Time Series Econometrics, 2ST111

Lecture 10. Unit Roots in Multivariate Time Series and Cointegration

Yukai Yang

Department of Statistics, Uppsala University

Outline of Today's Lecture

- Unit Roots in Multivariate Time Series
  - Asymptotic Results for Nonstationary Vector Processes
  - Vector Autoregressions Containing Unit Roots
  - Spurious Regressions
- Cointegration
  - Introduction

Multivariate Standard Brownian Motion

We introduce the multivariate standard Brownian Motion, the definition on pp.544 in Hamilton,

Definition

n-dimensional standard Brownian motion $\mathbf{W}(\cdot)$ is a continuous-time process associating each data $r \in [0, 1]$ with the $(n \times 1)$ vector $\mathbf{W}(r)$ satisfying the following:

1. $\mathbf{W}(0) = 0$;
2. For any dates $0 \leq r_1 < r_2 < \ldots < r_k \leq 1$, the changes $[\mathbf{W}(r_2) - \mathbf{W}(r_1)], [\mathbf{W}(r_3) - \mathbf{W}(r_2)], \ldots, [\mathbf{W}(r_k) - \mathbf{W}(r_{k-1})]$ are independent multivariate Gaussian with $\mathbf{W}(s) - \mathbf{W}(r) \sim N_n(\mathbf{0}, (s - r)\mathbf{I}_n)$;
3. For any given realization, $\mathbf{W}(r)$ is continuous in r with probability 1.

Multivariate Standard Brownian Motion

Remarks:

- The univariate standard Brownian motion (BM) is a special case of the multivariate standard Brownian motion (MBM or VBM), $n = 1$.
- The univariate BM can be easily extended to the VBM by adding more independent univariate BMs. Note that the covariance for $\mathbf{W}(s) - \mathbf{W}(r)$ is simply $(s - r)\mathbf{I}_n$ implying that they are independent.
- You can simply change all the scalars in the previous lecture to vectors, y_t to $\mathbf{y}_t$, ε_t to $\boldsymbol{\varepsilon}_t$, the same results hold for the multivariate Brownian motion.
- The matrix $\mathbf{W}(r)\mathbf{W}(r)'$ is Wishart distributed.

Functional Central Limit Theorem

Suppose that $\varepsilon_1, ..., \varepsilon_T$ is an *i.i.d.* n-vector sequence with mean zero and covariance $\mathbf{I}_n$.

Then we can construct a random vector $\mathbf{X}_T(r)$ that uses only the first rth fraction of the sample $\varepsilon_1, ..., \varepsilon_T$

$$\mathbf{X}_T(r) = \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} \varepsilon_t, \tag{1}$$

where $0 \leq r \leq 1$.

The corresponding functional central limit theorem

$$\sqrt{T}\mathbf{X}_T(\cdot) \xrightarrow{d} \mathbf{W}(\cdot) \tag{2}$$

for $0 \leq r \leq 1$. Note that $\mathbf{W}(r) \sim N_n(\mathbf{0}, r\mathbf{I}_n)$.

Functional Central Limit Theorem

Suppose that there is another vector sequence $\mathbf{v}_t$ such that $\mathbf{v}_t = \mathbf{P}\varepsilon_t$, and that $\mathbf{PP}' = \mathbf{\Omega}$ which is positive definite. The sequence of $\mathbf{v}_t$ is *i.i.d.* with mean zero and covariance $\mathbf{\Omega}$.

Let $\mathbf{X}_T^*(r)$ be

$$\mathbf{X}_T^*(r) = \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} \mathbf{v}_t, \tag{3}$$

where $0 \leq r \leq 1$.

Then the functional central limit theorem

$$\sqrt{T}\mathbf{X}_T^*(\cdot) \xrightarrow{d} \mathbf{P} \cdot \mathbf{W}(\cdot) \tag{4}$$

for $0 \leq r \leq 1$. Note that $\mathbf{PW}(r) \sim N_n(\mathbf{0}, r\mathbf{\Omega})$.

Vector I(0) process

Recall that

- A linear zero-mean vector I(0) process is a vector MA(∞) process

$$\mathbf{u}_t = \boldsymbol{\Psi}(L)\mathbf{v}_t, \ \boldsymbol{\Psi}(L) = \boldsymbol{\Psi}_0 + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 + \dots, \tag{5}$$

  satisfying 2 conditions:

  1. $\boldsymbol{\Psi}(1) \neq \mathbf{0}$ (but not necessarily of full rank), (ensures I(0)) and
  2. the matrix $\boldsymbol{\Psi}_s = (\psi_{ij}^{(s)})$ is one-summable, meaning $\sum_{s=0}^{\infty} s|\psi_{ij}^{(s)}| < \infty$ for all $i, j = 1, 2, \dots, n$ (we need it later).

