Notes on time series models

Gianluca Cubadda

Università di Roma "Tor Vergata"

AY 2023-24

1 The Beveridge-Nelson decomposition

By expanding the polynomial $\psi(L)$ on 1 we get

$$\psi(L) = \psi(1) + (1 - L)\psi^*(L) \tag{1}$$

where $\psi^*(L) = \sum_{j=0}^{\infty} \psi_j^* L^J$ is such that $\psi_j^* = -\sum_{i>j}^{\infty} \psi_i$.

In order to prove the above result, equate the coefficients corresponding to the same power of L on both sides of Equation (1) to get

$$L^0 \Rightarrow \psi_0 = 1 = \psi(1) + \psi_0^* \Rightarrow \psi_0^* = 1 - \psi(1) = -\sum_{i=1}^{\infty} \psi_i$$

$$L^1 \Rightarrow \psi_1 = \psi_1^* - \psi_0^* \Rightarrow \psi_1^* = \psi_1 + \psi_0^* = -\sum_{i=2}^{\infty} \psi_i$$

...

2 ADF tests

Consider the unobserved components (UC) representation:

$$\begin{aligned} y_t &= \alpha + \beta t + u_t, \\ \phi(L) u_t &= \varepsilon_t. \end{aligned} \tag{2}$$

where $\phi(L) = 1 - \phi L$ and $\phi \in (-1, 1]$. If we premultiply both sides of Equation (2) by $\phi(L)$ we get

$$y_t - \phi y_{t-1} = (1 - \phi)\alpha + \beta t - \phi\beta(t - 1) + \varepsilon_t$$

Subtracting y_{t-1} from both sides of the above equation and rearranging the terms we get

$$\Delta y_t = [(1 - \phi)\alpha - \phi\beta] + \beta(1 - \phi)t + (\phi - 1)y_{t-1} + \varepsilon_t$$

Assume that $\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$. By expanding the polynomial $\phi(L)$ on 0 and 1 we get

$$\phi(L) = \phi(1)L + \Delta\phi^{\dagger}(L), \tag{3}$$

where $\phi^{\dagger}(L) = 1 - \sum_{j=1}^{p-1} \phi_j^{\dagger} L^j$, and $\phi_j^{\dagger} = -\sum_{i>j}^{p} \phi_i$.

In order to prove the above result, equate the coefficients corresponding to the same power of L on both sides of Equation (3) to get

$$L^0 \Rightarrow \phi_0 = 1 = \phi_0^{\dagger}$$

$$L^1 \Rightarrow -\phi_1 = \phi(1) - \phi_1^{\dagger} + \phi_0^{\dagger} \Rightarrow \phi_1^{\dagger} = -\sum_{i=2}^{p} \phi_i$$

$$L^2 \Rightarrow -\phi_2 = -\phi_2^{\dagger} + \phi_1^{\dagger} \Rightarrow \phi_1^{\dagger} = -\sum_{i=3}^{p} \phi_i$$

...

In view of (3), if we premultiply both sides of Equation (2) by $\phi(L)$ we get

$$\Delta\phi^{\dagger}(L)y_t + \phi(1)y_{t-1} = \phi(1)\alpha + \phi(1)\beta(t-1) + \phi^{\dagger}(1)\beta + \varepsilon_t$$

which in turn can be rewritten as follows

$$\Delta y_t = [\phi(1)\alpha + \phi^{\dagger}(1)\beta - \phi(1)\beta] + \phi(1)\beta t - \phi(1)y_{t-1} + \sum_{j=1}^{p-1} \phi_j^{\dagger}\Delta y_{t-j} + \varepsilon_t$$

3 Stationarity conditions of the VAR

Consider the VAR model

$$\Pi(L)Y_t = c + \varepsilon_t \tag{4}$$

There are two equivalent ways to state the stationarity conditions:

The first condition states that the solutions of

$$\det[\Pi(z)] = 0 \tag{5}$$

must lie outside the complex unit circle.

Condition (5) comes from the following equation

$$\Pi(L)^{-1} = \text{Adj}[\Pi(L)]/\det[\Pi(L)] \tag{6}$$

Inverting both sides of (6) one gets

$$\Pi(L) = \det[\Pi(L)]\text{Adj}[\Pi(L)]^{-1} \tag{7}$$

Substituting Equation (7) into (4) one finally gets the so-called final equation representation of series Y_t:

$$\det[\Pi(L)]Y_t = \text{Adj}[\Pi(1)]c + \text{Adj}[\Pi(L)]\varepsilon_t \tag{8}$$

Notice Adj[$\Pi(L)$] is a polynomial matrix of order $(n-1)p$, hence the RHS of (8) is a VMA($(n-1)p$), which is always stationary.

