6 Cointegration

- **1 Introduction**

Consider the following bivariate system

$$\begin{aligned} y_{1t} &= \gamma y_{2t} + u_{1t} \\ y_{2t} &= y_{2,t-1} + u_{2t} \end{aligned}$$

with u_{1t} and u_{2t} uncorrelated white noice processes. Both y_{1t} and y_{2t} are I(1) processes:

$$\begin{aligned} \Delta y_{1t} &= \gamma u_{2t} + u_{1t} - u_{1,t-1} \\ \Delta y_{2t} &= u_{2t}, \end{aligned}$$

but the linear combination $(y_{1t} - \gamma y_{2t})$ is stationary. We say $y_t \equiv (y_{1t}, y_{2t})'$ is cointegrated with a vector $(1, -\gamma)'$. See Figure 1 in Ch6-ex1: $y_{1t} = \gamma y_{2t} + u_{1t},\ y_{2t} = y_{2,t-1} + u_{2t}$, with u_{1t} and u_{2t} independent N(0,1) variables, $\gamma = 1, 2, -1$.

Note: 1) If the vector y_t is cointegrated, it is not correct to fit a VAR to the differenced data. Denote $\varepsilon_{1t} = \gamma u_{2t} + u_{1t}$, $\varepsilon_{2t} = u_{2t}$. Then

$$\begin{aligned} \Delta y_t &\equiv \begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} - \varepsilon_{1,t-1} + \gamma \varepsilon_{2,t-1} \\ \varepsilon_{2t} \end{pmatrix} \\ &= \begin{pmatrix} 1-L & \gamma L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \equiv \Psi(L)\varepsilon_t. \end{aligned}$$

The matrix moving average operator has a unit root and is noninvertible, hence no finite-order VAR could describe Δy_t.

2) Error-correction:

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

or

$$\begin{aligned} \Delta y_t &= \begin{pmatrix} -1 & \gamma \\ 0 & 0 \end{pmatrix} y_{t-1} + u_t \\ &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} (y_{1,t-1} - \gamma y_{2,t-1}) + u_t \\ &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} (1, -\gamma)\, y_{t-1} + u_t \equiv \alpha\beta' y_{t-1} + u_t. \end{aligned}$$

Some examples:

1) Consumptiuon and income are I(1) processes, but over a long run consumptiuon c_t tend to be a roughly constant proportion of income y_t, so that $(\log c_t - \log y_t)$ appears to be stationary. That is $(\log c_t, \log y_t)'$ is cointegrated with $(1, -1)'$.

2) PPP: $P_t = S_t P_t^*$ or taking log, $p_t = s_t + p_t^*$. Each of the three variables p_t, s_t, and p_t^* is $I(1)$. A weak version of the hypothesis is that the variable $z_t \equiv (p_t - s_t - p_t^*)$ is stationary, i.e. $(p_t, s_t, p_t^*)'$ is cointegrated with $(1, -1, -1)'$.

3) Money demand: Money demand is proportional to the price level; as income increases, indviduals will want to hold increased money balances; money demand is negatively related to the interest rate. Hence in equilibrium: money demand = money supply,

$$m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t$$

where m_t is the the money supply, e_t is stationary. Here $\beta_1 = 1$, $\beta_2 > 0$ and $\beta_3 < 0$ by the behavioral assumptions. When all the variables are I(1), $(m_t, p_t, y_t, r_t)'$ is cointegrated with $(1, -1, -\beta_2, -\beta_3)'$. Also, suppose that the monetary authorities followed a feedback rule such that they decreased the money supply when nominal GDP was high and increased the money supply when nominal GDP was low. Then

$$\begin{aligned} m_t &= \gamma_0 - \gamma_1 (p_t + y_t) + e_{1t} \\ &= \gamma_0 - \gamma_1 p_t - \gamma_1 y_t + 0 \cdot r_t + e_{1t} \end{aligned}$$

where e_{1t} is stationary. Then $(m_t, p_t, y_t, r_t)'$ is also cointegrated with $(1, \gamma_1, \gamma_1, 0)'$.

- **2 Cointegration**

Long-run equilibrium: $\beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt} = 0$ or $\beta' x_t = 0$, where $\beta = (\beta_1, \beta_2, \cdots, \beta_n)'$, $x_t = (x_{1t}, x_{2t}, \cdots, x_{nt})'$.

Equilibrium error——the deviation from the long-run equilibrium: $\beta' x_t = e_t$, where $\{e_t\}$ is stationary.

Cointegration: If $z_t \sim I(d)$ and $y_t \sim I(d)$, it is generally true that $z_t - a y_t \sim I(d)$. Further, when $z_t - a y_t \sim I(b)$ $(b < d)$, we say that z_t and y_t are cointegrated. More formally, the components of vector x_t are said to be **Cointegrated of order** d, b, denoted $x_t \sim CI(d, b)$, if

1) all components of x_t are $I(d)$;

2) $\exists$ a vector $\beta = (\beta_1, \beta_2, \cdots, \beta_n)' \neq 0$ such that the **linear** combination $\beta' x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt} \sim I(d - b)$ where $b > 0$.

The vector β is called **the cointegrating vector,** which represents the long-run equilibrium relationship among variables.

Remarks: (1) Cointegration refers to a **linear combination** of nonstationary variables. (2) The cointegrating vector is not unique: the set of cointegrating vectors constitutes a vector subspace satisfying $\beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_n x_{nt} \sim I(d-b)$; that is,

$$\{\beta = (\beta_1, \beta_2, \cdots, \beta_n)' \in R^n \backslash \{0\} : \beta' x_t \sim I(d-b)\}$$

When β is a cointegrating vector, $\lambda\beta$ is also a cointegrating vector for all $\lambda \neq 0$. **A normalized integrating vector** is $\beta/\beta_1 = (1, \beta_2/\beta_1, \cdots, \beta_n/\beta_1)'$ if $\beta_1 \neq 0$. There may be at most $n-1$ linearly independent cointegrating vectors. The number of linearly independent cointegrating vectors is called **the cointegrating rank of** x_t. (3) Convention: Here assume that $x_t \sim CI(1,1)$ s.t. $\beta' x_t \sim I(0)$.

