Time Series Econometrics, 2ST111

Lecture 11. Cointegration

Yukai Yang

Department of Statistics, Uppsala University

Outline of Today's Lecture

- Engle-Granger's Procedure for the Cointegration Analysis
- Johansen's Procedure for the Cointegration Analysis
- The Parameter Estimation and Reduced Rank Regression
- The Trace Test

Engle-Granger's Procedure

- Suppose that x_t and $y_t \sim \mathrm{I}(1)$. This can be checked by for example the augmented DF test.
- If x_t and y_t can be cointegrated, then there exits a linear combination $\mathbf{a}$ of the two variables, which can be normalized as $\mathbf{a} = (1, -a)'$, such that $\mathbf{a}'(x_t, y_t)' = x_t - a y_t \sim \mathrm{I}(0)$.
- Therefore, if we regress the model as follows

$$x_t = c + b y_t + u_t,$$

then $\hat{b} \xrightarrow{p} a$ in a super consistent way, and in particular the residuals should satisfy $\hat{u}_t \sim \mathrm{I}(0)$, as u_t is $\mathrm{I}(0)$.

- We test whether $\hat{u}_t$ contains unit root, or $\hat{u}_t \sim \mathrm{I}(1)$, and then make the conclusion whether x_t and y_t can be cointegrated.

Engle-Granger's Procedure

Remarks:

- When x_t and y_t are not cointegrated, $\hat{u}_t$ is I(1) almost surely. And $\hat{b}$ will not converge to zero, though $b = 0$ (spurious regression).
- E-G procedure requires that the variables x_t, y_t and perhaps more, should be all I(1). This makes sense because if $x_t \sim$ I(0) and $y_t \sim$ I(1), then there is naturally a linear combination $\mathbf{a} = (1, 0)$ such that $\mathbf{a}'(x_t, y_t)' \sim$ I(0), which is not cointegration.
- However, E-G procedure suffers from the problem that, if there are more than two variables x_t, y_t and z_t, there may be more than one linear combination, say, two linearly independent linear combinations, that makes the cointegration, but E-G can only find one of them.

Engle-Granger's Procedure

Remarks:

- Consider

$$x_t = \rho_1 u_t + \epsilon_{1t}; \quad y_t = \rho_2 u_t + \epsilon_{2t}; \quad z_t = \rho_3 u_t + \epsilon_{3t},$$

  where $u_t \sim \mathrm{I}(1)$, $\epsilon_{it} \sim \mathrm{I}(0)$, and ρ_i are non-zero real numbers, $i = 1, 2, 3$.
- The three random variables share the same I(1) process u_t. We can say that the vector system, or the system (x_t, y_t, z_t), is pushed by the same underlying common stochastic trend. This gives normally nice economic insight.
- $\mathbf{a}_1 = (1, -\rho_1/\rho_2, 0)'$ is one cointegrating vector. $\mathbf{a}_2 = (1, 0, -\rho_1/\rho_3)'$ is another one. So is any non-zero linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$.
- E-G procedure just finds the cointegrating vector which produces the smallest residual-sum-squares for the regression model

$$x_t = c + b_1 y_t + b_2 z_t + u_t$$

$y_t \sim I(1)$ $\quad \alpha_1' y_t \sim I(0)$ $\quad \alpha_1 \alpha_1' + \alpha_2 \alpha_2'$

$\alpha_2' y_t \sim I(0)$

The Characteristic Polynomial

Before introducing the Johansen's procedure, we have to introduce

- The n-dimensional VAR(p)

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + \Phi D_t + \varepsilon_t \tag{1}$$

where D_t is vector containing the deterministic terms and any possible exogenous variables, ε_t is i.i.d. $\sim N_n(0, \Omega)$, t= 1, ..., T with initial variables $y_{-p+1}, ..., y_0$.

- It can be written as follows

$$A(L)y_t = \Phi D_t + \varepsilon_t \tag{2}$$

where $A(L) = I_n - A_1 L - A_2 L^2 - ... - A_p L^p$ is the lag polynomial. $A(z)$ is termed the characteristic polynomial of the dynamic system.

The Basic Assumption

- Johansen's cointegration procedure presumes that the data vector y_t is no more than I(1). More specifically, if $y_t = (y_{1t}, ..., y_{nt})'$, then either $y_{it} \sim \text{I}(0)$ or $\Delta y_{it} \sim \text{I}(0)$ holds for $i = 1, ..., n$.
- To ensure this, Johansen's procedure employs the following assumption explicitly

Assumption 1 in Johansen (1995)

The characteristic polynomial satisfies the condition that if $|A(z)| = 0$, then either $|z| > 1$ or $z = 1$.