It follows in turn that stationarity of series Y_t only depends on the roots det[$\Pi(L)$], which is a polynomial scalar of order np.

Before introducing the second condition, let us first look at the VAR(1)

$$Y_t = c + \Pi Y_{t-1} + \varepsilon_t \tag{9}$$

and notice that the matrix Π can be factorized as

$$\Pi = V \Lambda V^{-1}$$

where V and Λ are respectively the eigenvector matrix and the (diagonal) eigenvalue matrix of Π.
Multiplying both sides of (9) by V^{-1} one gets

$$\underbrace{V^{-1} Y_t}_{X_t} = V^{-1} c + \Lambda \underbrace{V^{-1} Y_{t-1}}_{X_{t-1}} + V^{-1} \varepsilon_t \tag{10}$$

Notice that elements of X_t are AR(1) processes, which are then stationary when the eigenvalues of Π lie in $(-1, 1)$. Given that $Y_t = V X_t$, the same conclusion applies to series Y_t as well.

Let us now rewrite a VAR(p) model with $p > 1$ into its companion form

$$\underbrace{\begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{bmatrix} \Pi_1 & \Pi_2 & \cdots & \Pi_p \\ I_n & 0_{n\times n} & \cdots & 0_{n\times n} \\ \vdots & \ddots & \cdots & 0_{n\times n} \\ 0_{n\times n} & \cdots & I_n & 0_{n\times n} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix}}_{\mathbf{Y}_{t-1}} + \underbrace{\begin{bmatrix} c + \varepsilon_t \\ 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \end{bmatrix}}_{\mathbf{E}_t}$$

which shows that any VAR(p) model can rewritten as a np-dimensional VAR(1) model. *In Italian: Varone!* ;-)

From the previous result on the stationarity conditions for the VAR(1), it follows that series $\mathbf{Y}_t$ (and Y_t by implication) are stationary when the eigenvalues of F lie in $(-1, 1)$.

4 The VAR model in matrix format

The VAR(p) model

$$\begin{aligned} Y_t &= c + \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \varepsilon_t \\ t &= 1, 2, ..., T \end{aligned}$$

can be rewritten compactly as follows

$$Y_t = \underbrace{[c, \Pi_1, \cdots, \Pi_p]}_{\Pi'} \underbrace{\begin{bmatrix} 1 \\ Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix}}_{Z_t} + \varepsilon_t$$

The T observations of series Y_t and Z_t can be gathered in the following matrices

$$\underset{T\times n}{\mathbf{Y}} = \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_T' \end{bmatrix}; \underset{T\times k}{\mathbf{Z}} = \begin{bmatrix} Z_1' \\ Z_2' \\ \vdots \\ Z_T' \end{bmatrix}$$

where $k = np + 1$.

Hence, the VAR model can be rewritten in matrix format as

$$\mathbf{Y} = \mathbf{Z}\Pi + \varepsilon \tag{11}$$

where $\varepsilon = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T]'$.

Now are going to meet two new (for some of you) algebraic operators:

- For an $n \times k$ matrix $A = [a_{\cdot 1}, \cdots, a_{\cdot k}]$, where a_j is an n-vector for $j = 1, ..., k$, the Vec operator is defined as

$$\underset{kn \times 1}{\text{Vec}(A)} = \begin{bmatrix} a_{\cdot 1} \\ \vdots \\ a_{\cdot k} \end{bmatrix}$$

- For an $n \times k$ matrix A and an $m \times q$ matrix M the Kronecker product $A \otimes M$ is

$$\underset{nm \times kq}{A \otimes M} = \begin{bmatrix} a_{11}M & \cdots & a_{1k}M \\ \vdots & \ddots & \vdots \\ a_{n1}M & \cdots & a_{nk}M \end{bmatrix}$$

The two operators are linked by the property

$$\text{Vec}(ABC) = (C' \otimes A)\text{Vec}(B) \tag{12}$$

where B is a $k \times m$ matrix and C is an $m \times q$ matrix. You can easily check (12) by verifying that the corresponding generic elements of both sides of Equation (12) are the same.