- **Cointegration and Trend:** The parameters of the cointegration vector purge the trend from the linear combination of the cointegrated variables while any other linear combination up to normalization can not achieve this. Three examples:

1) Suppose $\varepsilon_{yt}, \varepsilon_{zt}, \varepsilon_t$ are i.i.d. white noise processes, and

$$\begin{aligned} y_t &= \mu_t + \varepsilon_{yt}, \quad z_t = \mu_t + \varepsilon_{zt} \\ \mu_t &= \mu_{t-1} + \varepsilon_t. \end{aligned}$$

Then $y_t - z_t = \varepsilon_{yt} - \varepsilon_{zt} \sim I(0)$, i.e. $(1, -1)\begin{pmatrix} y_t \\ z_t \end{pmatrix} \sim I(0)$. $\begin{pmatrix} y_t \\ z_t \end{pmatrix}$ is integrated with $(1, -1)'$. Here the stochastic trend in the cointegration is purged. And $\beta_3 y_t + \beta_4 z_t \sim I(0) \Longleftrightarrow \beta_3/\beta_4 = -1$. See Figure 2 in Ch6-ex2.

2) Suppose $\varepsilon_{yt}, \varepsilon_{zt}, \varepsilon_{wt}, \varepsilon_t$ are i.i.d. white noise processes, and

$$\begin{aligned} y_t &= \mu_{yt} + \varepsilon_{yt}, \quad z_t = \mu_{zt} + \varepsilon_{zt}, \quad w_t = \mu_{wt} + \varepsilon_{wt}, \\ \mu_{yt} &= \mu_{y,t-1} + \varepsilon_t, \ \mu_{zt} = \mu_{z,t-1} + \varepsilon_t, \ \mu_{wt} = \mu_{yt} + \mu_{zt}. \end{aligned}$$

Then $y_t + z_t - w_t = \varepsilon_{yt} + \varepsilon_{zt} - \varepsilon_{wt} \sim I(0)$, i.e. $(1, 1, -1)(y_t, z_t, w_t)' \sim I(0)$. $(y_t, z_t, w_t)'$ is integrated with $(1, 1, -1)'$. Here the stochastic trend in the cointegration is also purged.

3) Consider the vector representation:

$$x_t = \mu_t + e_t$$

where $x_t = (x_{1t}, \cdots, x_{nt})'$, $\mu_t = (\mu_{1t}, \cdots, \mu_{nt})'$ is the vector of stochastic trends, and e_t is an $n \times 1$ vector of stationary components. If one trend can be expressed as a linear combination of the other trends in the system, i.e. there exists a vector $\beta = (\beta_1, \cdots, \beta_n)'$ such that $\beta' \mu_t = 0$, then $\beta' x_t = \beta' e_t \sim I(0)$. That is, x_t is integrated with $(\beta_1, \cdots, \beta_n)'$.

- **Cointegration and Error correction:** The vector x_t has an error-correction representation if it can be expressed in the form

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \pi_2 \Delta x_{t-2} + \cdots + \pi_p \Delta x_{t-p} + \varepsilon_t$$

where $\pi_0 = (\pi_{i0})_{n\times 1}$, $\pi = (\pi_{jk})_{n\times n} \neq 0$, $\pi_i = (\pi_{jk}(i))_{n\times n}, i = 1, 2, \cdots, p$. The components in the error term vector $\varepsilon_t = (\varepsilon_{it})_{n\times 1}$ may be correlated with each other, but are stationary.

Remarks:
(i) Suppose that $x_t \sim I(1)$. Since

$$\pi x_{t-1} = \Delta x_t - \pi_0 - \pi_1 \Delta x_{t-1} - \pi_2 \Delta x_{t-2} - \cdots - \pi_p \Delta x_{t-p} - \varepsilon_t,$$

πx_{t-1} is stationary. Each row of π is a cointegrating vector of x_t.

If $\pi = 0$, the model is a VAR in first difference of x_t. There is no error correction term, implying that Δx_t does not respond to the previous period's deviation from long-run equilibrium (or disequilibrium error).

If $\pi \neq 0$, Δx_t responds to the previous period's deviation from long-run equilibrium and estimating x_t as a VAR in the first difference by omitting the error correction term πx_{t-1} is inappropriate.

Example: Case 1. two-variable case VAR(1):

$$\begin{cases} y_t = a_{11} y_{t-1} + a_{12} z_{t-1} + \varepsilon_{yt} \\ z_t = a_{21} y_{t-1} + a_{22} z_{t-1} + \varepsilon_{zt} \end{cases}$$

where ε_{yt} and ε_{zt} are white-noise which may be correlated with each other. Write

$$\begin{cases} (1 - a_{11}L) y_t - a_{12} L z_t = \varepsilon_{yt} \\ -a_{21} L y_t + (1 - a_{22}L) z_t = \varepsilon_{zt} \end{cases}$$

and further

$$y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12} L \varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12} a_{21} L^2}$$

$$z_t = \frac{a_{21} L \varepsilon_{yt} + (1 - a_{11}L)\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12} a_{21} L^2}.$$

The two series $\{y_t\}$ and $\{z_t\}$ have the same characteristic equation:

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11} a_{22} - a_{12} a_{21}) = 0.$$

We can also write the model as

$$\begin{cases} \Delta y_t = (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \\ \Delta z_t = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt}. \end{cases} \tag{1}$$

Denote the two characteristic roots be λ_1 and λ_2. We are interested in the case that y_t, $z_t \sim CI(1, 1)$.