- $|z| < 1$ is termed the explosive root. $|z| = 1$ is not necessary a unit root. Instead, we call $|z| = 1$ but $z \neq 1$ a seasonal root which is ruled out.

The VECM Form

The corresponding VECM is:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \Phi D_t + \varepsilon_t \tag{3}$$

where $\Pi = \sum_{i=1}^{p} A_i - I_n$ and $\Gamma_i = -\sum_{j=i+1}^{p} A_j$.

Remarks:

- If $y_t \sim \mathrm{I}(1)$ and can be cointegrated, Π has reduced rank $r < n$ and can be decomposed as $\Pi = \alpha\beta'$. Note that the CI rank is r now!
- In estimation, we assume that we know r, though you can try to estimate using every possible r.
- I'd like to stress again, we presume that the system is at most I(1).

Remarks

- $r = 0$ implies that $\Pi = 0$, and hence $r = 0$, I(1) but no cointegration! The number of stochastic trends are $n - r = n$, and hence there are n unit roots in the system.
- $r = n$ implies that Π is full rank, and hence the num her of stochastic trends are $n - r = 0$, NO unit root! Thus, I(0) asymptotically stable! (It is not I(1) but at most I(1), which means it can only be I(0)! If you check the VECM, you will get the same conclusion.)
- A cointegrated VAR and its VECM should have a Π whose rank is an integer satisfying $1 \leq r < n$.
- If the rank r is testable (yes), based on the Johansen's procedure, we can test $r = 0$ (no cointegration), $r = n$ (I(0), a more general DF test), and $r = r_0$ (the number of the linearly independent long-run relations), compared to E-G procedure.

The nonlinear regression model

- Denoting $Z_{0t} = \Delta y_t$, $Z_{1t} = y_{t-1}$ and $Z_{2t} = (\Delta y'_{t-1}, \Delta y'_{t-2}, ..., \Delta y'_{t-p+1}, D'_t)'$, (3) can be written as:

$$Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \varepsilon_t, \tag{4}$$

  where $\Psi = (\Gamma_1, \Gamma_2, ..., \Gamma_{p-1}, \Phi)$.

  nuisance parameter

- (4) is the regression model we will consider in the estimation.
- It is nonlinear regression due to the reduced rank property of Π.
- α and Ψ are called "short-run adjustments".
- People are more interested in the long-run parameter β, the short-run adjustment ΨZ_{2t} is more or less "nuisance parameter".
- ΨZ_{2t} can be removed, so that the it looks like a simple VAR(1) model without deterministic terms.

The log-likelihood function

- The log-likelihood function is given by

$$\log L(\alpha, \beta, \Psi, \Omega) = -\frac{1}{2} T \log |\Omega| - \frac{1}{2} \sum_{t=1}^{T} (Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t})' \Omega^{-1} (Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t}) \quad (5)$$

- The Maximum Likelihood estimators of the parameters $(\alpha, \beta, \Psi, \Omega)$ are obtained by maximizing this function.
- When ε_t is not Gaussian distributed, it is called "Quasi-Maximum Likelihood".
- In addition, I introduce the notation for the product moment matrices: $M_{ij} = T^{-1} \sum_{t=1}^{T} Z_{it} Z_{jt}'$, $i, j = 0, 1, 2$. Note that $M_{ij} = M_{ji}'$.

Estimating Ψ given α and β

- The first order condition (FOC) is for estimating Ψ are given by

$$\sum_{t=1}^{T}(Z_{0t} - \alpha\beta' Z_{1t} - \hat{\Psi} Z_{2t})Z_{2t}' = 0 \tag{6}$$

- It can be rewritten as

$$M_{02} = \alpha\beta' M_{12} + \hat{\Psi} M_{22}$$

such that

$$\hat{\Psi}(\alpha, \beta) = M_{02} M_{22}^{-1} - \alpha\beta' M_{12} M_{22}^{-1} \tag{7}$$

- Note that Z_{it} are from the data set, and hence M_{ij} are always known.

A tricky way to get rid of Ψ

- Replace the Ψ by $\hat{\Psi}(\alpha, \beta)$ in (5), and get concentrated LL function

$$\log L(\alpha, \beta, \Omega) = \frac{1}{2} \sum_{t=1}^{T} (Z_{0t} - \alpha\beta' Z_{1t} - (M_{02} M_{22}^{-1} - \alpha\beta' M_{12} M_{22}^{-1}) Z_{2t})'$$
$$\Omega^{-1} (Z_{0t} - \alpha\beta' Z_{1t} - (M_{02} M_{22}^{-1} - \alpha\beta' M_{12} M_{22}^{-1}) Z_{2t}) + \ldots$$
$$= \frac{1}{2} \sum_{t=1}^{T} ((Z_{0t} - M_{02} M_{22}^{-1} Z_{2t}) - \alpha\beta' (Z_{1t} - M_{12} M_{22}^{-1} Z_{2t}))'$$
$$\Omega^{-1} ((Z_{0t} - M_{02} M_{22}^{-1} Z_{2t}) - \alpha\beta' (Z_{1t} - M_{12} M_{22}^{-1} Z_{2t})) + \ldots$$

- This implies two simple linear regressions! The residuals of the two regressions are given by

$$R_{0t} = Z_{0t} - M_{02} M_{22}^{-1} Z_{2t} \quad (8)$$
$$R_{1t} = Z_{1t} - M_{12} M_{22}^{-1} Z_{2t} \quad (9)$$

A tricky way to get rid of Ψ (cont.)