By taking the Vec of both sides of Equation (11) and using the property (12) in the RHS we get

$$\underbrace{\text{Vec}(\mathbf{Y})}_{\mathbf{Y}^*} \overset{(12)}{=} \underbrace{(I_n \otimes \mathbf{Z})}_{\mathbf{Z}^*}\underbrace{\text{Vec}(\Pi)}_{\Pi^*} + \underbrace{\text{Vec}(\varepsilon)}_{\varepsilon^*}$$

Since it is easy to see that $\text{E}(\varepsilon^*\varepsilon^{*\prime})$ is not a scalar matrix, GLS should be used to optimally estimate Π^*. However, it can be proven (but we don't do it) that GLS and OLS coincide in this case.

Hence, we estimate Π^* as follows

$$\widehat{\Pi}^* = (\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{Y}^*$$

Some additional algebra (that we do not show) reveals that

$$\widehat{\Pi}^* = \text{Vec}(\widehat{\Pi})$$

where $\widehat{\Pi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ is the OLS estimator of Π in Equation (11). Finally, rewriting $\widehat{\Pi}$ as

$$\widehat{\Pi} = [\underset{k\times 1}{\widehat{\Pi}_{\cdot 1}}, \cdots, \underset{k\times 1}{\widehat{\Pi}_{\cdot n}}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\underset{T\times 1}{\mathbf{Y}_{\cdot 1}}, \cdots, \underset{T\times 1}{\mathbf{Y}_{\cdot n}}]$$

we see that $\widehat{\Pi}$ is equivalent to apply OLS to each equation of the VAR (i.e. column of matrix $\mathbf{Y}$):

$$\widehat{\Pi}_{\cdot j} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{\cdot j}; \quad \text{for } j = 1, ..., n]$$

5 Identifying restrictions in SVAR models

The n-dimensional SVAR(p) model is

$$B(L)Y_t = \gamma_0 + \varepsilon_t$$

where

$$B(L) = \underbrace{B}_{\text{the diag. is } 1_n} - \sum_{j=1}^{p} B_j L^j \Rightarrow n(n-1) + n^2 p \text{ parameters}$$

$$\gamma_0 \Rightarrow n \text{ parameters}$$

$$\mathsf{E}(\varepsilon_t \varepsilon_t') = D = \mathsf{diag}(\sigma_1^2, ..., \sigma_n^2) \Rightarrow n \text{ parameters}$$

Hence, the overall number of parameters is $n^2(p+1) + n$.

The reduced-form (RF) representation is

$$\underbrace{B^{-1}B(L)}_{A(L)}Y_t = \underbrace{B^{-1}\gamma_0}_{a_o} + \underbrace{B^{-1}\varepsilon_t}_{u_t}$$

where

$$A(L) = I_n - \sum_{j=1}^{p} B_j L^j \Rightarrow n^2 p \textit{ parameters}$$

$$a_0 \Rightarrow n \textit{ parameters}$$

$$\mathsf{E}(u_t u_t') = \Omega \Rightarrow n(n+1)/2 \textit{ parameters}$$

Hence, the overall number of parameters is $n^2 p + n + n(n+1)/2$.

The number of identifying restrictions on the SVAR is then given by

$$\#\ \textit{parıs of the SVAR} - \#\ \textit{parıs of the RF} = n(n-1)/2$$

Typical identifying restrictions include to assume that either matrix B is a lower triangular

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ b_{n,1} & \cdots & b_{n,n-1} & 1 \end{bmatrix}$$

or that matrix $B(1)$ is lower triangular.

Remark: the SVAR model can be equivalently rewritten as

$$\underbrace{B^*(L)}_{D^{-1/2}B(L)} Y_t = \underbrace{\gamma_0^*}_{D^{-1/2}\gamma_0} + \underbrace{\varepsilon_t^*}_{D^{-1/2}\varepsilon_t}$$

where

$$B^*(L) = B^* - \sum_{j=1}^{p} B_j^* L^j \Rightarrow n^2(p+1) \textit{ parameters}$$

$$\gamma_0^* \Rightarrow n \textit{ parameters}$$

$$\mathsf{E}(\varepsilon_t^* \varepsilon_t^{*\prime}) = I_n \Rightarrow 0 \textit{ parameters}$$

Hence, the overall number of parameters is still $n^2(p+1)+n$.