If λ_1 and λ_2 lie inside the unit circle, $\{y_t\}$ and $\{z_t\}$ cannot be integrated of order $(1, 1)$.

If either of λ_1 and λ_2 lies outside the unit circle, the solution is explosive. Neither variable is difference stationary and cannot be $CI(1, 1)$.

If both λ_1 and λ_2 are unity, $\{y_t\}$ and $\{z_t\}$ cannot be integrated of order $(1, 1)$. Note that $a_{12} = a_{21} = 0$ will make $\lambda_1 = \lambda_2 = 1$.Hence if y_t, $z_t \sim CI(1, 1)$, then $a_{12} \cdot a_{21} \neq 0$.

For $\{y_t\}$ and $\{z_t\}$ to be $CI(1, 1)$, **it is necessary that** $\lambda_1 = 1$, $|\lambda_2| < 1$. Then

$$a_{11} = 1 - \frac{a_{12}a_{21}}{1 - a_{22}}. \tag{2}$$

From (1) and (2), for y_t, $z_t \sim CI(1, 1)$,

$$\begin{cases} \Delta y_t = -\frac{a_{12}a_{21}}{1-a_{22}} y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} = -\frac{a_{12}a_{21}}{1-a_{22}} \left(y_{t-1} - \frac{1-a_{22}}{a_{21}} z_{t-1} \right) + \varepsilon_{yt} \\ \Delta z_t = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt} = a_{21} \left(y_{t-1} - \frac{1-a_{22}}{a_{21}} z_{t-1} \right) + \varepsilon_{zt} \end{cases}$$

or

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t$$

where

$$\begin{aligned} x_t &= \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \varepsilon_t = \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix}, \\ \pi &= \begin{pmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{pmatrix} = \begin{pmatrix} -\frac{a_{12}a_{21}}{1-a_{22}} & a_{12} \\ a_{21} & a_{22} - 1 \end{pmatrix} = \alpha\beta', \end{aligned}$$

where $\alpha = \left(-\frac{a_{12}a_{21}}{1-a_{22}}, a_{21} \right)'$, $\beta = (1, -\frac{1-a_{22}}{a_{21}})'$ with $rank(\beta) = 1$.

Remarks: 1) $\{y_{t-1} - \frac{1-a_{22}}{a_{21}} z_{t-1}\}$ is stationary and $y_t - \frac{1-a_{22}}{a_{21}} z_t = 0$ is the long-run equilibrium. We can see that y_t and z_t change in response to the previous period's deviation $y_{t-1} - \frac{1-a_{22}}{a_{21}} z_{t-1}$ from the long-run equilibrium. Here the normalized cointegrating vector is $(1, -\frac{1-a_{22}}{a_{21}})'$.

2) Granger representation theorem—error correction and cointegration are equivalent representation: $CI(1, 1)$ guarantees the existence of an error-correction model and an error-correction model for $I(1)$ variables implies cointegration.

3) A cointegrated system can be viewed as a restricted form of a general VAR model. It is inappropriate to estimate a VAR of cointegrated variables using only first differences by ignoring the error-correction portion of the model.

4) If $rank(\pi) = 0$, we have $a_{12} = a_{21} = 0, a_{11} = a_{22} = 1$, and $\Delta x_t = \varepsilon_t$ or $x_t \nsim CI(1,1)$. If the variables are cointegrated, the rows of π must be linearly dependent, and hence $\det(\pi) = 0$ or the rank of π is 1, since if $rank(\pi) = 2$, there is no unit root for y_t and z_t, hence $x_t \nsim CI(1,1)$.

Example: Case 2. n-variable case VAR(1):

$$x_t = A_1 x_{t-1} + \varepsilon_t$$

where $x_t = (x_{1t}, x_{2t}, \cdots, x_{nt})'$, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{nt})', i.i.d.(0, \Omega)$, A_1 is an $n \times n$ matrix of parameters.

$$\begin{aligned} \Delta x_t &= (A_1 - I)x_{t-1} + \varepsilon_t \\ &= \pi x_{t-1} + \varepsilon_t. \end{aligned}$$

The rank of π determine the number of cointegration vectors. For example,

1) $rank(\pi) = 0 : \pi = 0$ and $\Delta x_t = \varepsilon_t$. All the sequences $\{x_{it}\}$ are unit root processes and there is no linear combination of the variables that is stationary. Hence $x_t \nsim CI(1,1)$.

2) $rank(\pi) = n : det(\pi) \neq 0$ (there is no unit root) and each row of $\pi x_{t-1} = 0$ is an independent restriction on the long-run solution of the variables. Each of the n variables in x_t must be stationary with the corresponding long-run value constrain in $\pi x_{t-1} = 0$. Hence $x_t \nsim CI(1,1)$.

3) $rank(\pi) = 1 : there$ is a single cointegrating vector given by any row of π, e.g. for

$$\Delta x_{1t} = \pi_{11}(x_{1t-1} + \beta_{12}x_{2t-1} + \cdots + \beta_{1n}x_{nt-1}) + \varepsilon_{1t},$$

the normalized cointegrating vector is $(1, \beta_{12}, \cdots, \beta_{1n})$, where $\beta_{ij} = \pi_{ij}/\pi_{11}$.

4) $0 < rank(\pi) = r < n$: there are r cointegrating vectors and $n - r$ stochastic trends in the system.

