the annual rate, $y = \log Y$, and Y is the net national product.

and $y_t = \log Y_t$. Consider annual US data for the period 1900-1989 shown in Figure 3.

By estimation of a VAR(2) model which allows for a linear trend for $X_t = ((m-p)_t, y_t, R_t)$ the following rank test statistics were found,

$$\mathrm{LR}_{0,l} = 48.5, \quad \mathrm{LR}_{1,l} = 18.3, \quad \mathrm{LR}_{2,l} = 3.2. \tag{IV.91}$$

Hence $\hat{r} = 1$, with

$$\hat{\beta}' X_t + \hat{\tau}_l t = (m-p)_t - 1.61 y_t + 0.11 R_t + 0.02t. \tag{IV.92}$$

The LR test statistic of the exclusion of the linear trend, $\tau_l = 0$, and $(m - p - y)$ cointegrating is given by

$$\mathrm{LR}\,(\tau_l = 0) = 3.3, \tag{IV.93}$$

and hence as the LR test is asymptotically χ^2_2 distributed this is accepted. The cointegrating relation is given by

$$\hat{\beta}' X_t = (m-p)_t - y_t + 0.11 R_t. \tag{IV.94}$$

References

[1] Hayashi, F.(2000), Econometrics, Princeton University Press.

[2] Johansen, S. (1996), Likelihood-based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press.

[3] Johansen, S. (2005), Cointegration: A Survey, Handbook of Econometrics.

[4] Paruolo, P. (1997), Asymptotic Inference on the Moving Average Impact Matrix in Cointegrated I(1) VAR Systems, *Econometric Theory,* 13:79-118.

A Results from Linear Algebra

Some well-known results that are used in the text are briefly mentioned here.

Let M be a $m \times n$ matrix of rank r. Then with $rank(M) = r(M) = r$, it holds that:

1. $r(M) = r \leq \min(m, n)$
2. $r(M) = r(MM') = r(M'M) = r(M') = r$
3. $r(MB) \leq \min(r(M), r(B))$
4. $r(MB) = r(M) = r$ if B is a $n \times n$ matrix of full rank

For the column space of M,

$$sp(M) = \{Mx \mid x \in \mathbb{R}^n\},$$

it holds that,

$$\begin{aligned} \dim(sp(M)) &= r(M) = r \\ sp(M) &= sp(MM') \end{aligned}$$

Also the direct sum applies, that is:

$$\begin{aligned} m &= \dim(sp(M)) + \dim(sp(M)_\perp) \\ &= r + (m - r) \end{aligned}$$

This is used explicitly in cointegration in the following way. Let α be a $m \times r$ matrix with rank r so that $sp(\alpha) = sp(M)$. Then define $\alpha_\perp$ as a $m \times (m - r)$ matrix of full rank $(m - r)$, such that $sp(\alpha_\perp) = sp(M)_\perp$ ie.

$$\alpha' \alpha_\perp = 0 \quad \text{and} \quad sp(\alpha, \alpha_\perp) = \mathbb{R}^m$$

In particular, the orthogonal projection is often applied:

$$I_m = \alpha(\alpha'\alpha)^{-1}\alpha' + \alpha_\perp(\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'.$$

A.1 Diagonalization

With M be a symmetric $p \times p$ matrix, then M is diagonalizable. The eigenvalue problem is given by,

$$\det(\lambda I_p - M) = 0, \tag{IV.95}$$

which is solved for eigenvalues $\lambda_1 \geq \lambda_2 \geq .. \geq \lambda_p$ with corresponding eigenvectors $v_1, .., v_p$. Then with $V = (v_1, .., v_p)$ and Λ =diag$(\lambda_1, .., \lambda_p)$,

$$V'MV = \Lambda \tag{IV.96}$$
$$MV = V\Lambda \tag{IV.97}$$
$$VV' = V'V = I_p. \tag{IV.98}$$

A generalized version of the eigenvalue problem, used in cointegration analysis and reduced rank regression (RRR), is given by:

$$\det(\rho N - M) = 0, \tag{IV.99}$$

where M is as before, while N is a positive definite (and hence in particular symmetric) $p \times p$ matrix. The generalized eigenvalue problem has eigenvalues $\rho_1 \geq ... \geq \rho_p$, with $W = (w_1, ..., w_p)$ the corresponding eigenvectors, and it holds that

$$W'MW = R \tag{IV.100}$$
$$MW = NWR \tag{IV.101}$$
$$W'NW' = I_p, \tag{IV.102}$$

where R =diag$(\rho_1, ..., \rho_p)$.

That the generalized eigenvalue problem has this solution, follows easily by noting that $N^{-1/2}MN^{-1/2}$ is symmetric and hence diagonalizable, and,

$$\det(\rho N - M) = 0 \Leftrightarrow \det\left(\rho I - N^{-1/2}MN^{-1/2}\right) = 0.$$

B General Decomposition

Let Π be a $p \times p$ matrix with rank r possibly less than p.

Then a 'Singular Value Decomposition' holds:

Lemma B.1 *With Π a $p \times p$ matrix of rank r, there exist $p \times r$ matrices A, B with full rank r and a $r \times r$ diagonal matrix $\Lambda = diag(\lambda_1, .., \lambda_r)$, where $\lambda_i > 0$ such that*

$$\Pi = A\Lambda^{1/2}B' \tag{IV.103}$$

It holds that $sp(B) = sp(\Pi')$ (row space) and $sp(A) = sp(\Pi)$ (column space).

The decomposition in cointegration of Π as $a\beta'$ then holds as a corollary:

Corollary B.1 *With* Π *a* $p \times p$ *matrix of rank* r, $p \times r$ *matrices* α, β *of rank* r *exist such that*

$$\Pi = \alpha\beta' \tag{IV.104}$$

where $sp(\beta) = sp(\Pi')$*.*

Proof of Lemma B.1:

$\Pi\Pi'$ is symmetric and positive definite with rank r. Diagonalize $\Pi\Pi'$, which has eigenvalues and eigenvectors,

$$\begin{aligned} \lambda_1 \geq .. \geq \lambda_r \ , \ V_1 = (v_1, .., v_r) \\ \lambda_{r+1} = ... = \lambda_p = 0, \ V_2 = (v_{r+1}, .., v_p) \end{aligned}$$

Then with $\Lambda = diag(\lambda_1, .., \lambda_r)$,

$$\Pi\Pi' V_1 = V_1 \Lambda \tag{IV.105}$$

$$\Pi\Pi' V_2 = 0 \tag{IV.106}$$

Note that (IV.106) implies $V_2 V_2' \Pi = 0$ (use e.g.. $r(V_2'\Pi) = r(V_2'\Pi\Pi' V_2) = 0$). Define $B = \Pi' V_1 \Lambda^{-1/2}$ and set $A = V_1$. Then the desired decomposition holds by

$$\begin{aligned} \Pi &= (V_1, V_2)(V_1, V_2)'\Pi = V_1 V_1' \Pi \\ &= V_1 \Lambda^{1/2} \Lambda^{-1/2} V_1' \Pi = A \Lambda^{1/2} B' \end{aligned}$$

$\square$