6 Identification through a triangular system

Consider for simplicity the following bivariate VAR(1) model

$$y_{1t} = a_{10} + a_{11}y_{1t-1} + a_{12}y_{2t-1} + u_{1t} \tag{13}$$

$$y_{2t} = a_{20} + a_{21}y_{1t-1} + a_{22}y_{2t-1} + u_{2t} \tag{14}$$

where $\mathsf{E}(u_t u_t') = \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}$.

Let us consider the conditional expectation

$$\mathsf{E}(y_{2t}|Y_{t-1}, u_{1t}) = a_{20} + a_{21}y_{1t-1} + a_{22}y_{2t-1} + \underbrace{\mathsf{E}(u_{2t}|Y_{t-1}, u_{1t})}_{\frac{\omega_{21}}{\omega_{11}}u_{1t} \equiv \beta u_{1t}} \tag{15}$$

In view of Equation (13), we can write

$$\beta u_{1t} = \beta\left(y_{1t} - a_{10} - a_{11}y_{1t-1} - a_{12}y_{2t-1}\right) \tag{16}$$

Putting together Equations (13), (15) and (16), we obtain the conditional model for series y_{2t} as

$$\begin{aligned} y_{2t} &= a_{20} + a_{21}y_{1t-1} + a_{22}y_{2t-1} + \beta u_{1t} + \underbrace{[u_{2t} - \beta u_{1t}]}_{\varepsilon_{2t}} \\ &= \underbrace{(a_{20} - a_{10}\beta)}_{\gamma_{20}} + \underbrace{\beta}_{-b_{21}} y_{1t} + \underbrace{(a_{21} - a_{11}\beta)}_{\gamma_{21}} y_{1t-1} \\ &+ \underbrace{(a_{22} - a_{12}\beta)}_{\gamma_{22}} y_{2t-1} + \varepsilon_{2t} \end{aligned}$$

where $\mathsf{E}(\varepsilon_{2t}u_{1t}) = 0$ by construction.
Thus, b_{21} is obtained as

$$b_{21} = -\frac{\omega_{21}}{\omega_{11}}$$

and consequently

$$B = \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix}; \varepsilon_t = Bu_t = \begin{bmatrix} u_{1t} \\ u_{2t} + b_{21}u_{1t} \end{bmatrix}$$

7 Identification by the Cholesky factorization

Recall that the RF errors are such that $\mathsf{E}(u_t u_t') = \Omega$, a symmetric positive definite (SPD) matrix that it is not generally diagonal.
The problem of identifying the structural errors is related to the one of finding a linear transformation $z_t = \Upsilon u_t$ such that

$$\mathsf{E}(z_t z_t') = \Upsilon \Omega \Upsilon' = I_n$$

Since Ω is SPD, it can be factorized as

$$\Omega = V \Lambda V'$$

where V and Λ are respectively its eigenvector matrix and its (diagonal) eigenvalue matrix.

Since $V' = V^{-1}$ and the eigenvalues of Ω are positive, we get the factorization

$$\Omega = \underbrace{V\Lambda^{1/2}V'}_{\Omega^{1/2}}\underbrace{V\Lambda^{1/2}V'}_{\Omega^{1/2}}$$

Matrix $\Omega^{1/2}$ is called the square root of Ω, and its inverse is equal to

$$\Omega^{-1/2} = V\Lambda^{-1/2}V'$$

as you may easily check by yourself.
Consider now the linear transformation $z_t = \Omega^{-1/2}u_t$. We have that

$$\mathsf{E}(z_t z_t') = \Omega^{-1/2}\Omega\Omega^{-1/2} = \Omega^{-1/2}\Omega^{1/2}\Omega^{1/2}\Omega^{-1/2} = I_n \qquad (17)$$

However, a problem arises since z_t is not the only linear transformation of u_t that satisfies Equation (17).

Indeed, consider an $n \times n$ matrix M such that $MM' = I_n$ and define $v_t = Mz_t$. Then we have

$$\mathsf{E}(v_t v_t') = M\mathsf{E}(z_t z_t')M' = I_n$$

In words, there are infinite ways to orthogonalize the RF shocks u_t. However, there is a unique lower triangular matrix C with positive diagonal elements that satisfies

$$CC' = \Omega \tag{18}$$

Equation (18) is the Cholesky decomposition of matrix Ω and the factor C is indicated as $C = \mathsf{Chol}(\Omega)$. Notice that C^{-1} is lower triangular as well. Hence, the transformation $\varepsilon_t^* = C^{-1}u_t$ is unique and it satisfies

$$\mathsf{E}(\varepsilon_t^* \varepsilon_t^{*\prime}) = C^{-1}\Omega(C^{-1})' = C^{-1}CC'(C')^{-1} = I_n$$

Equivalently, one may construct $\varepsilon_t = D^{1/2}\varepsilon_t^* = \underbrace{D^{1/2}C^{-1}}_{B}u_t$, where $D^{-1/2}$ is a diagonal matrix with the same diagonal elements as C^{-1}.