- The Long-Run coVariance (LRV) matrix of $\mathbf{u}_t$ is

$$\mathrm{LRV}(\mathbf{u}_t) \equiv \lim_{T\to\infty} \mathrm{Var}[\sqrt{T}\bar{\mathbf{u}}_T] = \boldsymbol{\Psi}(1)\boldsymbol{\Omega}\boldsymbol{\Psi}(1)'. \tag{6}$$

Vector I(1) process

- A vector I(1) process is defined as

$$\Delta \mathbf{y}_t = \boldsymbol{\delta} + \mathbf{u}_t \tag{7}$$

where $\mathbf{u}_t = \boldsymbol{\Psi}(L)\mathbf{v}_t$ and $\boldsymbol{\delta} = \mathrm{E}(\Delta \mathbf{y}_t)$ is a vector of constants. Hence,

$$\Delta \mathbf{y}_t = \boldsymbol{\delta} + \boldsymbol{\Psi}(L)\mathbf{v}_t \tag{8}$$

is the vector moving average (VMA) representation of a vector I(1) process.

- In levels, $\mathbf{y}_t$ can be written as

$$\mathbf{y}_t = \mathbf{y}_0 + \boldsymbol{\delta} t + \sum_{s=1}^{t} \mathbf{u}_s \tag{9}$$

Beveridge-Nelson decomposition

- Using $\boldsymbol{\Psi}(L) = \boldsymbol{\Psi}(1) + \Delta\boldsymbol{\alpha}(L)$ where $\boldsymbol{\alpha}(L) = \sum_{j=0}^{\infty} \boldsymbol{\alpha}_j L^j$, with $\boldsymbol{\alpha}_j = -(\boldsymbol{\Psi}_{j+1} + \boldsymbol{\Psi}_{j+2} + \ldots)$ for $j = 0, 1, \ldots$, we can write

$$\mathbf{u}_t = \boldsymbol{\Psi}(1)\mathbf{v}_t + \boldsymbol{\eta}_t - \boldsymbol{\eta}_{t-1}, \tag{10}$$

  where $\boldsymbol{\eta}_t = \boldsymbol{\alpha}(L)\mathbf{v}_t$ is a zero-mean I(0) process, and $\boldsymbol{\alpha}(L)$ is absolutely summable (ensured by the one-summability).
- Substitution of (10) in (9) gives

$$\mathbf{y}_t = \underbrace{\delta t}_{\text{linear trend}} + \underbrace{\boldsymbol{\Psi}(1)\sum_{s=1}^{t}\mathbf{v}_s}_{\text{stochastic trend}} + \underbrace{\boldsymbol{\eta}_t}_{\text{cycle}} + \underbrace{\mathbf{y}_0 - \boldsymbol{\eta}_0}_{\text{initial condition}} \tag{11}$$

Now suppose that $\boldsymbol{\delta} = \mathbf{0}$ and $\mathbf{y}_0 = \mathbf{0}$.

$$\mathbf{y}_t = \sum_{s=1}^{t} \mathbf{u}_s = \boldsymbol{\Psi}(1) \sum_{s=1}^{t} \mathbf{v}_s + \boldsymbol{\eta}_t - \boldsymbol{\eta}_0 \tag{12}$$

Let $\mathbf{X}_T^{**}(r)$ be

$$\mathbf{X}_T^{**}(r) = \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} \mathbf{u}_t = \mathbf{y}_{\lfloor rT \rfloor} / T, \tag{13}$$

where $0 \leq r \leq 1$.

Then the functional central limit theorem

$$\sqrt{T} \mathbf{X}_T^{**}(\cdot) \xrightarrow{d} \boldsymbol{\Psi}(1) \cdot \mathbf{P} \cdot \mathbf{W}(\cdot) \tag{14}$$

for $0 \leq r \leq 1$. Note that $\boldsymbol{\Psi}(1)\mathbf{P}\mathbf{W}(r) \sim N_n(\mathbf{0}, r\boldsymbol{\Psi}(1)\boldsymbol{\Omega}\boldsymbol{\Psi}(1)')$.

Asymptotic Results for Nonstationary Vector Processes

- Proposition 18.1 on pp.547 in Hamilton summarizes the results.
- Note that we use different notations compared to the ones in Hamilton.
- The differences are mainly, in Hamilton,
  - $\mathbf{v}_t$ for *i.i.d.* error vectors with covariance $\mathbf{I}_n$ while we use ε_t;
  - ε_t for $\mathbf{P}\mathbf{v}_t$ while we use $\mathbf{v}_t$;
  - $\boldsymbol{\xi}_t = \sum_{s=1}^{t} \mathbf{u}_s$ while we use $\mathbf{y}_t$.

An Alternative Representation of a VAR(p) Process

Consider the following VAR(p) process

$$\mathbf{\Phi}(L)\mathbf{y}_t = \boldsymbol{\alpha} + \mathbf{v}_t, \tag{15}$$

where $\mathbf{\Phi}(L) = \mathbf{I}_n - \mathbf{\Phi}_1 L - ... - \mathbf{\Phi}_p L^p$, and $\mathbf{v}_t$ is defined as before.