- **3 Test for Cointegration——Engle-Granger Method: based on residuals**

Consider the linear regression model

$$y_t = z_t'\beta + e_t$$

where the $k \times 1$ vector $z_t \sim I(1)$ and elements of z_t are not cointegrated. Further, assume

$$e_t = a_1 e_{t-1} + v_t$$

where $v_t \sim I(0)$. Testing the null

$$H_0 : a_1 = 1$$

amounts to testing the null of non-cointegration of y_t and z_t.

Four steps to test for the cointegration of y_t and z_t : (take $k = 1$ as an example):

1) pretest each variable to determine its integration order (Augmented Dickey-Fuller test infers the number of unit roots);

2) Estimate the long-run equilibrium relationship. If the above test indicates that both y_t and z_t are $I(1)$, estimate the long-run equilibrium relationship:

$$y_t = \beta_0 + \beta_1 z_t + e_t. \tag{3}$$

Conduct D-F test for the AR(1) model of the above residuals $\hat{e}_t$:

$$\Delta \hat{e}_t = a_1 \hat{e}_{t-1} + \varepsilon_t, \quad H_0 : a_1 = 0$$

or, if the the residuals ε_t from the above regression exhibits serial correlation,

$$\Delta \hat{e}_t = a_1 \hat{e}_{t-1} + \sum_{i=1}^{p} a_{i+1} \Delta \hat{e}_{t-i} + \varepsilon_t, \quad H_0 : a_1 = 0 \text{ (no cointegration)}$$

The OLS estimator a_1 and its test τ-statistic for $H_0 : a_1 = 0$ have no standard asymptotic distributions. The asymptotic distribution can be derived by the similar argument to that in Chapter 4. The critical value for the test should be obtained by simulation. If rejecting the null H_0, we conclude that the variables are cointegrated. Here the critical values for the test are provided in **Table C** at the end of the text (for the Engle-Granger cointegration test).

3) Estimate the error correction model. Use the residuals from the equilibrium regression (3) as the deviation from the long-run equilibrium in the error correction model

$$\begin{aligned}
\Delta y_t &= \alpha_1 + \alpha_y \hat{e}_{t-1} + \sum_{i=1} \alpha_{11}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{12}(i) \Delta z_{t-i} + \varepsilon_{yt} \\
\Delta z_t &= \alpha_2 + \alpha_z \hat{e}_{t-1} + \sum_{i=1} \alpha_{21}(i) \Delta y_{t-i} + \sum_{i=1} \alpha_{22}(i) \Delta z_{t-i} + \varepsilon_{zt}
\end{aligned}$$

which constitutes VAR in first difference and can be estimated using the conventional method for a VAR model.

4) Assess model adequacy. Estimate the error correction model by adjusting the lag lengths such that the errors are serially uncorrelated.

Example: see Ch6-ex3. Test the cointegration of y_t, z_t and w_t, where the data are generated from

$$\begin{aligned} y_t &= \mu_{yt} + \delta_{yt}, \quad \mu_{yt} = \mu_{y,t-1} + \varepsilon_{yt}, \ \delta_{yt} = 0.5\delta_{y,t-1} + \eta_{yt} \\ z_t &= \mu_{zt} + \delta_{zt} + 0.5\delta_{yt}, \quad \mu_{zt} = \mu_{t,t-1} + \varepsilon_{zt}, \ \delta_{zt} = 0.5\delta_{z,t-1} + \eta_{zt} \\ w_t &= \mu_{wt} + \delta_{wt} + 0.5\delta_{yt} + +0.5\delta_{zt}, \quad \mu_{wt} = \mu_{yt} + \mu_{zt}, \ \delta_{wt} = 0.5\delta_{w,t-1} + \eta_{wt} \end{aligned}$$

where $\varepsilon_{yt}, \eta_{yt}, \varepsilon_{zt}, \eta_{zt}$ and η_{wt} are all white noise processes. The true relationship is that $(y_t, z_t, w_t)'$ is cointegrated with $(1, 1, -1)$.

Remark: The Engle-Granger method has the following shortcomings: (i) In Step 2 above, the long-run equilibrium relationship (3) can also be set as

$$z_t = \beta_0 + \beta_1 y_t + e_t.$$

As the sample size $T \to \infty$, the two settings give equivalent results in testing for a unit root of the residuals. However, in practice, it is possible to find that one regression indicates that the variables are cointegrated, whereas reversing the order indicates no cointegration. The test for more than two variables becomes more troubled. (ii) Also in Step 2, the residuals from (3) are used to estimate $\hat{a}_1$ in the regression. Any error introduced from the regression (3) is carried into the estimation of a_1 and the no-cointegration test. Hence the following Johansen-Stock-Watson Method.

- **4 Test for Cointegration——Johansen-Stock-Watson Method**

Some forms of the models:

1) The model without a drift is

$$Model\ 1 : x_t = A_1 x_{t-1} + \varepsilon_t \ \text{ or } \ \Delta x_t = \pi x_{t-1} + \varepsilon_t,$$

where $\pi = A_1 - I$. The rank of π equals the number of cointegrating vectors (also called the cointegration rank, denoted as $r \equiv rank(\pi)$). If $r = 0$, all the $\{x_{it}\}$, $i = 1, 2, \cdots, n$, are unit root processes. There is no linear combination of $\{x_{it}\}$ that is stationary, and hence x_t is not cointegrated. If $r = n$, there is no unit root in each x_{it}, and hence $x_{it} \notin CI(1, 1)$. If $1 \leq r < n$, there are r cointegrating vectors and $n - r$ stochastic trends in the system.