8 Identification through long-run restrictions

Economic theory is often more informative on the long-run behavior of the economy (i.e., equilibria) rather than on the short-run one.
An alternative identification scheme imposes a triangular structure on the long-run impact matrix $\Theta(1)$ rather than on Θ_0. This goal is achieved as follows

$$\Theta(L) = \underbrace{\Psi(L)}_{A(L)^{-1}} \underbrace{A(1)C(1)}_{B^{*-1}}$$

where

$$C(1) = \text{Chol}[A(1)^{-1}\Omega A(1)^{-1\prime}]$$

Hence, we have that $\Theta(1)$ is lower triangular since

$$\Theta(1) = \underbrace{\Psi(1)}_{A(1)^{-1}} A(1)C(1) = C(1)$$

Consequently, the (standardized) structural shocks ε_t^* are obtained as

$$\varepsilon_t^* = \underbrace{C(1)^{-1}A(1)^{-1}}_{B^*} u_t$$

and they are orthogonal each other given that

$$\mathsf{E}(\varepsilon_t^* \varepsilon_t^{*\prime}) = C(1)^{-1} \underbrace{A(1)^{-1}\Omega A(1)^{-1\prime}}_{C(1)C(1)'} \underbrace{C(1)^{-1\prime}}_{[C(1)']^{-1}} = I_n$$

Since $B^*(L) = B^*A(L)$, we have that

$$B^*(1) = \underbrace{C(1)^{-1}A(1)^{-1}}_{B^*} A(1) = C(1)^{-1}$$

so even $B^*(1)$ is lower triangular being the inverse of a lower triangular matrix.

9 Interpretation of lower-triangular identification schemes

Given the structural VMA representation$Y_t = \mu + \Theta(L)\varepsilon_t^*$where $\Theta(L) = \Psi(L)B^{*-1}$, we have a causal ordering interpretation when the matrix B^* induces a lower-triangular structure.

In particular, when

$$B^{*-1} = \text{Chol}(\Omega)$$

we have that

$$\Theta_0 = \begin{bmatrix} \vartheta_{11}^0 & 0 & 0 & \cdots & 0 \\ \vartheta_{21}^0 & \vartheta_{22}^0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vartheta_{n1}^0 & \vartheta_{n2}^0 & \vartheta_{n3}^0 & \cdots & \vartheta_{nn}^0 \end{bmatrix}$$

Hence, we see that ε_{1t}^* hits contemporaneously all the elements of Y_t, ε_{2t}^* hits contemporaneously $[y_{2t}, ..., y_{nt}]'$, ... and ε_{nt}^* hits contemporaneously y_{nt} only.

Instead, when

$$B^{*-1} = \Psi(1)^{-1}\text{Chol}[\Psi(1)\Omega\Psi(1)']$$

we have that

$$\Theta(1) = \text{Chol}[\Psi(1)\Omega\Psi(1)'] = \begin{bmatrix} \vartheta_{11}(1) & 0 & 0 & \cdots & 0 \\ \vartheta_{21}(1) & \vartheta_{22}(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vartheta_{n1}(1) & \vartheta_{n2}(1) & \vartheta_{n3}(1) & \cdots & \vartheta_{nn}(1) \end{bmatrix}$$

Hence, ε_{1t}^* has a long-run impact on all the elements of Y_t, ε_{2t}^* has a long-run impact on $[y_{2t}, ..., y_{nt}]'$, ... and ε_{nt}^* has a long-run impact on y_{nt} only.