The lag polynomial can be rewritten as

$$\begin{aligned}\mathbf{\Phi}(L) &= \mathbf{I}_n - \mathbf{\Phi}_1 L - ... - \mathbf{\Phi}_p L^p \\ &= (\mathbf{I}_n - \boldsymbol{\rho} L) - (\zeta_1 L + \zeta_2 L^2 + ... + \zeta_{p-1} L^{p-1})(1 - L)\end{aligned}$$

where $\boldsymbol{\rho} = \sum_{s=1}^{p} \mathbf{\Phi}_s$ and $\zeta_s = -(\mathbf{\Phi}_{s+1} + \mathbf{\Phi}_{s+2} + ...\mathbf{\Phi}_p)$.

Then it follows that

$$\mathbf{y}_t = \boldsymbol{\rho}\mathbf{y}_{t-1} + \zeta_1 \Delta\mathbf{y}_{t-1} + \zeta_2 \Delta\mathbf{y}_{t-2} + ... + \zeta_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \mathbf{v}_t. \tag{16}$$

A Very Strong Assumption

- If $\mathbf{l}_n = \boldsymbol{\rho}$, we can equivalently analyze the first-order difference of $\mathbf{y}_t$, which is a VAR$(p-1)$ process.

$$\Delta \mathbf{y}_t = \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \ldots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \mathbf{v}_t. \quad (17)$$

- By assuming that $\zeta(L)$ is stable, the model can be

$$\Delta \mathbf{y}_t = \zeta(1)^{-1} \boldsymbol{\alpha} + \zeta(L)^{-1} \mathbf{v}_t = \boldsymbol{\delta} + \mathbf{u}_t, \quad (18)$$

  where $\zeta(L)^{-1} = \boldsymbol{\Psi}(L)$, $\boldsymbol{\delta} = \boldsymbol{\Psi}(1)\boldsymbol{\alpha}$ and $\mathbf{u}_t = \boldsymbol{\Psi}(L)\mathbf{v}_t$.
- The assumption $\mathbf{l}_n = \boldsymbol{\rho}$ is so strong that we could rarely find it in reality.
- Note that $\mathbf{l}_n = \boldsymbol{\rho}$ implies $|\mathbf{l}_n - \boldsymbol{\rho}| = 0$, but the other way around does not hold.
- First we consider the testing for the case $\mathbf{l}_n = \boldsymbol{\rho}$, and then the more interesting case $|\mathbf{l}_n - \boldsymbol{\rho}| = 0$ follows.

The Case with No Drift

- The regression model

$$\mathbf{y}_t = \boldsymbol{\rho}\mathbf{y}_{t-1} + \zeta_1 \Delta \mathbf{y}_{t-1} + \zeta_2 \Delta \mathbf{y}_{t-2} + \ldots + \zeta_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \mathbf{v}_t.$$

- Assumptions: the lag polynomial $\zeta(L)$ is stable, or equivalently the roots of the polynomial

$$|\mathbf{I}_n - \zeta_1 z - \zeta_2 z^2 - \ldots - \zeta_{p-1} z^{p-1}| = 0 \tag{19}$$

are all outside the unit disk.

- Null hypothesis $H_0 : \boldsymbol{\rho} = \mathbf{I}_n$ and $\boldsymbol{\alpha} = \mathbf{0}$
- From [18.2.18] on pp.551, we see that the estimators have different rates of convergence. In particular, $\hat{\boldsymbol{\rho}} - \mathbf{I}_n = O_p(T^{-1})$
- The test is given by [18.2.25] on pp.552. Note that the parameters are split into two parts, ζs and $(\boldsymbol{\alpha}, \boldsymbol{\rho})$.

The Case with Drift

- The regression model

$$\mathbf{y}_t = \boldsymbol{\rho}\mathbf{y}_{t-1} + \boldsymbol{\zeta}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\zeta}_2 \Delta \mathbf{y}_{t-2} + \ldots + \boldsymbol{\zeta}_{p-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\alpha} + \mathbf{v}_t.$$

- Assumptions: the lag polynomial $\zeta(L)$ is stable and $\boldsymbol{\alpha} \neq \mathbf{0}$.
- Null hypothesis $H_0 : \boldsymbol{\rho} = \mathbf{I}_n$
- Note that there is a reparametrization in [18.2.43] on pp.556!
- The rates of convergence of the estimators are shown in [18.2.45] on pp.556.
- The test for equation i is given by [18.2.49] on pp.557. Note that the parameters are split into two parts, ζ_is and $(\boldsymbol{\alpha}_i^*, \boldsymbol{\rho}_i^*, \gamma_i)$.

