Ch. 23 Cointegration

(May 28, 2018)

1 Introduction

An important property of $I(1)$ variables is that there can be linear combinations of theses variables that are $I(0)$. If this is so then these variables are said to be *cointegrated*. Suppose that we consider two variables Y_t and X_t that are $I(1)$. (For example, $Y_t = Y_{t-1} + \zeta_t$ and $X_t = X_{t-1} + \eta_t$.) Then, Y_t and X_t are said to be cointegrated if there exists a β such that $Y_t - \beta X_t$ is $I(0)$. What this mean is that the regression equation

$$Y_t = \beta X_t + u_t$$

make sense because Y_t and X_t do not drift too far apart from each other over time. Thus, there is a long-run equilibrium relationship between them. If Y_t and X_t are not cointegrated, that is, $Y_t - \beta X_t = u_t$ is also $I(1)$, then Y_t and X_t would drift apart from each other over time. In this case the relationship between Y_t and X_t that we obtain by regressing Y_t and X_t would be spurious.

Let us continue the cointegration with the spurious regression setup in which X_t and Y_t are independent random walks, consider what happens if we take a nontrivial linear combination of X_t and Y_t:

$$a_1 Y_t + a_2 X_t = a_1 Y_{t-1} + a_2 X_{t-1} + a_1 \zeta_t + a_2 \eta_t,$$

where a_1 and a_2 are not both zero. We can write this as

$$Z_t = Z_{t-1} + v_t,$$

where $Z_t = a_1 Y_t + a_2 X_t$ and $v_t = a_1 \zeta_t + a_2 \eta_t$. Thus, Z_t is again a random walk process, as v_t is *i.i.d.* with mean zero and finite variance, given that ζ_t and η_t each are *i.i.d.*

with mean zero and finite variance. No matter what coefficients a_1 and a_2 we choose, the resulting linear combination is again a random walk, hence an integrated or unit root process.

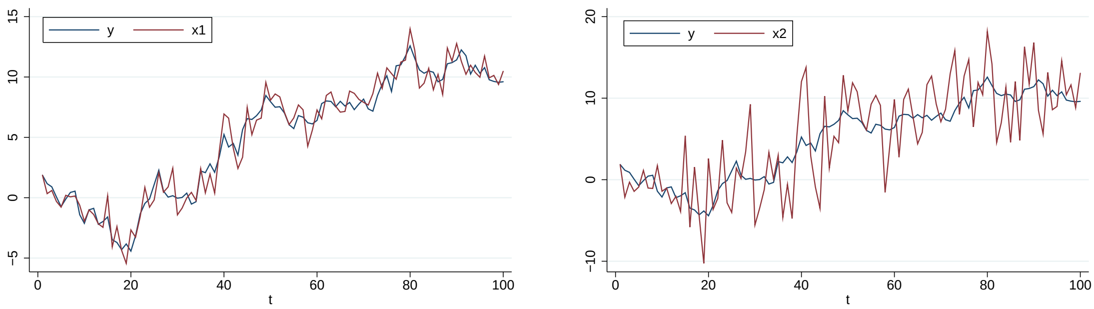
Now consider what happens when X_t is a random walk as before, but Y_t is instead generated according to $Y_t = \beta X_t + u_t$, with u_t again *i.i.d.*. By itself, Y_t is an integrated process, because

$$Y_t - Y_{t-1} = (X_t - X_{t-1})\beta + u_t - u_{t-1},$$

so that

$$\begin{aligned} Y_t &= Y_{t-1} + \beta\eta_t + u_t - u_{t-1} \\ &= Y_{t-1} + \varepsilon_t, \end{aligned}$$

where $\varepsilon_t = \beta\eta_t + u_t - u_{t-1}$ is readily verified to be $I(0)$ process.



The cointegration with smaller variance of cointegrating error   The cointegration with larger variance of cointegrating error

Fig.1 Examples of Cointegrated Variables

Despite the fact that both X_t and Y_t are integrated processes, the situation is very different from that considered at last chapter. Here, there is indeed a linear combination of X_t and Y_t that are not an integrated process: putting $a_1 = 1$ and $a_2 = -\beta$ we have

$$a_1 Y_t + a_2 X_t = Y_t - \beta X_t = u_t,$$

which is *i.i.d.* This is an example of a pair $\{X_t, Y_t\}$ of cointegrated process.