10 Normalized cointegration vectors

Given that $0 < \text{Rank}(\beta) = r < n$ (if cointegration exists) there exists a $r \times r$ submatrix of β (i.e. a matrix formed by r rows of β) that is full-rank. Hence, without loss of generality, we can write

$$\beta' = [\underset{r\times r}{\beta_1}, \quad \underset{r\times(n-r)}{\beta_2}]$$

where $\text{Rank}(\beta_1) = r$. Then let us write

$$\beta_c' = \beta_1^{-1}[\beta_1, \beta_2] = [I_r, \underbrace{\beta_1^{-1}\beta_2}_{-B}]$$

Let us partition Y_t conformably to β_c', i.e. $Y_t' = [\underset{1\times r}{Y_{1t}'}, \quad \underset{1\times(n-r)}{Y_{2t}'}]$. Then we can write

$$\beta_c' Y_t = Y_{1t} - BY_{2t} \backsim I(0)$$

Remark: in empirical applications, one should pay attention in allocating in the first rows of Y_t the variables that are most likely to enter in equilibrium relations.
Consider e.g. the case $Y_t' = [y_{1t}, y_{2t}, y_{3t}]'$ with $r = 1$ and assume that y_{1t} has a null coefficient in the cointegration vector, i.e.

$$\beta' = [0, b_2, b_3]$$

It is then clear that we cannot normalize the first coefficient of β to 1, unless we permute the rows of Y_t in a way such that y_{1t} is not allocated in the 1st row anymore.

11 Reduced-rank regression and cointegration

Consider the VECM representation

$$\Gamma(L)\Delta Y_t = \Phi D_t + \alpha\beta' Y_{t-1} + \varepsilon_t \tag{19}$$

1. In order to concentrate Model (19) on the parameters α and β, remove the linear influence of $X_t = \left[D_t, \Delta Y_{t-1}, ..., \Delta Y_{t-p+1}\right]'$ from both ΔY_t and Y_{t-1} to get, respectively, $R_{0,t}$ and $R_{1,t}$. The concentrated model reads

$$R_{0,t} = \alpha\beta' R_{1,t} + \varepsilon_t \tag{20}$$

   Model (20) is known in statistics as the Reduced-Rank Regression model (Anderson, 1951).

2. In order to make the method invariant to linear transformations of the variables, standardize both $R_{0,t}$ and $R_{1,t}$ and transform Model (20) into the following one:

$$\underbrace{S_{00}^{-1/2}R_{0,t}}_{R_{0,t}^*} = \underbrace{S_{00}^{-1/2}\alpha}_{\alpha_*}\underbrace{\beta' S_{11}^{1/2}}_{\beta_*'}\underbrace{S_{11}^{-1/2}R_{1,t}}_{R_{1,t}^*} + \underbrace{S_{00}^{-1/2}\varepsilon_t}_{\varepsilon_t^*} \tag{21}$$

where $S_{ij} = \mathsf{E}(R_{i,t}R_{j,t}')$ for $i, j = 0, 1$.

3. Post-multiply both sides of Equation (21) by $R_{1,t}^{*\prime}$ and take the expected value on both sides to get

$$S_{01}^* = \alpha_*\beta_*' \tag{22}$$

where $S_{ij}^* = \mathsf{E}(R_{i,t}^*R_{j,t}^{*\prime})$ for $i, j = 0, 1$ (notice that $S_{11}^* = I_n$ by construction).

4. In view of Equations (21) and (22), get the equation

$$S_{10}^{*}S_{01}^{*} = \beta_{*}\alpha_{*}'\alpha_{*}\beta_{*}' = \underbrace{S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}}_{S} \tag{23}$$

5. Since matrix S is symmetric and semi-definite positive, get the eigen-decomposition

$$S = V\Lambda V' \tag{24}$$

where Λ is the diagonal eigenvalue matrix and V is the eigenvector matrix.

6. Since Rank(S) $= r$, then $n-r$ eigenvalues of S are zero and Equation (24) boils down to

$$S = [V_r, V_{n-r}] \begin{bmatrix} \underset{r\times r}{\Lambda_r} & \underset{r\times(n-r)}{0} \\ \underset{(n-r)\times r}{0} & \underset{(n-r)\times(n-r)}{0} \end{bmatrix} \begin{bmatrix} V_r' \\ V_{n-r}' \end{bmatrix} = V_r \Lambda_r V_r' \quad (25)$$

where Λ_r denotes the diagonal matrix of the r positive eigenvalues of S, and V_r $[V_{n-r}]$ denote the eigenvectors associated with the positive [zero] eigenvalues.

7. In view of Equations (23) and (25), we conclude that V_r spans the same space as $\beta_* = S_{11}^{1/2}\beta$. Hence, $S_{11}^{-1/2}V_r$ spans the same space as the cointegration matrix β.

- Johansen (1991) shows that the sample analogous of the eigendecomposition (24) solves the Gaussian ML problem for Model (19).