Remarks:

- We can test all the parameters including ζs. The "null hypothesis" in previous pages stresses that they are somewhat "assumed" but still the zero parameters are put inside the regression.
- We see that the limiting distributions of these parameters depend closely on the assumptions or (better say) beliefs.
- Though these tests are not so often used in reality, but the tests for the case $n = 1$ is widely used, which are exactly the augmented Dickey-Fuller tests.

Spurious Regressions

- Consider the I(1) vector sequence $\mathbf{y}_t$ whose difference is simply

$$\Delta \mathbf{y}_t = \varepsilon_t,$$

  where ε_t has covariance $\mathbf{I}_n$.
- For simplicity, assume $n = 2$. Let us regress the following model

$$y_{1t} = \alpha + \gamma y_{2t} + \epsilon_t \tag{20}$$

- We know that $\alpha = 0$ and $\gamma = 0$.
- However,

$$\begin{pmatrix} T^{-1/2}\hat{\alpha}_T \\ \hat{\gamma}_T \end{pmatrix} \xrightarrow{d} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \tag{21}$$

  where h_1 and h_2 are given by [18.3.9] on pp.559.
- We see that neither of them are consistent because h_1 and h_2 have random variables with non-zero expectation and non-zero variances.
- Even worse, $\hat{\alpha}_T$ diverges.

Engel and Granger's cointegration:

- The I(1) process $\mathbf{y}_t$ defined in (8) is cointegrated with cointegrating vector $\mathbf{a} \neq 0$ (of dimension $n \times 1$), if $\mathbf{a}'\mathbf{y}_t$ is trend-stationary.
- Multiplying (11) on both sides by $\mathbf{a}'$, we obtain

$$\mathbf{a}'\mathbf{y}_t = \mathbf{a}'\delta t + \mathbf{a}'\mathbf{\Psi}(1)\sum_{s=1}^{t}\mathbf{v}_s + \mathbf{a}'\eta_t + \mathbf{a}'(\mathbf{y}_0 - \eta_0) \tag{22}$$

  and $\mathbf{a}'\mathbf{y}_t$ is trend-stationary if $\mathbf{a}'\mathbf{\Psi}(1) = \mathbf{0}'$.
- The example on pp.572 about the purchasing power parity (PPP).
- The sufficient condition for $\mathbf{a}'\mathbf{\Psi}(1) = \mathbf{0}'$ holds for a non-zero vector $\mathbf{a}$ is that $\mathbf{\Psi}(1)$ has reduced rank.
- Suppose that the null space of $\mathbf{\Psi}(1)$ has dimension h. There exist h $\mathbf{a}$ vectors who are linearly independent such that $\mathbf{A}'\mathbf{\Psi}(1) = \mathbf{0}'$, where $\mathbf{A} = (\mathbf{a}_1, ..., \mathbf{a}_h)$. $\mathbf{A}$ is not unique!

Related Concepts

- The cointegrating (CI) rank is the number of linearly independent cointegrating vectors. If the CI rank is equal to h, then $\mathrm{rank}(\Psi(1)) = n - h$.
- The CI space is the space spanned by the cointegrating vectors.
- It may also happen that $\mathbf{a}'\boldsymbol{\delta} = 0$. Then $\mathbf{a}'\mathbf{y}_t$ is stationary, rather than trend stationary.
- In that case $\mathrm{rank}([\boldsymbol{\delta} \vdots \boldsymbol{\Psi}(1)]) = \mathrm{rank}(\boldsymbol{\Psi}(1))$: since the left matrix has $1 + n$ columns and n rows, its rank can't exceed $\mathrm{rank}(\boldsymbol{\Psi}(1))$. This implies that $\boldsymbol{\delta}$ is a linear combination of the columns of $\boldsymbol{\Psi}(1)$.

Implications

- If $h = n$, $\mathrm{rank}(\mathbf{\Psi}(1)) = n - h = 0 \Rightarrow \mathbf{\Psi}(1) = \mathbf{0}$, which is ruled out since $\mathbf{u}_t$ is I(0).
- $\mathrm{LRV}(\Delta \mathbf{y}_t) = \mathrm{LRV}(\mathbf{u}_t) = \mathbf{\Psi}(1)\mathbf{\Omega}\mathbf{\Psi}(1)'$ is positive definite if and only if $\mathbf{\Psi}(1)$ is of full rank.
- Therefore, $\mathbf{y}_t$ cannot be cointegrated if $\mathrm{LRV}(\Delta \mathbf{y}_t)$ is positive definite. In this case, each element of $\Delta \mathbf{y}_t$ has its $\mathrm{LRV} > 0$ and is a univariate I(1) process.
- Let $\mathbf{y}_t' = (y_{1t}, \mathbf{y}_{2t}')$ and $\mathbf{a}' = (a_1, \mathbf{a}_2')$. If $h = 1$ and $a_1 \neq 0$, then y_{1t} is cointegrated with some elements in $\mathbf{y}_{2t}$ (i.e. $\mathbf{a}_2 \neq 0$).
  But $\mathbf{y}_{2t}$ is not cointegrated (itself) without y_{1t}, i.e. there is no CI vector such as $(0, \mathbf{b}')$.
  If $h > 1$ (more than 1 CI vector), then $\mathbf{y}_{2t}$ is also cointegrated (itself).