The concept of *cointegration* was introduced by Granger (1981). This paper and that of Engle and Granger (1987) have had a major impact on modern econometrics. Following Engle and Granger (1987), we have the definition of cointegration formally as follows.

Definition 1.
The components of the vector $\mathbf{x}_t$ are said to be co-integrated of order d, b, denoted $\mathbf{x}_t \sim CI(d,b)$, if

(a). all components of $\mathbf{x}_t$ are $I(d)$;

(b). there exists a vector $\mathbf{a}(\neq \mathbf{0})$ so that $z_t = \mathbf{a}'\mathbf{x}_t \sim I(d-b),\ b > 0$. The vector $\mathbf{a}$ is called the co-integrating vector. ■

For ease of exposition, only the value $d = 1$ and $b = 1$ will be considered in this chapter. For the case that d and b are fractional value, this is called fractional cointegration. We will consider this case in Chapter 25.

Clearly, the cointegrating vector $\mathbf{a}$ is not unique, for if $\mathbf{a}'\mathbf{x}_t$ is $I(0)$, then so is $b \cdot \mathbf{a}'\mathbf{x}_t$ for any nonzero scalar b; if $\mathbf{a}$ is a cointegrating vector, then so is $b\mathbf{a}$.

Furthermore, if $\mathbf{x}_t$ has k components, then there may be more than one cointegrating vector $\mathbf{a}$. Indeed, there may be $h < k$ linear independent $(k \times 1)$ vectors $(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_h)$ such that $\mathbf{A}'\mathbf{x}_t$ is a $I(0)$ $(h \times 1)$ vector, where $\mathbf{A}'$ is the following $(h \times k)$ matrix:

$$\mathbf{A}'_{h\times k} = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}'_h \end{bmatrix},$$

and is called the cointegrating matrix.

Again, the vector $(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_h)$ are not unique; if $\mathbf{A}'\mathbf{x}_t$ is a $I(0)$, then for any nonzero $(1 \times h)$ vector $\mathbf{b}'$, the scalar $\mathbf{b}'\mathbf{A}'\mathbf{x}_t$ is also $I(0)$. Then the $(k \times 1)$ vector $\boldsymbol{\phi}$ given by $\boldsymbol{\phi}' = \mathbf{b}'\mathbf{A}'$ could also be described as a cointegrating vector.[1]

It is seen that the paramtetrs in $\mathbf{A}$ are not identified. What can be determined by the model is the space spanned by $\mathbf{A}$, the cointegration space $sp(\mathbf{A})$. Suppose that there exists an $(h \times k)$ matrix $\mathbf{A}'$ whose rows are linearly independent such that $\mathbf{A}'\mathbf{x}_t$ is a $(h \times 1)$ $I(0)$ vector. Suppose further that if $\mathbf{c}'$ is any $(1 \times k)$ vector that is linearly independent of the rows of $\mathbf{A}'$, then $\mathbf{c}'\mathbf{x}_t$ is a $I(1)$ scalar. Then we say that there are

[1] Recalling the results that $rank(\mathbf{AB}) \leq \min(rank(\mathbf{A}), rank(\mathbf{B}))$.

exactly h cointegrating relations among the elements of $\mathbf{x}_t$ and that ($\mathbf{a}_1$, $\mathbf{a}_2$,..., $\mathbf{a}_h$) form a *basis* for the space of the cointegrating vectors.

Example.

Let P_t denote an index of the price level in the United States, P_t^* a price index for Italy, and S_t the exchange rate between the currency. Then purchasing power parity holds that

$$P_t = S_t P_t^*,$$

or, taking logarithms,

$$p_t = s_t + p_t^*,$$

where $p_t \equiv \log P_t$, $s_t \equiv \log S_t$, and $p_t^* \equiv \log P_t^*$. In equilibrium we need $p_t - s_t - p_t^* = 0$. However, in practice, error in measuring price, transportation costs, and differences in quality prevent purchasing power parity from holding exactly at every date t. A weaker form of the hypothesis is that the variable z_t defined by

$$z_t = p_t - s_t - p_t^*$$

is $I(0)$, even though the individual elements $\mathbf{y}_t = (p_t\ s_t\ p_t^*)'$ are all $I(1)$. In this case, we have a single cointegrating vector $\mathbf{a} = (1 \quad -1 \quad -1)'$. The term $z_t = \mathbf{a}'\mathbf{y}_t$ is interpreted as the *equilibrium error*; although it is not always zero, but it can not be apart from zero too often and too far to make sense the equilibrium concept. ■