- In particular, let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_n$ be the ordered eigenvalues of

$$\widehat{S} = \widehat{S}_{11}^{-1/2} \widehat{S}_{10} \widehat{S}_{00}^{-1} \widehat{S}_{01} \widehat{S}_{11}^{-1/2}$$

  where $\widehat{S}_{ij}$ is the sample covariance matrix between residuals $\widehat{R}_{i,t}$ and $\widehat{R}_{j,t}$ for $i, j = 0, 1$. These eigenvalues are also equal to the squares of the canonical correlations between ΔY_t and Y_{t-1} corrected for X_t (i.e., the correlations coefficients between the linear combinations of $\widehat{R}_{0,t}$ and $\widehat{R}_{1,t}$ that are mostly correlated) and so lie between 0 and 1.

- Recall, $r = \text{Rank}(\Pi) = \text{Rank}(S)$ and so r equals the number of non-zero eigenvalues of S.

12 Identification of the VECM through Information Criteria

Identification criteria (IC) may be used to simultaneously determine the lag order p and the cointegration rank r. The approach requires to estimate a VECM with $p = 1, ..., p_{\text{max}}$ and, for each p, $r = 0, 1, ..., n$, and choose the couple (p^*, r^*) that minimizes:

$$IC(p, r) = \widehat{\Sigma}(p, r) + c(T)\varphi(n, p, r)$$

where $\widehat{\Sigma}(p, r)$ is the residual covariance matrix estimate from the reduced rank regression estimation of a VECM with p lags and rank r, $c(T)$ is a function of the sample size T, and

$$\varphi(n, p, r) = (p - 1)n^2 + r(2n - r)$$

is the number of parameters of the model (*why*?)

For the most common IC, we have

$$c(T) = \begin{cases} 2/T & \Rightarrow AIC \\ 2\ln[\ln(T)]/T & \Rightarrow HQIC \\ \ln(T)/T & \Rightarrow AIC \end{cases}$$

Simulation studies indicate that simultaneous determination of (p, r) through IC, in particular the HQIC, outperforms the sequential procedure (first p, then r) through the LR tests, in particular when the sample size T is small (e.g., $T = 50$) or the number of series n is large (e.g., $n = 9$).

13 Granger representation theorem

Consider the VAR representation

$$\Pi(L)(Y_t - \mu_t) = \varepsilon_t \tag{26}$$

where $\mu_t = \mathsf{E}(y_t)$, and the Wold representation

$$(1 - L)(Y_t - \mu_t) = \Psi(L)\varepsilon_t \tag{27}$$

Premultiply both sides of (26) by $\Psi(L)$ to get

$$\Psi(L)\Pi(L)(Y_t - \mu_t) = \Psi(L)\varepsilon_t = (1 - L)(Y_t - \mu_t)$$

Premultiply both sides of (27) by $\Pi(L)$ to get

$$(1 - L)\Pi(L)(Y_t - \mu_t) = \Pi(L)\Psi(L)\varepsilon_t = (1 - L)\varepsilon_t$$

From the two previous equations we see

$$\Psi(L)\Pi(L) = \Pi(L)\Psi(L) = I_n(1-L),$$

which for $L = 1$ gives

$$\Psi(1)\Pi(1) = \Pi(1)\Psi(1) = 0$$

Since $\Pi(1) = -\alpha\beta'$ and α and β are full-column rank matrices, we see from the equation above that

$$\Psi(1)\alpha = 0, \quad \beta'\Psi(1) = 0$$

and hence

$$\Psi(1) = \beta_\perp \Xi \alpha'_\perp$$

where Ξ is full-rank square matrix of dimension $(n-r)$, and $A_\perp$ indicates the orthogonal complement of a matrix A such that $A'_\perp A = 0$.

Moreover, from the relation

$$\Pi(L)\Psi(L) = (1 - L)I_n$$

it follows that

$$\Pi^{\dagger}(L)\Psi(L) + \Pi(L)\Psi^{\dagger}(L) = -I_n$$

where $A^{\dagger}(L)$ indicates the first derivative of a polynomial matrix $A(L)$. If we evaluate the above equation for $L = 1$ we get

$$\Pi^{\dagger}(1)\Psi(1) + \Pi(1)\Psi^{\dagger}(1) = -I_n$$

which can be rewritten as

$$\Pi^{\dagger}(1)\beta_{\perp}\Xi\alpha'_{\perp} - \alpha\beta'\Psi^{\dagger}(1) = -I_n \tag{28}$$