The Stock-Watson Common Trend Representation

- Fact 1: if $\mathrm{rank}(\boldsymbol{\Psi}(1)) = n - h$, there exists a non-singular $n \times n$ matrix $\mathbf{G}$, and a $n \times (n-h)$ matrix $\mathbf{F}$ of full column rank, such that $\boldsymbol{\Psi}(1)\mathbf{G} = [\mathbf{F} \vdots \mathbf{0}_{n\times h}]$.
- Then the stochastic trend component of y_t in (11) can be written

$$\boldsymbol{\Psi}(1)\sum_{s=1}^{t}\mathbf{v}_s = \boldsymbol{\Psi}(1)\mathbf{G}\mathbf{G}^{-1}\sum_{s=1}^{t}\mathbf{v}_s$$

$$= [\mathbf{F} \vdots \mathbf{0}_{n\times h}]\begin{pmatrix}\boldsymbol{\tau}_t \\ \dots \\ \boldsymbol{\upsilon}_t\end{pmatrix} = \mathbf{F}\boldsymbol{\tau}_t$$

Therefore, an I(1) system with a CI rank equal to h has $h - h$ "common stochastic trends", which are the elements of $\boldsymbol{\tau}_t$.

Cointegrated VAR

- We don't use the VMA form to model cointegration. Normally we use the VAR model.
- We need to model $\mathbf{y}_t$, not just $\Delta \mathbf{y}_t$!
- We transform the VAR into its VECM form.

Stationary VAR

- A VAR(p) model with stable lag polynomial, implies that y_t is stationary, therefore not cointegrated. What conditions must be imposed if we want the VAR to allow for cointegration of $\mathbf{y}_t$?
- We write the VAR(p) as

$$\mathbf{y}_t - \mathbf{a} - \mathbf{d}t = \mathbf{w}_t \tag{23}$$

$$\mathbf{\Phi}(L)\mathbf{w}_t = \mathbf{v}_t \tag{24}$$

which is equivalent to

$$\mathbf{\Phi}(L)\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\gamma} t + \mathbf{v}_t \tag{25}$$

for $\boldsymbol{\alpha} = \mathbf{\Phi}(1)\mathbf{a} - (\sum_{j=1}^{p} j\mathbf{\Phi}_j)\mathbf{d}$ and $\boldsymbol{\gamma} = \mathbf{\Phi}(1)\mathbf{d}$.

I(1) VAR and Reduced Rank Condition

- $\mathbf{w}_t \sim \mathrm{I}(1)$ and $\mathbf{v}_t \sim \mathrm{I}(0)$.
- Multiply both sides of (24) by $\Delta = 1 - L$:
  $\mathbf{\Phi}(L)\Delta\mathbf{w}_t = (1 - L)\mathbf{v}_t$, and substitute $\mathbf{\Psi}(L)\mathbf{v}_t$ (Wold representation) for $\Delta\mathbf{w}_t$: $\mathbf{\Phi}(L)\mathbf{\Psi}(L)\mathbf{v}_t = (1 - L)\mathbf{v}_t$.
  This must be true for any $\mathbf{v}_t$, hence

$$\mathbf{\Phi}(L)\mathbf{\Psi}(L) = (1 - L)\mathbf{I}_n \tag{26}$$

- Let $L = 1$, we see that $\mathbf{\Phi}(1)\mathbf{\Psi}(1) = 0$.
- For cointegration, we need $\mathbf{\Psi}(L)$ to be one-summable and $\mathrm{rank}(\mathbf{\Psi}(1)) = n - h$.
- The essential condition for this is that $\mathrm{rank}(\mathbf{\Phi}(1)) = h < n$ (reduced rank).
- Denote $\mathbf{\Pi} = -\mathbf{\Phi}(1)$ hereafter. $\mathrm{rank}(\mathbf{\Pi}) = \mathrm{rank}(\mathbf{\Phi}(1))$.