- The two regressions are just: 1. regress Z_{0t} on Z_{2t}; 2. regress Z_{1t} on Z_{2t}; and then collect the residuals R_{0t} and R_{1t}.
- The concentrated log-likelihood function becomes

$$\log L(\alpha, \beta, \Omega) = \frac{1}{2} \sum_{t=1}^{T} (R_{0t} - \alpha\beta' R_{1t})' \Omega^{-1} (R_{0t} - \alpha\beta' R_{1t}) + ...$$

- This concentrated log-likelihood function implies a Gaussian model (a new regression)

$$R_{0t} = \alpha\beta' R_{1t} + \tilde{\varepsilon}_t, \tag{10}$$

  with $\tilde{\varepsilon}_t \sim N(0, \Omega)$. This implies that Π can be estimated by regressing R_{0t} on R_{1t} if it is full rank.
- The key point is that we can forget about Ψ by using R_{0t} and R_{1t}.
- Recall that Z_{0t} contains Δy_t, Z_{1t} contains y_{t-1} and Z_{2t} the others.
- If no Z_{2t}, $R_{0t} = Z_{0t}$, $R_{1t} = Z_{1t}$! Simple VAR(1) without anything else.

Estimating α and Ω given β

- Due to the reduced rank restriction, (10) can not be estimated directly! We resort to the concentrated likelihood again.
- Denote $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}' = M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $i = 0, 1$. (This notation will be used very often!)
- Supposing that β is known, α and Ω can be estimated by simple linear regression:

$$\begin{aligned} \hat{\alpha}(\beta) &= S_{01}\beta(\beta' S_{11}\beta)^{-1}, && (11) \\ \hat{\Omega}(\beta) &= S_{00} - \hat{\alpha}(\beta)(\beta' S_{11}\beta)^{-1}\hat{\alpha}(\beta)' \\ &= S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}. && (12) \end{aligned}$$

Reduced Rank Regression

- Actually this is the method suggested by Anderson (1951), which is termed Reduced Rank Regression (RRR). It solves the ML problem when there is a reduced rank parameter matrix in the vector system.
- We see that,

$$\arg\max_{\beta} \log L(\beta) = \arg\min_{\beta} \log L(\beta)^{-\frac{2}{\pi}}, \tag{13}$$

and apart from the constant term in the log-likelihood function,

$$\log L(\beta)^{-\frac{2}{\pi}} = |\hat{\Omega}(\beta)| = |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}|. \tag{14}$$

Reduced Rank Regression

- We apply the following identity

$$\begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} = |\Sigma_{11}||\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}| = |\Sigma_{22}||\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}|. \tag{15}$$

Thus,

$$\begin{aligned} &|S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}| = \\ &\quad |\beta' S_{11}\beta|^{-1}|S_{00}||\beta' S_{11}\beta - \beta' S_{10}(S_{00})^{-1}S_{01}\beta| \\ &\quad \propto |\beta' S_{11}\beta|^{-1}|\beta'[S_{11} - S_{10}(S_{00})^{-1}S_{01}]\beta| \end{aligned} \tag{16}$$

- Minimizing this ratio implies an eigenvalue problem!

$|\lambda I - A| = 0$, $|B\lambda - A| = 0$

Reduced Rand Regression

- (16) is minimized by solving the eigenvalue problem:

$$|\rho S_{11} - [S_{11} - S_{10}(S_{00})^{-1}S_{01}]| = 0, \tag{17}$$

and choose $\hat{\beta} = (v_1, v_2, ..., v_r)$ the eigenvectors corresponding to the smallest r eigenvalues.

- Or equivalently, for $\lambda = 1 - \rho$, by solving the eigenvalue problem:

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0 \tag{18}$$

and choose $\hat{\beta} = (v_1, v_2, ..., v_r)$ the eigenvectors corresponding to the biggest r eigenvalues.