Premultiplying both sides of (28) by $\alpha'_\perp$ we get

$$\alpha'_\perp \Pi^\dagger(1)\beta_\perp \Xi \alpha'_\perp \overset{(\alpha'_\perp \alpha = 0)}{=} -\alpha'_\perp$$

Premultiplying both sides of the above equation by $(\alpha'_\perp \Pi^\dagger(1)\beta_\perp)^{-1}$ [*Remark*: Johansen (1995) proved that $\alpha'_\perp \Pi^\dagger(1)\beta_\perp$ is invertible iff elements of Y_t are, at most, $I(1)$] we get

$$\Xi \alpha'_\perp = -(\alpha'_\perp \Pi^\dagger(1)\beta_\perp)^{-1}\alpha'_\perp$$

and finally postmultipliying both sides of the above equation by $\alpha_\perp(\alpha'_\perp \alpha_\perp)^{-1}$ [*Remark*: $\alpha'_\perp \alpha_\perp$ is invertible because we assumed that $\alpha_\perp$ has full rank] we get

$$\Xi = -(\alpha'_\perp \Pi^\dagger(1)\beta_\perp)^{-1} \tag{29}$$

Notice that from the expansion

$$\Pi(L) = (1 - L)\Gamma(L) - \alpha\beta' L$$

we easily get

$$\Pi^{\dagger}(L) = (1 - L)\Gamma^{\dagger}(L) - \Gamma(L) - \alpha\beta'$$

If we evaluate the above equation for $L = 1$ we get

$$\Pi^{\dagger}(1) = -\Gamma(1) - \alpha\beta'$$

which can be substituted into equation (29) to finally obtain

$$\Xi = (\alpha'_{\perp}\Gamma(1)\beta_{\perp})^{-1}$$

and hence

$$\Psi(1) = \beta_{\perp}(\alpha'_{\perp}\Gamma(1)\beta_{\perp})^{-1}\alpha'_{\perp}$$

Assuming that $\mu_t = Y_0 + \mu t$, the multivariate Beveridge-Nelson (BN) decomposition

$$Y_t = Y_0 + \mu t + \Psi(1) \sum_{i=0}^{t-1} \varepsilon_{t-i} + \underbrace{\widetilde{\Psi}(L)\varepsilon_{t-j}}_{\widetilde{\varepsilon}_t \backsim I(0)}$$

where $\widetilde{\Psi}(L) = \sum_{j=0}^{\infty} \widetilde{\Psi}_j L^J$ is such that $\widetilde{\Psi}_j = -\sum_{i>j}^{\infty} \Psi_i$ and $\lim_{j\to\infty} |\widetilde{\Psi}_j| = 0$, can hence be rewritten as the following Common Trend (CT) representation

$$Y_t = Y_0 + \mu t + \gamma \tau_t + \widetilde{\varepsilon}_t \tag{30}$$

where $\gamma = \beta_\perp(\alpha'_\perp \Gamma(1)\beta_\perp)^{-1}$, and the $(n-r)$ common stochastic trends are

$$\tau_t = \alpha'_\perp \sum_{i=0}^{t-1} \varepsilon_{t-i}$$

Notice that by premultiplying both sides of (30) by β' one gets

$$\beta' Y_t = \beta' Y_0 + \beta' \mu t + \beta' \widetilde{\varepsilon}_t$$

given that $\beta' \gamma = \beta' \beta_\perp (\alpha'_\perp \Gamma(1) \beta_\perp)^{-1} = 0$.
Hence, the same linear combinations that stationarize variables Y_t kill the random walk components in the multivariate BN decomposition.

Finally, notice that all the parameters of the CT representation are functions of the parameters of the VECM.
This implies that we can estimate the CT representation through the estimated VECM.

14 Empirical applications

- Monthly data (1959.01-2019.12) for:
  - S&P 500S&P's Common Stock Price Index: Composite
  - S&P: industS&P's Common Stock Price Index: Industrials
  - S&P div yield S&P's Composite Common Stock: Dividend Yield
  - S&P PE ratio S&P's Composite Common Stock: Price-Earnings Ratio

- Monthly data (1959.01-2020.11) for:
  - 3-Month Treasury Constant Maturity rate (C) Minus Federal Funds rate (FEDFUNDS)
  - 6-Month Treasury C Minus FEDFUNDS
  - 1-Year Treasury C Minus FEDFUNDS
  - 5-Year Treasury C Minus FEDFUNDS5
  - 10-Year Treasury C Minus FEDFUNDS