CI Vectors

- If $\mathrm{rank}(\mathbf{\Pi}) = h$, there exist two $n \times h$ matrices $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$, each of of rank h, such that
$$\mathbf{\Pi} = \tilde{\boldsymbol{\alpha}}\boldsymbol{\beta}' \tag{27}$$
Hence, $\tilde{\boldsymbol{\alpha}}\boldsymbol{\beta}'\mathbf{\Psi}(1) = \mathbf{0} \Rightarrow \boldsymbol{\beta}'\mathbf{\Psi}(1) = \mathbf{0}$, which shows that the rows of $\boldsymbol{\beta}'$ are cointegrating vectors.
- The matrices $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ are not uniquely defined, since $\tilde{\boldsymbol{\alpha}}\boldsymbol{\beta}' = \tilde{\boldsymbol{\alpha}}\mathbf{H}\mathbf{H}^{-1}\boldsymbol{\beta}'$ for any non-singular matrix $\mathbf{H}$ (of dimension $h \times h$).
- For estimation, h^2 identification restrictions need to be imposed. For example, if $h = 1$, $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1, \ \boldsymbol{\beta}_2)$ must be normalized to e.g. $(1, \ -b)$ where $b = -\boldsymbol{\beta}_2/\boldsymbol{\beta}_1$.

VECM Representation

- We have

$$\Delta \mathbf{y}_t = \tilde{\alpha}\boldsymbol{\beta}' y_{t-1} + \boldsymbol{\alpha} + \boldsymbol{\gamma} t + \boldsymbol{\zeta}_1 \Delta \mathbf{y}_{t-1} + \ldots + \boldsymbol{\zeta}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{v}_t \quad (28)$$

  which is the vector error-correction model (VECM).
- The variables $\boldsymbol{\beta}' y_{t-1}$ are the "cointegrating errors" (or "disequilibrium terms") which are corrected for in each equation of the system through the "loading coefficients" in the matrix $\tilde{\alpha}$.
- If $\boldsymbol{\beta}' y_t$ has no trend, then $\boldsymbol{\beta}'\mathbf{d} = 0$ and $\gamma = -\tilde{\alpha}\boldsymbol{\beta}'\mathbf{d} = 0$. In this case, the VECM does not include the linear trend term although it may be present in some elements of y_t as we see in equation (23).

A Cointegraed VAR(1)

- Consider the simple VAR(1) process:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{v}_t, \tag{29}$$

with $\mathbf{A}(z) = \mathbf{I} - \mathbf{A}z$ has at least one root at $z = 1$ if $|\mathbf{A}(1)| = 0$. Equivalently, the corresponding $\mathbf{\Pi}$ has reduced rank where $\mathbf{\Pi} = \mathbf{A} - \mathbf{I}$.

- We assume that $\mathbf{y}_t \sim \mathrm{I}(1)$. Very important!
- The corresponding VECM is

$$\Delta\mathbf{y}_t = \mathbf{\Pi}\mathbf{y}_{t-1} + \mathbf{v}_t \tag{30}$$

with $\mathbf{\Pi} = \tilde{\boldsymbol{\alpha}}\boldsymbol{\beta}'$, h is the CI rank.

- Since $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$ are both $n \times h$ full rank matrices, there exist $\tilde{\boldsymbol{\alpha}}_\perp$ and $\boldsymbol{\beta}_\perp$, which are both $n \times (n-h)$ matrices, s.t. $\tilde{\boldsymbol{\alpha}}'_\perp\tilde{\boldsymbol{\alpha}} = 0$ and $\boldsymbol{\beta}'_\perp\boldsymbol{\beta} = 0$.

Roots and Eigenvalues

- If (29) is stable (the lag polynomial $\mathbf{A}(L)$ is stable), the roots of $|\mathbf{A}(z)| = |\mathbf{I} - \mathbf{A}z| = 0$ are all outside the unit circle, i.e. $|z| > 1$.
- This is equivalent to the eigenvalue problem $|\lambda\mathbf{I} - \mathbf{A}| = 0$, which implies that the modulus of all the eigenvalues of $\mathbf{A}$ matrix are all smaller than 1, i.e. $|\lambda| < 1$.
- If the system contains unit roots, we say, some of the roots or eigenvalues are equal to one.

- Multiplying both sides of (30) by β' yields

$$\begin{aligned}\beta'(\mathbf{y}_t - \mathbf{y}_{t-1}) &= \beta'\tilde{\alpha}\beta'\mathbf{y}_{t-1} + \beta'\mathbf{v}_t \\ \beta'\mathbf{y}_t &= (\beta'\tilde{\alpha} + \mathbf{I})\beta'\mathbf{y}_{t-1} + \beta'\mathbf{v}_t \\ \rightarrow \mathbf{s}_t &= \mathbf{B}\mathbf{s}_{t-1} + \boldsymbol{\eta}_t = \mathbf{B}^t\mathbf{s}_0 + \sum_{i=0}^{t-1}\mathbf{B}^i\boldsymbol{\eta}_{t-i} \qquad (31) \\ &= \sum_{i=0}^{\infty}\mathbf{B}^i\boldsymbol{\eta}_{t-i}, \qquad (32)\end{aligned}$$

where $\mathbf{s}_t = \beta'\mathbf{y}_t \sim \mathrm{I}(0)$, $\mathbf{B} = \beta'\tilde{\alpha} + \mathbf{I}$ and $\boldsymbol{\eta}_t = \beta'\mathbf{v}_t \sim \mathrm{I}(0)$.
- This process contains the linear combinations of $\mathbf{y}_t$, which are stationary or asymptotically stable process over time.
- β consists of h linearly independent vectors, and it is called long-run relations or cointegrating relations if $\beta'\mathbf{y}_t \sim \mathrm{I}(0)$.
- These linear combinations are not unique. For any $K \neq 0$, $K\beta'\mathbf{y}_t$ is also stationary.