2 Granger Representation Theorem

Let each elements of the $(k \times 1)$ vector, $\mathbf{y}_t$ is $I(1)$ with the $(k \times h)$ cointegrating matrix, $\mathbf{A}$, such that each elements of $\mathbf{A}'\mathbf{y}_t$ is $I(0)$. Then Granger (1983) have the following fundamental results when $\mathbf{y}_t$ are cointegrated.

2.1 Implication of Cointegration For the VMA Representation

We now discuss the general implications of cointegration for the moving average representation of a vector system. Since it is assumed the $\triangle\mathbf{y}_t$ is $I(0)$, let $\boldsymbol{\delta} \equiv E(\triangle\mathbf{y}_t)$, and define

$$\mathbf{u}_t = \triangle\mathbf{y}_t - \boldsymbol{\delta}. \tag{23-1}$$

Suppose that $\mathbf{u}_t$ has the Wold representation:[2]

$$\mathbf{u}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2\boldsymbol{\varepsilon}_{t-1} + ... = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t,$$

where $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and

$$E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_\tau) = \begin{cases} \boldsymbol{\Omega} & for\ t = \tau \\ \mathbf{0} & otherwise. \end{cases}$$

Let $\boldsymbol{\Psi}(1)$ denotes the $(k \times k)$ matrix polynomial $\boldsymbol{\Psi}(z)$ evaluated at $z = 1$; that is,

$$\boldsymbol{\Psi}(1) \equiv \mathbf{I}_k + \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2 + \boldsymbol{\Psi}_3 + ...,$$

Then the following holds.
(a). $\mathbf{A}'\boldsymbol{\Psi}(1) = \mathbf{0}$,
(b). $\mathbf{A}'\boldsymbol{\delta} = \mathbf{0}$. ■

To verify this claim, note that as long as $\{s\boldsymbol{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable, the difference equation (23-1) implies that (from multivariate B-N decomposition):

$$\begin{aligned} \mathbf{y}_t &= \mathbf{y}_0 + \boldsymbol{\delta} \cdot t + \mathbf{u}_1 + \mathbf{u}_2 + ... + \mathbf{u}_t \\ &= \mathbf{y}_0 + \boldsymbol{\delta} \cdot t + \boldsymbol{\Psi}(1) \cdot (\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + ... + \boldsymbol{\varepsilon}_t) + \boldsymbol{\eta}_t - \boldsymbol{\eta}_0, \end{aligned} \tag{23-2}$$

[2] Recalling from Chapter 15 that Wold's decomposition is important for us because it provides an explanation of the sense in which $ARMA$ model (stochastic difference equation) provide a general model for the indeterministic part of any stationary stochastic process.

where $\boldsymbol{\eta}_t$ is a stationary process. Premultiplying (23-2) by $\mathbf{A}'$ results in

$$\mathbf{A}'\mathbf{y}_t = \mathbf{A}'(\mathbf{y}_0 - \boldsymbol{\eta}_0) + \mathbf{A}'\boldsymbol{\delta}\cdot t + \mathbf{A}'\boldsymbol{\Psi}(1)\cdot(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + ... + \boldsymbol{\varepsilon}_t) + \mathbf{A}'\boldsymbol{\eta}_t \sim I(0). \tag{23-3}$$

If $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')$ is nonsingular, then $\mathbf{c}'(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + ... + \boldsymbol{\varepsilon}_t)$ is $I(1)$ for every nonzero $(k \times 1)$ vector $\mathbf{c}$. Moreover, if some of the series exhibit nonzero drift ($\boldsymbol{\delta} \neq \mathbf{0}$), the linear combination $\mathbf{A}'\mathbf{y}_t$ will grow deterministically at rate $\mathbf{A}'\boldsymbol{\delta}$. Thus if the underlying hypothesis suggesting the possibility of cointegration is that certain linear combination of $\mathbf{y}_t$ are $I(0)$, this require that both conditions that $\mathbf{A}'\boldsymbol{\Psi}(1) = \mathbf{0}$ and $\mathbf{A}'\boldsymbol{\delta} = \mathbf{0}$ hold.