2) Add a drift term, if the variables exhibit a decided tendency to increase or decrease:

$$Model\ 2 : x_t = A_0 + A_1 x_{t-1} + \varepsilon_t \ \text{ or } \Delta x_t = A_0 + \pi x_{t-1} + \varepsilon_t.$$

Here $A_0 = (a_{10}, a_{20}, \cdots, a_{n0})'$ allows for the possibility of a linear trend in the data generating process. $rank(\pi)$ = the number of cointegrating relationships existing in the "detrended" data. For the i-th series x_{it},

$$\Delta x_{it} = a_{0i} + (\pi_{i1}x_{1,t-1} + \cdots + \pi_{n1}x_{n,t-1}) + \varepsilon_{it}.$$

In the long run, $\pi_{i1}x_{1,t-1} + \cdots + \pi_{n1}x_{n,t-1} = 0$, and $E\Delta x_{it} = a_{0i}$. Aggregating all such changes over t yields the deterministic expression $a_{0i}t$.See Figure 6.3 in Text: P350.

3) Include a constant in the cointegrating relationship: e.g. assume that $rank(\pi) = 1$ and $a_{i0} = s_i a_{10}$

$$\begin{aligned} Model\ 3 \ : \ \ \Delta x_{1t} &= (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10}) + \varepsilon_{1t} \\ \Delta x_{2t} &= s_2 (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10}) + \varepsilon_{2t} \\ &\cdots \\ \Delta x_{1t} &= s_n (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + a_{10}) + \varepsilon_{nt} \end{aligned}$$

or $\Delta x_t = \pi^* x^*_{t-1} + \varepsilon_t$, where $x^*_{t-1} = (x_{1t-1}, x_{2t-1}, \cdots, x_{nt-1}, 1)'$ and $\pi^* = (\pi, A_0)$.Here the linear trend is purged from the system.

4) Include an intercept term in the cointegrating vector along with a drift term: e.g. assume that $rank(\pi) = 1$,

$$\begin{aligned} Model\ 4 \ : \ \ \Delta x_{1t} &= (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + b_{10}) + b_{11} + \varepsilon_{1t} \\ \Delta x_{2t} &= s_2 (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + b_{10}) + b_{21} + \varepsilon_{2t} \\ &\cdots \\ \Delta x_{1t} &= s_n (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1} + b_{10}) + b_{n1} + \varepsilon_{nt} \end{aligned}$$

where $s_i b_{10} + b_{i1} = a_{i0}$. EViews identifies the portion belonging in the cointegrating vector as the amount necessary to force the error correction term to have a sample mean of zero.

5) Higher-order AR process:

$$Model\ 5 : x_t = A_1 x_{t-1} + A_2 x_{t-2} + \cdots + A_p x_{t-p} + e_t$$

or

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t \tag{4}$$

where $\pi = \sum_{i=1}^{p} A_i - I$, $\pi_i = -\sum_{j=i+1}^{p} A_j$, and $rank(\pi)$ = the number of independent cointegrating vector (i.e. the cointegration rank). If $rank(\pi) = 0$, $\pi = 0$ and the model is the usual VAR model in the first difference. If $rank(\pi) = n$, the vector process is stationary. If $rank(\pi) = 1$, there is a single cointegration vector; if $1 < r \equiv rank(\pi) < n$, there are r cointegration vectors, and πx_{t-1} is the error correction term. For $1 \leq r < n$, π can be decomposed as

$$\pi = \alpha\beta'$$

where α and β are two $n \times r$ matrix with $rank(\beta) = r$. We can say $\beta' x_t = 0$ is the long run equilibrium.

Likelihood ratio test of cointegration rank

(A) Log-likelihood function

Consider the VAR(p) model (4), where $e_t \sim iidN(0, \Sigma)$, $t = 1, 2, \cdots, T$, and $X_{1-p}, \cdots, X_0$ are given constant vectors. The log-likelihood function is derived from the multivariate normal distribution:

$$\ln L(x_1, x_2, \cdots, x_T; \pi_1, \pi_2, \cdots, \pi_{p-1}, \pi, \Sigma) = \frac{-nT}{2} \ln(2\pi) - \frac{T}{2} \ln|\Sigma| - \frac{1}{2} \sum_{t=1}^{T} e_t' \Sigma^{-1} e_t.$$

The following steps will be taken for the likelihood ratio test:

(i) Concentrate $\ln L$ with respect to Σ : by $\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} e_t e_t'$,

$$\ln L^*(x_1, x_2, \cdots, x_T; \pi_1, \pi_2, \cdots, \pi_{p-1}, \pi) = C - \frac{T}{2} \ln \left| \sum_{t=1}^{T} e_t e_t' \right|.$$

(ii) Concentrate $\ln L^*$ with respect to $\pi_1, \pi_2, \cdots, \pi_{p-1}$. Use the partial-out approach. Let

$$q_t = (\Delta x_{t-1}, \Delta x_{t-2}, \cdots, \Delta x_{t-p+1})'.$$

Denote

$$R_{0t} = \Delta x_t - \sum_{i=1}^{p-1} \hat{\pi}_i \Delta x_{t-i}$$

$$R_{1t} = x_{t-1} - \sum_{i=1}^{p-1} \tilde{\pi}_i \Delta x_{t-i}$$

where

$$(\hat{\pi}_1, \hat{\pi}_2, \cdots, \hat{\pi}_{p-1}) = \left(\sum_{t=1}^{T} q_t q_t' \right)^{-1} \left(\sum_{t=1}^{T} \Delta x_t q_t' \right)$$

$$(\tilde{\pi}_1, \tilde{\pi}_2, \cdots, \tilde{\pi}_{p-1}) = \left(\sum_{t=1}^{T} q_t q_t' \right)^{-1} \left(\sum_{t=1}^{T} x_{t-1} q_t' \right).$$