Why $\beta' y_t$ is asymptotically stable?

- Due to the important assumption: $|\mathbf{A}(z) = 0|$ has $n - h$ unit roots and the other roots are outside the unit circle.
- $\mathbf{\Pi} = \mathbf{A} - \mathbf{I} = -\mathbf{A}(1)$ has reduced rank and can be decomposed by $\tilde{\alpha}\beta'$. CI rank is h. And $\mathbf{A} = \mathbf{I} + \tilde{\alpha}\beta'$.
- Check the following derivation carefully

$$
\begin{aligned}
|\mathbf{A}(z)| = 0 \quad & \Longrightarrow \quad |(\boldsymbol{\beta}, \boldsymbol{\beta}_\perp)'\mathbf{A}(z)(\boldsymbol{\beta}, \boldsymbol{\beta}_\perp)| = 0 \\
& \Longrightarrow \quad \begin{vmatrix} \boldsymbol{\beta}'\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta} z & -\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta}_\perp z \\ -\boldsymbol{\beta}_\perp'\mathbf{A}\boldsymbol{\beta} z & \boldsymbol{\beta}_\perp'\boldsymbol{\beta}_\perp - \boldsymbol{\beta}_\perp'\mathbf{A}\boldsymbol{\beta}_\perp z \end{vmatrix} = 0 \quad (33) \\
& \Longrightarrow \quad |\mathbf{I}_h - (\mathbf{I}_h + \boldsymbol{\beta}'\tilde{\alpha})z||\mathbf{I}_{n-h} - \mathbf{I}_{n-h}z| = 0 \qquad (34)
\end{aligned}
$$

  where $\mathbf{I}_h + \boldsymbol{\beta}'\tilde{\alpha} = \mathbf{B}$
- The other roots ($|\mathbf{I}_h - (\mathbf{I}_h + \boldsymbol{\beta}'\tilde{\alpha})z| = 0$) are outside the unit circle as assumed...

The Pushing Force

- Multiplying both sides of (30) by $\tilde{\alpha}'_\perp$ yields

$$\begin{aligned} \tilde{\alpha}'_\perp \Delta \mathbf{y}_t &= \tilde{\alpha}'_\perp \mathbf{v}_t \\ \tilde{\alpha}'_\perp \mathbf{y}_t &= \tilde{\alpha}'_\perp \sum_{i=0}^{t-1} \mathbf{v}_{t-i} + \tilde{\alpha}'_\perp \mathbf{y}_0. \end{aligned} \tag{35}$$

- $\tilde{\alpha}'_\perp \sum_{i=0}^{t-1} \mathbf{v}_{t-i}$ in (35) is the common stochastic trends of the I(1) VAR(1) process. We see that there are $n-h$ common stochastic trends, or unit roots in the vector system.
- The common stochastic trends are not unique. For any full rank $(n-h) \times (n-h)$ matrix K, $K\tilde{\alpha}'_\perp \sum_{i=0}^{t-1} \mathbf{v}_{t-i}$ common trends as well.
- $\tilde{\alpha}'_\perp \sum_{i=0}^{t-1} \mathbf{v}_{t-i}$ is also called the pushing force.

- The beautiful identity:

$$\beta_\perp(\tilde{\alpha}'_\perp\beta_\perp)^{-1}\tilde{\alpha}'_\perp + \tilde{\alpha}(\beta'\tilde{\alpha})^{-1}\beta' = \mathbf{I}. \tag{36}$$

Thus, we have

$$\begin{aligned}\mathbf{y}_t &= (\beta_\perp(\tilde{\alpha}'_\perp\beta_\perp)^{-1}\tilde{\alpha}'_\perp + \tilde{\alpha}(\beta'\tilde{\alpha})^{-1}\beta')\mathbf{y}_t \\ &= (\beta_\perp(\tilde{\alpha}'_\perp\beta_\perp)^{-1})\tilde{\alpha}'_\perp\mathbf{y}_t + (\tilde{\alpha}(\beta'\tilde{\alpha})^{-1})\beta'\mathbf{y}_t\end{aligned}$$

Replace the red parts by the common trends and the long-run relations:

$$y_t = \mathbf{C}\sum_{i=0}^{t-1}\mathbf{v}_{t-i} + \mathbf{C}\mathbf{y}_0 + \tilde{\alpha}(\beta'\tilde{\alpha})^{-1}(\sum_{i=0}^{\infty}\mathbf{B}^i\eta_{t-i},) \tag{37}$$

where $\mathbf{C} = \beta_\perp(\tilde{\alpha}'_\perp\beta_\perp)^{-1}\tilde{\alpha}'_\perp$.