The second condition means that despite the presence of a drift term in the process generating $\mathbf{y}_t$, there is **no linear trend** in the cointegrated combination.[3] To the implication of the first condition, from partitioned matrix production we have

$$\mathbf{A}'\boldsymbol{\Psi}(1) = \begin{bmatrix} \mathbf{a}'_{\mathbf{1}(1\times k)} \\ \mathbf{a}'_{\mathbf{2}(1\times k)} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_{\mathbf{h}(1\times k)} \end{bmatrix} \cdot \boldsymbol{\Psi}(1)_{(k\times k)} = \begin{bmatrix} \mathbf{a}'_{\mathbf{1}}\boldsymbol{\Psi}(1) \\ \mathbf{a}'_{\mathbf{2}}\boldsymbol{\Psi}(1) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_{\mathbf{h}}\boldsymbol{\Psi}(1) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{0} \end{bmatrix},$$

which implies

$$\mathbf{a}'_i\boldsymbol{\Psi}(1) = \begin{bmatrix} a_{1i} & a_{2i} & . & . & . & a_{ki} \end{bmatrix} \begin{bmatrix} \boldsymbol{\psi}(1)'_{1(1\times k)} \\ \boldsymbol{\psi}(1)'_{2(1\times k)} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{\psi}(1)'_{k(1\times k)} \end{bmatrix} = \sum_{s=1}^{k} a_{si}\boldsymbol{\psi}(1)'_s = \mathbf{0}_{(1\times k)}$$

$$for\ i = 1, 2, ..., k, \tag{23-4}$$

where a_{si} is the sth elements of the row vector $\mathbf{a}'_i$ and $\boldsymbol{\psi}(1)'_i$ is the i th row of the matrix $\boldsymbol{\Psi}(1)$

Equation (23-4) implies that certain linear combination of the rows (columns ?) of $\boldsymbol{\Psi}(1)$ are zero, meaning that the row vector of $\boldsymbol{\Psi}(1)$ are linearly dependent. That is, $\boldsymbol{\Psi}(1)$ is a singular matrix, or equivalently, the determinant of $\boldsymbol{\Psi}(1)$ are zero, i.e. $|\boldsymbol{\Psi}(1)| = 0$, that is $|\boldsymbol{\Psi}(z)| = 0$ has a root $z = 1$.[4] This in turn means that the

[3] See Banerjee et.al (1993) p. 151 for details.

[4] Recall from Theorem 4 on page 7 of Chapter 22, this condition violate the proof of spurious regression

matrix operator $\mathbf{\Psi}(L)$ is non-invertible.[5] Thus, *a cointegrated system can never be represented by a finite-order vector autoregression in the differenced data* $\triangle \mathbf{y}_t$ from the non-invertibility of $\mathbf{\Psi}(L)$ of the following equations:

$$\triangle \mathbf{y}_t = \boldsymbol{\delta} + \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t.$$

2.2 Implication of Cointegration For the VAR Representation

Suppose that the level of $\mathbf{y}_t$ can be represented as a non-stationary pth-order vector autoregression:[6]

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + ... + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{23-5}$$

or

$$\mathbf{\Phi}(L)\mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t. \tag{23-6}$$

where

$$\mathbf{\Phi}(L) \equiv [\mathbf{I}_k - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - ... - \mathbf{\Phi}_p L^p].$$

Suppose that $\triangle \mathbf{y}_t$ has the Wold representation

$$(1-L)\mathbf{y}_t = \boldsymbol{\delta} + \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t. \tag{23-7}$$

Premultiplying (23-7) by $\mathbf{\Phi}(L)$ results in

$$(1-L)\mathbf{\Phi}(L)\mathbf{y}_t = \mathbf{\Phi}(1)\boldsymbol{\delta} + \mathbf{\Phi}(L)\mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t. \tag{23-8}$$