Then

$$\begin{aligned}\ln L^{**}(x_1, x_2, \cdots, x_T; \pi) &= C_0 - \frac{T}{2} \ln \left| \sum_{t=1}^{T} (R_{0t} - \pi R_{1t}) (R_{0t} - \pi R_{1t})' \right| \\ &= C_{00} - \frac{T}{2} \ln |s_{00} - \pi s_{10} - s_{01}\pi' + \pi s_{11}\pi'| ,\end{aligned}$$

where

$$s_{ij} = \frac{1}{T} \sum_{t=1}^{T} R_{it} R'_{jt}, \quad i, j = 0, 1.$$

(iii) Imposing the restriction $\pi = \alpha\beta'$, we have

$$\ln L^{**}(x_1, x_2, \cdots, x_T; \alpha, \beta) = C_{00} - \frac{T}{2} \ln |s_{00} - \alpha\beta' s_{10} - s_{01}\beta\alpha' + \alpha\beta' s_{11}\beta\alpha'| .$$

By

$$\frac{\partial \ln L^{**}(x_1, x_2, \cdots, x_T; \alpha, \beta)}{\partial \alpha} = 0 \Longrightarrow \hat{\alpha} = s_{01}\beta \left(\beta' s_{11}\beta\right)^{-1} ,$$

we have, by normalizing $\beta' s_{11}\beta = I$ and substitititing α with $\hat{\alpha}$,

$$\begin{aligned}\ln L^{***}(x_1, x_2, \cdots, x_T; \beta) &= \ln L^{**}(x_1, x_2, \cdots, x_T; \hat{\alpha}, \beta) \\ &= C_1 - \frac{T}{2} \ln \left| s_{00} - s_{01}\beta \left(\beta' s_{11}\beta\right)^{-1} \beta' s_{10} \right| \\ &= C_1 - \frac{T}{2} \ln \left[|\beta' s_{11}\beta|^{-1} |s_{00}| \left| \beta' \left(s_{11} - s_{10}s_{00}^{-1}s_{01}\right) \beta \right| \right] \\ &= C_1 - \frac{T}{2} \ln |s_{00}| - \frac{T}{2} \ln \left| \beta' \left(s_{11} - s_{10}s_{00}^{-1}s_{01}\right) \beta \right| .\end{aligned}$$

Therefore, MLE requires

$$\begin{aligned}&\text{minimize } \left| \beta' \left(s_{11} - s_{10}s_{00}^{-1}s_{01}\right) \beta \right| \\ &\text{subject to } \beta' s_{11}\beta = I.\end{aligned}$$

(iv) Suppose $rank(\pi) = r$. Order the characteristic roots of π such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_n = 0$. If the variables in x_t are not cointegrated, $rank(\pi) = 0$, all $\lambda_i = 0$ and $\ln(1-\lambda_i) = 0$. If $rank(\pi) = 1, 0 < \lambda_1 < 1$, then $\ln(1-\lambda_1) < 0$ and $\ln(1-\lambda_i) = 0$ for $i = 2, 3, \cdots, n$. It is deduced from the theory of matrix charactristic root that the maximized value of the likelihood function is given by

$$\begin{aligned}\ln L^{****}(x_1, x_2, \cdots, x_T; r) &= C_2 - \frac{T}{2} \ln \left| \beta' \left(s_{11} - s_{10}s_{00}^{-1}s_{01}\right) \beta \right| \\ &= K - \frac{T}{2} \sum_{i=1}^{r} \ln(1 - \lambda_i),\end{aligned}$$

where K is a constant.

(B) Test two null hypothesis

1) The first statistic tests the null hypothesis that the number of distinct cointegrating vectors is less than or equal to r against a general alternative. That is, to test

$$H_0 : rank(\pi) \leq r \quad (that\ is,\ \lambda_{r+1} = \cdots = \lambda_n = 0)$$

against $H_1 : rank(\pi) > r$ (i.e. at least one of $\lambda_{r+1}, \cdots, \lambda_n$ is not equal to zero). When π is unrestricted, the log-likelihood function is

$$L_u = K - \frac{T}{2} \sum_{i=1}^{n} \ln(1 - \lambda_i).$$

Under H_0, $\lambda_{r+1} = \cdots = \lambda_n = 0$, the log likelihood function is

$$L_r = K - \frac{T}{2} \sum_{i=1}^{r} \ln(1 - \lambda_i).$$

Therefore, the likelihood ratio is

$$2(L_u - L_r) = -T \sum_{i=r+1}^{n} \ln(1 - \lambda_i).$$

This is the trace-statistic $\lambda_{trace}(r)$:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i).$$

When all $\lambda_i = 0$, $\lambda_{trace}(r) = 0$; The further the estimated characteristic roots are from zero, the more neagtive is $\ln(1-\hat{\lambda}_i)$, and the larger is the trace-statistic $\lambda_{trace}(r)$. Critical values of $\lambda_{trace}(r)$ are obtained by simulation, see Table E in Enders's textbook.

2) The second statistic tests the null hypothesis that the number of distinct cointegrating vectors is r against the alternative of $r + 1$ cointegrating vectors. That is, to test

$$H_0 : rank(\pi) = r$$

against the alternative hypothesis $H_1 : rank(\pi) = r+1$. Under H_1, $\lambda_{r+2} = \cdots = \lambda_n = 0$, so the log likelihood function is

$$L_u = K - \frac{T}{2} \sum_{i=1}^{r+1} \ln(1 - \lambda_i).$$

Under the restriction H_0, $\lambda_{r+1} = \cdots = \lambda_n = 0$, so the likelihood function is

$$L_r = K - \frac{T}{2}\sum_{i=1}^{r} \ln(1 - \lambda_i).$$

Therefore, the likelihood ratio is

$$2(L_u - L_r) = -T \log(1 - \lambda_{r+1}).$$

This is the maximum-eigenvalue statistic $\lambda_{\max}(r, r+1)$:

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}).$$

When the estimated $\hat{\lambda}_{r+1}$ is close to zero, $\lambda_{\max}(r, r+1)$ will be small. The further the estimated characteristic root are from zero, the more neagtive is $\ln(1 - \hat{\lambda}_{r+1})$, and the larger is $\lambda_{\max}(r, r+1)$. Critical values of the test are in Table E (the case without any deterministic regressors: the first cell) at the end of the text.