Cointegrated VAR with Intercept and Trend

- We consider the following VAR(1) model

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\delta} t + \mathbf{v}_t, \tag{38}$$

  where $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$ may not be zero. $\mathbf{A}(z)$ contains r unit roots and the others roots are outside the unit circle.
- The corresponding VECM

$$\Delta\mathbf{y}_t = \mathbf{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\delta} t + \mathbf{v}_t, \tag{39}$$

  where $\mathbf{\Pi} = \tilde{\alpha}\beta'$.
- $\beta'\mathbf{y}_t$ is trend stationary. But the "one-summability" should be carefully checked.
- The pushing force may contains quadratic trend:

$$\begin{aligned} \tilde{\alpha}_\perp' \Delta\mathbf{y}_t &= \tilde{\alpha}_\perp' \boldsymbol{\mu} + \tilde{\alpha}_\perp' \boldsymbol{\delta} t + \tilde{\alpha}_\perp' \mathbf{v}_t \\ \tilde{\alpha}_\perp' \mathbf{y}_t &= \tilde{\alpha}_\perp' \sum_{i=0}^{t-1} (\boldsymbol{\mu} + \boldsymbol{\delta}(t-i) + \mathbf{v}_{t-i}) + \tilde{\alpha}_\perp' \mathbf{y}_0. \end{aligned} \tag{40}$$

The role of deterministic terms

- The Granger VMA representation is

$$\mathbf{y}_t = \mathbf{C}\mathbf{y}_0 + \mathbf{C}\sum_{i=0}^{t-1}(\boldsymbol{\mu} + \boldsymbol{\delta}(t-i) + \mathbf{v}_{t-i}) \quad (41)$$

$$+\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}'\tilde{\boldsymbol{\alpha}})^{-1}(\sum_{i=0}^{\infty}\mathbf{B}^i\boldsymbol{\beta}'(\boldsymbol{\mu} + \boldsymbol{\delta}(t-i) + \mathbf{v}_{t-i})) \quad (42)$$

  where $\mathbf{C} = \boldsymbol{\beta}_\perp(\tilde{\boldsymbol{\alpha}}'_\perp\boldsymbol{\beta}_\perp)^{-1}\tilde{\boldsymbol{\alpha}}'_\perp$ and the last term is trend stationary.
- If $\boldsymbol{\delta} = \tilde{\boldsymbol{\alpha}}\boldsymbol{\kappa}$, where $\boldsymbol{\kappa}$ is an $h \times h$ matrix, there will be no quadratic trend in the system!
- Given $\boldsymbol{\delta} = 0$, if $\boldsymbol{\mu} = \tilde{\boldsymbol{\alpha}}\gamma$, where γ is an $h \times h$ matrix, there will be no deterministic trend in the system!

- Given CI h, the deterministic terms can be written as $\mathbf{d}_t = \boldsymbol{\mu} + \boldsymbol{\delta} t = \tilde{\boldsymbol{\alpha}}\kappa_0 + \tilde{\boldsymbol{\alpha}}_\perp \kappa_1 + (\tilde{\boldsymbol{\alpha}}\gamma_0 + \tilde{\boldsymbol{\alpha}}_\perp \gamma_1)t$
- the following models (hypotheses) have nested relations:

$$
\begin{aligned}
H(h) &: \quad \mathbf{d}_t = \tilde{\boldsymbol{\alpha}}\kappa_0 + \tilde{\boldsymbol{\alpha}}_\perp \kappa_1 + (\tilde{\boldsymbol{\alpha}}\gamma_0 + \tilde{\boldsymbol{\alpha}}_\perp \gamma_1)t && (43)\\
H^*(h) &: \quad \mathbf{d}_t = \tilde{\boldsymbol{\alpha}}\kappa_0 + \tilde{\boldsymbol{\alpha}}_\perp \kappa_1 + \tilde{\boldsymbol{\alpha}}\gamma_0 t \text{ (no quadratic trend)} && (44)\\
H_1(h) &: \quad \mathbf{d}_t = \tilde{\boldsymbol{\alpha}}\kappa_0 + \tilde{\boldsymbol{\alpha}}_\perp \kappa_1 \text{ (no trend in } \boldsymbol{\beta}' y_t) && (45)\\
H_1^*(h) &: \quad \mathbf{d}_t = \tilde{\boldsymbol{\alpha}}\kappa_0 \text{ (no trend)} && (46)\\
H_2(h) &: \quad \mathbf{d}_t = 0 \text{ (no deterministic terms)} && (47)
\end{aligned}
$$

To be continued! Thank you!