Substituting (23-6) into (23-8), we have

$$(1-L)\boldsymbol{\varepsilon}_t = \mathbf{\Phi}(1)\boldsymbol{\delta} + \mathbf{\Phi}(L)\mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t, \tag{23-9}$$

[5] If the determinant of an $(n \times n)$ matrix $\mathbf{H}$ is not equal zero, its inverse is found by dividing the adjoint by the determinant: $\mathbf{H}^{-1} = (1/|\mathbf{H}|) \cdot [(-1)^{i+j}|\mathbf{H}_{ji}|]$.

[6] The is not the only model for $I(1)$. See Saikkonen and Luukkonen (1997) infinite VAR and ? VARMA model

since $(1-L)\mathbf{c} = \mathbf{0}$. Now, equation (23-9) has to hold for all realizations of $\boldsymbol{\varepsilon}_t$, which require that

$$\boldsymbol{\Phi}(1)\boldsymbol{\delta} = \mathbf{0} \quad (a\ vector) \tag{23-10}$$

and that $(1-L)\mathbf{I}_k$ and $\boldsymbol{\Phi}(L)\boldsymbol{\Psi}(L)$ represent the identical polynomials in L. In particular, for $L = 1$, equation (23-9) implies that

$$\boldsymbol{\Phi}(1)\boldsymbol{\Psi}(1) = \mathbf{0}. \quad (a\ matrix) \tag{23-11}$$

Let $\boldsymbol{\phi}'_i$ denote ith row of $\boldsymbol{\Phi}(1)$. Then (23-10) and (23-11) state that $\boldsymbol{\phi}'_i\boldsymbol{\Psi}(1) = \mathbf{0}'$ (a row of zero) and $\boldsymbol{\phi}'_i\boldsymbol{\delta} = 0$ (a zero scalar). Recalling conditions (a) and (b)of section 2.1, this mean that $\boldsymbol{\phi}_i$ is a cointegrating vector. If $\mathbf{a}_1$, $\mathbf{a}_2$,..., $\mathbf{a}_h$ form a basis for the space of cointegrating vectors, then it must be possible to express $\boldsymbol{\phi}_i$ as a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$,..., $\mathbf{a}_h$–that is, there exist an $(h \times 1)$ vector $\mathbf{b}_i$ such that

$$\boldsymbol{\phi}_i = [\mathbf{a}_1\ \mathbf{a}_2\ ....\ \mathbf{a}_h]\mathbf{b}_i = \mathbf{A}\mathbf{b}_i = \mathbf{a}_1 b_{i1} + ... + \mathbf{a}_h b_{ih} \ \ (linear\ combination\ of\ \mathbf{a}_j, j = 1, ..., h.)$$

or that

$$\boldsymbol{\phi}'_i = \mathbf{b}'_i\mathbf{A}'$$

for $\mathbf{A}'$ the $(h \times k)$ matrix whose ith row is $\mathbf{a}'_i$. Applying this reasoning to each of the rows of $\boldsymbol{\Phi}(1)$, i.e.

$$\boldsymbol{\Phi}(1) = \begin{bmatrix} \boldsymbol{\phi}'_1 \\ \boldsymbol{\phi}'_2 \\ . \\ . \\ . \\ \boldsymbol{\phi}'_k \end{bmatrix} = \begin{bmatrix} \mathbf{b}'_1\mathbf{A}' \\ \mathbf{b}'_2\mathbf{A}' \\ . \\ . \\ . \\ \mathbf{b}'_k\mathbf{A}' \end{bmatrix} = \mathbf{B}\mathbf{A}', \tag{23-12}$$

where $\mathbf{B}$ is an $k \times h$ matrix. However, it is seen that the matrix $\mathbf{A}$ and $\mathbf{B}$ is not identified since for any choice of $h \times h$ matrix $\boldsymbol{\Upsilon}$, the matrix $\boldsymbol{\Phi}(1) = \mathbf{B}\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Upsilon}\mathbf{A}' = \mathbf{B}^*\mathbf{A}^{*\prime}$ implies the same distribution with $\boldsymbol{\Phi}(1) = \mathbf{B}\mathbf{A}'$. What can be determined is the space spanned by $\mathbf{A}$ the cointegrating space which need the concept of the *basis*.