(C) Extensions of the Johansen test

1) Cointegration test with unrestricted intercept. Consider the VAR model with an intercept

$$\Delta x_t = \mu + \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t$$

where μ is unrestricted. Replacing R_{0t} and R_{1t}, respectively, by

$$\begin{aligned} R_{0t} &= \Delta x_t - \sum_{i=1}^{p-1} \hat{\pi}_i \Delta x_{t-i} - \hat{\mu} \\ R_{1t} &= x_{t-1} - \sum_{i=1}^{p-1} \tilde{\pi}_i \Delta x_{t-i} - \tilde{\mu}, \end{aligned}$$

we proceed in the same manner. The limiting distributions of the trace and maximum-eigenvalue tests change. Their critical values are listed in Table E (the case with drift: the second cell) at the end of the text according to the form of the vector A_0.

2) Cointegration test with restricted intercept, that is, with a constant in the cointegrating vector. Consider

$$\Delta x_t = \mu + \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t$$

with $\pi = \alpha\beta'$,where α and β are two $n \times r$ matrix with $rank(\beta) = r$. The intercept is restricted: $\mu = \alpha\beta_0'$ for any arbitrary $1 \times r$ vector β_0. Then

$$\begin{aligned}
\Delta x_t &= \alpha\left(\beta_0' + \beta' x_{t-1}\right) + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t \\
&= \alpha\left(\beta_0', \beta'\right)\begin{pmatrix} 1 \\ x_{t-1} \end{pmatrix} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t \\
&\equiv \pi^* x_{t-1}^* + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t.
\end{aligned}$$

Note the presence of intercept in the cointegrating vectors as opposed to the unrestricted drift in 1) above. Formulate the trace and maximum-eigenvalue tests in the same manner as before, but replace R_{1t} by

$$R_{1t}^* = x_{t-1}^* - \sum_{i=1}^{p-1} \tilde{\pi}_i \Delta x_{t-i}.$$

Critical values for these tests should change, and are listed in Table E (the case with a constant in the cointegrating vector: the third cell) at the end of the text.

(D) Hypothesis testing about some restricted forms of cointegrating vectors

Note that if there are r cointegrating vectors, only these r linear combinations of the variables are stationary. All other linear combinations are nonstationary. Suppose you re-estimate the model restricting the parameters of π. If the restrictions are not binding, it should be founded that the number of cointegrating vectors has not diminished. Some cases:

1) Test H_0 : including an intercept in the cointegrating vector as opposed to the unrestricted drift A_0 (i.e. a linear time trend). Estimate the two forms of the model and obtain the ordered characteristic roots:

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_n$$

(without an intercept in the cointegrating vector, i.e. with unrestricted π)

$$\hat{\lambda}_1^* \geq \hat{\lambda}_2^* \geq \cdots \geq \hat{\lambda}_n^*$$

(with an intercept in the cointegrating vector, i.e. with restricted π).

Assume that the unrestricted model has r nonzero characteristic roots. Then the likelihood ratio statistic is

$$\begin{aligned}
& 2\left[\left(-\frac{T}{2}\sum_{i=1}^{r}\ln(1-\hat{\lambda}_i)\right)-\left(-\frac{T}{2}\sum_{i=1}^{r}\ln(1-\hat{\lambda}_i^*)\right)\right] \\
= \;& T\sum_{i=1}^{r}\ln\left(\frac{1-\hat{\lambda}_i^*}{1-\hat{\lambda}_i}\right) \\
\sim \;& \chi^2(n-r) \text{ asymptotically.}
\end{aligned}$$

The intuition is that: if the restriction is true, $\ln(1-\hat{\lambda}_i^*)$ and $\ln(1-\hat{\lambda}_i)$ should be equivalent. Hence, small values for the above statistic imply the restriction (i.e. including an intercept in the cointegrating vector); large values imply that it is possible to reject the null of including an intercept in the cointegrating vectors, and that there is a linear trend in the variables.

2) Test for restrictions on γ or α, where $\pi = \alpha\beta'$, α $(n \times r)$ is the matrix of the speed of adjustment parameters, and β' $(r \times n)$ is the matrix of cointegrating parameters. Use MLE to estimate VAR(p) model (4), where $e_t \sim iidN(0, \Sigma)$, $t = 1, 2, \cdots, T$, and $X_{1-p}, \cdots, X_0$ are given constant vectors, determine the rank of π, use the r most significant cointegrating vectors to form α, then select γ such that $\pi = \alpha\beta'$.For example, if $rank(\pi) = 1$, then

$$\begin{aligned}
\Delta x_{1t} &= (\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}) + \cdots + e_{1t} \\
\Delta x_{2t} &= s_2\left(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}\right) + \cdots + e_{2t} \\
&\cdots \\
\Delta x_{1t} &= s_n\left(\pi_{11}x_{1t-1} + \pi_{12}x_{2t-1} + \cdots + \pi_{1n}x_{nt-1}\right) + \cdots + e_{nt}.
\end{aligned}$$