- Therefore, we have

$$\begin{aligned} \max \log L^{-\frac{2}{\pi}}(r) &= |S_{00}||\hat{\beta}' S_{11}\hat{\beta}|^{-1}|\hat{\beta}'[S_{11} - S_{10}(S_{00})^{-1}S_{01}]\hat{\beta}| \\ &= |S_{00}|\prod_{i=1}^{r}\hat{\rho}_i = |S_{00}|\prod_{i=1}^{r}(1-\hat{\lambda}_i) \end{aligned} \tag{19}$$

The ML estimator and the LR test

- There is an identification problem for $\hat{\beta}$, since the value of the likelihood will not be changed for any $\tilde{\beta} = \hat{\beta}\kappa$, for any $r \times r$ full rank κ. Eigenvalue decomposition is not unique in the same manner. We can just use the eigenvectors the software gives, or choose some normalization scheme.
- Immediately, we have the Likelihood ratio test for $H(r) : \mathrm{rank}(\Pi) = r$ against $H(n) : \mathrm{rank}(\Pi) = n$ based on (19)

$$Q(H(r)|H(n))^{-\frac{2}{T}} = \frac{|S_{00}| \prod_{i=1}^{r}(1-\hat{\lambda}_i)}{|S_{00}| \prod_{i=1}^{M}(1-\hat{\lambda}_i)}. \tag{20}$$

This is the well-known Johansen's "trace" test statistic (it is an LR test):

$$-2 \log Q(H(r)|H(n)) = -T \sum_{i=r+1}^{M} (1-\hat{\lambda}_i). \tag{21}$$

The Johansen's trace test

- One may ask why (24) is called "trace". The reason is that, as $T \to \infty$, it converges in distribution to a trace (DF test with $M - r$ degrees of freedom):

$$-2 \log Q(H(r)|H(n)) \xrightarrow{d} \operatorname{tr}\left\{ \int_0^1 (\mathrm{d}W) F' \left(\int_0^1 FF' \mathrm{d}u \right)^{-1} \int_0^1 F (\mathrm{d}W)' \right\} \tag{22}$$

  where W stands for the standard multivariate Wiener process with dimension $n - r$. F is a random function whose dimension is depends on the deterministic term.
- The distribution of this trace is not necessarily standard, and it depends greatly on the functional form of F.
- The functional form of F is determined by the deterministic term.
- Recall the possible forms of the deterministic term in last section. There are five possible forms of F.
- Critical values are generated by means of simulation.

The Johansen's trace test (cont.)

- In particular, when $H_2(r) : d_t = 0$, $F = W$. A simple DF-type test, easy to simulate.
- The LR test does not depend on the choice of Γ_j, i.e. the lag length, which means there is no augmented DF!
- The test procedure for determining the rank is applied in a sequential way.
- Estimate using the RRR method only once, and find the λ_i, $i = 1, ..., n$. Sort λ_i in decreasing order!
- $\Pi(r_1)$ is a special case of $\Pi(r_0)$ if $r_1 < r_0$, and of course is nested by the latter one. The test procedure is conducted by the sequence: test $r = n - 1$ against $r = n$, if accepted then test $r = n - 2$ against $r = n$ ... until rejection.
- This procedure can also be understood as: test $r \leq n - 1$ against $r > n - 1$, test $r \leq n - 2$ against $r > n - 2$ ... until rejection.

The Johansen's trace test (cont.)

- Different from the Granger's procedure, Johansen's procedure can find more than one cointegration relation.
- It is flexible in the sense that some variables can be I(0) and the limit distribution is independent of the choice of the lag.
- Univariate unit root test finds one root, whereas the Johansen's test find how many common stochastic trends (several roots).
- It suffers from the size distortion problem in finite sample cases when the dimension grow up. The possible solutions are: 1. apply the possible Bartlett corrections; or 2. use the bootstrapping method.
- For details, see theorem 11.1 in Johansen (1995).

Hypothesis testing for long-run parameters

- The null hypothesis $H_0 : \beta = H\varphi$ is a very tricky way to represent the equivalent hypothesis $R'\beta = 0$, since $R'H = 0$ and they are actually orthogonal to each other.
- φ contains the freely varying parameters in β.
- We impose $n - s$ restrictions on the space spanned by β, $\implies$ R is $n \times (n - s)$, and hence, $\implies$ H is $n \times s$. H and R span the null space to each other.
- By replacing $\beta = H\varphi$, the model can be estimated by means of RRR and the null hypothesis can be tested.
- The new eigenvalue problem:

$$|\lambda^* H' S_{11} H - H' S_{10} S_{00}^{-1} S_{01} H| = 0 \tag{23}$$

- The LR test statistic for $\beta = H\varphi$:

$$-2 \log Q(H_0|H(r)) = T \sum_{i=1}^{r} \log\{(1 - \lambda_i^*)/(1 - \hat{\lambda}_i)\}. \tag{24}$$

which is asymptotically χ^2 distributed with degrees of freedom

Thank you for taking the course!