Note that (23-12) implies that the $k \times k$ matrix $\boldsymbol{\Phi}(1)$ is a singular matrix because

$$rank(\boldsymbol{\Phi}(1)) = rank(\mathbf{B}\mathbf{A}') \leq \min(rank(\mathbf{B}), rank(\mathbf{A}')) = h < k.$$

2.3 Vector Error Correction Representation

A final representation for a cointegrated system is obtained by recalling from equation (22-1) of Chapter 22 that any VAR (not necessary cointegrated at this stage) in the form of (22-5) can be equivalently be written as

$$\triangle \mathbf{y}_t = \boldsymbol{\xi}_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{\xi}_2 \triangle \mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1} \mathbf{y}_{t-p+1} + \mathbf{c} + \boldsymbol{\xi}_0 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{23-13}$$

where

$$\boldsymbol{\xi}_0 \equiv \boldsymbol{\rho} - \mathbf{I} = -(\mathbf{I} - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - ... - \boldsymbol{\Phi}_p) = -\boldsymbol{\Phi}(1).$$

Note that if $\mathbf{y}_t$ has h cointegrating relations, then substitution of (23-12) into (23-13) results in

$$\triangle \mathbf{y}_t = \boldsymbol{\xi}_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{\xi}_2 \triangle \mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1} \triangle \mathbf{y}_{t-p+1} + \mathbf{c} - \mathbf{B}\mathbf{A}' \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{23-14}$$

Denote $\mathbf{z}_t \equiv \mathbf{A}'\mathbf{y}_t$, noticing that $\mathbf{z}_t$ is a stationary $(h \times 1)$ vector. Then (23-14) can be written as

$$\triangle \mathbf{y}_t = \boldsymbol{\xi}_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{\xi}_2 \triangle \mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1} \triangle \mathbf{y}_{t-p+1} + \mathbf{c} - \mathbf{B}\mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{23-15}$$

Expression (23-15) is known as the **vector error-correction** representation of the cointegrated system. It is interesting to see that while a cointegrated system can never be represented by a finite-order vector autoregression in the differenced data $\triangle \mathbf{y}_t$, it has a vector error correction representation; the difference is in that the former has ignored the error correction term, $-\mathbf{B}\mathbf{z}_{t-1}$.

Example.

Let the individual elements $(p_t \ \ s_t \ \ p_t^*)'$ are all $I(1)$ and have a single cointegrating vector $\mathbf{a} = (1 \quad -1 \quad -1)'$ among them. Then these three variables has a $VECM$ representation:

$$\begin{bmatrix} \triangle p_t \\ \triangle s_t \\ \triangle p_t^* \end{bmatrix} = \begin{bmatrix} \xi_{11}^{(1)} & \xi_{12}^{(1)} & \xi_{13}^{(1)} \\ \xi_{21}^{(1)} & \xi_{22}^{(1)} & \xi_{23}^{(1)} \\ \xi_{31}^{(1)} & \xi_{32}^{(1)} & \xi_{33}^{(1)} \end{bmatrix} \begin{bmatrix} \triangle p_{t-1} \\ \triangle s_{t-1} \\ \triangle p_{t-1}^* \end{bmatrix} + \begin{bmatrix} \xi_{11}^{(2)} & \xi_{12}^{(2)} & \xi_{13}^{(2)} \\ \xi_{21}^{(2)} & \xi_{22}^{(2)} & \xi_{23}^{(2)} \\ \xi_{31}^{(2)} & \xi_{32}^{(2)} & \xi_{33}^{(2)} \end{bmatrix} \begin{bmatrix} \triangle p_{t-2} \\ \triangle s_{t-2} \\ \triangle p_{t-2}^* \end{bmatrix} + ...$$
$$+ \begin{bmatrix} \xi_{11}^{(p-1)} & \xi_{12}^{(p-1)} & \xi_{13}^{(p-1)} \\ \xi_{21}^{(p-1)} & \xi_{22}^{(p-1)} & \xi_{23}^{(p-1)} \\ \xi_{31}^{(p-1)} & \xi_{32}^{(p-1)} & \xi_{33}^{(p-1)} \end{bmatrix} \begin{bmatrix} \triangle p_{t-p+1} \\ \triangle s_{t-p+1} \\ \triangle p_{t-p+1}^* \end{bmatrix}$$
$$+ \begin{bmatrix} c_p \\ c_s \\ c_{p^*} \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ s_{t-1} \\ p_{t-1}^* \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(p)} \\ \varepsilon_t^{(s)} \\ \varepsilon_t^{(p*)} \end{bmatrix},$$