Define $\alpha_1 = \pi_{11}, \alpha_i = s_i\pi_{11}, \beta_i = \pi_{1i}/\pi_{11}, i = 2, 3, \cdots, n$. Then the model can be written as

$$\Delta x_{it} = \alpha_i\left(x_{1t-1} + \beta_2 x_{2t-1} + \cdots + \beta_n x_{nt-1}\right) + \cdots + e_{it},\ i = 1, \cdots, n$$

or

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\pi_i\Delta x_{t-i} + e_t$$

where

$$\begin{aligned}
\alpha &= (\alpha_1, \alpha_2, \cdots, \alpha_n)',\ \beta = (1, \beta_2, \cdots, \beta_n)', \\
\pi &= \begin{pmatrix} \alpha_1 & \alpha_1\beta_2 & \cdots & \alpha_1\beta_n \\ \alpha_2 & \alpha_2\beta_2 & \cdots & \alpha_2\beta_n \\ \vdots & \vdots & & \vdots \\ \alpha_n & \alpha_n\beta_n & \cdots & \alpha_n\beta_n \end{pmatrix}.
\end{aligned}$$

So, we can test various restrictions on α and β by comparing the number of cointegrating vectors under the null and alternative hypotheses. The test statistic is

$$\begin{aligned}
& \left(-T\sum_{i=1}^{n}\ln(1-\hat{\lambda}_i)\right) - \left(-T\sum_{i=1}^{n}\ln(1-\hat{\lambda}_i^*)\right) \\
= \;& T\sum_{i=1}^{n}\left[\ln(1-\hat{\lambda}_i^*) - \ln(1-\hat{\lambda}_i)\right] \\
\sim \;& \chi^2(\text{the number of restrictions}),\ \text{asymptotically}.
\end{aligned}$$

If $\alpha_i = 0$, the variable x_{it} is weakly exogenous. The practice importance is that a weakly exogenous variable does not experience the type of feedback, i.e. it does not respond to the deviation from the long-run equilibrium.

Remark: (i) Note that, in testing for the restriction (about β) on **multiple cointegration vectors**, the number of restrictions imposed on the system is not the number of the equations for the restriction. For eaxmple, if $n = 4, r = 2$, and we normalize each cointegration vector with respect to x_{1t}, we can write $\beta' x_t$ as

$$\begin{pmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 1 & -\beta_{22} & -\beta_{23} & -\beta_{24} \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \\ x_{4t} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since any linear combination of these two cointegrating vectors is also a cointegrating vector, by

$$\begin{pmatrix} 1 & 0 & -\beta_{13}^* & -\beta_{14}^* \\ 0 & 1 & -\beta_{23}^* & -\beta_{24}^* \end{pmatrix},$$

(x_{1t}, x_{3t}, x_{4t}) are cointegrated such that $x_{1t} = \beta_{13}^* x_{3t} + \beta_{14}^* x_{4t}$; (x_{2t}, x_{3t}, x_{4t}) are cointegrated such that $x_{2t} = \beta_{23}^* x_{3t} + \beta_{24}^* x_{4t}$. More generally, $n-r+1$ variables are cointegrated. A testable exclusion restriction entails the exclusion of r or more variables from a cointegrating vector. Hence, excluding $r_0(\geq r)$ variables from a cointegrating vector entails only $(r_0 - r + 1)$ restriction, and the degree freedom for the χ^2 statistic is only $(r_0 - r + 1)$ instead of r_0.

(ii) It is possible to test for the restriction on one cointegrating vector **conditional on** the values of all other cointegrating vectors. For example, $n = 4, r = 2$. We want to test if $(1, \beta_{22}, 0, \beta_{24})'$ is a cointegrating vector for the fixed normalized values of β_{12}, β_{13},and β_{14}. That is,

$$\begin{pmatrix} 1 & -\beta_{12} & -\beta_{13} & -\beta_{14} \\ 1 & -\beta_{22} & 0 & -\beta_{24} \end{pmatrix} \begin{matrix} \leftarrow \text{fixed} \\ \leftarrow \text{test } \beta_{23} = 0 \end{matrix}$$

3) Test lag length p in the model: Since all of the Δx_{t-i} are stationary, we can use χ^2 statistic similar to t or F-test. Let Σ_r and Σ_u be the variance-covariance matrices of the restricted and unrestricted systems, respectively. Use the test statistic

$$(T - c)\left(\log|\Sigma_r| - \log|\Sigma_u|\right) \sim \chi^2(number\ of\ restrictions)$$

where c is the maximum number of regressors contained in the longest equation. AIC and SBC can be also applied. An F-test test for the lag lengths in a single equation can be used, too.

4) Granger causality test is invalid if the variables in the model are cointegrated. The coefficients π of nonstationary variables are blamed for this failure.

- **Four Steps in Johansen-Stock-Watson Method:**

  1. Pretest all variables to assess their order of integration.
  2. Estimate the model and determine the rank of π.
  3. Analyze the normalized cointegrating vectors and speed of adjustment coefficients and test the restrictions about each of both.
  4. Conduct innovation accounting and causality tests on the error-correction model to identify a structural model and determine whether the estimated model appears to be reasonable.

- Difference or Not Difference? VAR or ECM?

If the I(1) variables x_t are cointegrated, differencing them and estimating a VAR:

$$\Delta x_t = \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + e_t$$

will lead to a misspecification error since it excludes the long-run equilibrium relationship among the variables that are included in πx_{t-1}.If the I(1) variables x_t are not cointegrated, it is preferable to estimate the VAR in first differences.

$$\begin{aligned}&\text{If I(1) } x_t \text{ are not cointegrated} \implies \text{estimate VAR in first differences}\\&\text{If I(1) } x_t \text{ are cointegrated} \implies \text{estimate the error-correction model}\end{aligned}$$

Hence the importance of testing unit root of each variable and cointegration of the variables.

- Linear or Nonlinear?