from which we see that the dynamics of changes in each variable is not only according to the lags of its own and other variable's change but also to the **levels of each of the elements** of $\mathbf{z}_{t-1}$ by the speed $\mathbf{B}$:

$$\begin{aligned}\triangle p_t &= \xi_{11}^{(1)}\triangle p_{t-1}+\xi_{12}^{(1)}\triangle s_{t-1}+\xi_{13}^{(1)}\triangle p^*_{t-1}+\xi_{11}^{(2)}\triangle p_{t-2}+\xi_{12}^{(2)}\triangle s_{t-2}+\xi_{13}^{(2)}\triangle p^*_{t-2}\\ &\quad +...+\xi_{11}^{(p-1)}\triangle p_{t-p+1}+\xi_{12}^{(p-1)}\triangle s_{t-p+1}+\xi_{13}^{(p-1)}\triangle p^*_{t-p+1}+c_p\\ &\quad -b_1(p_{t-1}-s_{t-1}-p^*_{t-1})+\varepsilon_t^{(p)}\\ &= \xi_{11}^{(1)}\triangle p_{t-1}+\xi_{12}^{(1)}\triangle s_{t-1}+\xi_{13}^{(1)}\triangle p^*_{t-1}+\xi_{11}^{(2)}\triangle p_{t-2}+\xi_{12}^{(2)}\triangle s_{t-2}+\xi_{13}^{(2)}\triangle p^*_{t-2}\\ &\quad +...+\xi_{11}^{(p-1)}\triangle p_{t-p+1}+\xi_{12}^{(p-1)}\triangle s_{t-p+1}+\xi_{13}^{(p-1)}\triangle p^*_{t-p+1}+c_p\\ &\quad -b_1 z_{t-1}+\varepsilon_t^{(p)}.\end{aligned}$$

■

From economics equilibrium, when there is a positive equilibrium error happen in previous period, i.e. $z_{t-1} = p_{t-1} - s_t - p^*_{t-1} > 0$, at time t, the changes in p_t, i.e. $\triangle p_t = p_t - p_{t-1}$ should be negatively related with this equilibrium error. Therefore, the parameters of equilibrium error adjustment should be positive in (23-15).

3 Johansen's Granger Representation Theorem

Hohansen (1991) analysis the VAR model for cointegration with Gaussian error and constant. Consider a general k-dimensional VAR model with Gaussian error written in the error correction form:

$$\triangle \mathbf{y}_t = \boldsymbol{\xi}_1 \triangle \mathbf{y}_{t-1} + \boldsymbol{\xi}_2 \triangle \mathbf{y}_{t-2} + ... + \boldsymbol{\xi}_{p-1} \triangle \mathbf{y}_{t-p+1} + \mathbf{c} + \boldsymbol{\xi}_0 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{23-16}$$

where

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t) &= \mathbf{0} \\ E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') &= \begin{cases} \boldsymbol{\Omega} & for\ t = s \\ \mathbf{0} & otherwise. \end{cases} \end{aligned}$$

The model defined by (23-16) is rewritten as

$$\boldsymbol{\xi}(L)\mathbf{y}_t = -\boldsymbol{\xi}_0 \mathbf{y}_t + \boldsymbol{C}(L)\triangle \mathbf{y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t,$$

where

$$\boldsymbol{\xi}(L) = (1-L)\mathbf{I} - \sum_{i=1}^{p-1} \boldsymbol{\xi}_i (1-L) L^i - \boldsymbol{\xi}_0 L^1 \tag{23-17}$$

and

$$\boldsymbol{C}(L) = (\boldsymbol{\xi}(L) - \boldsymbol{\xi}(1))/(1-L) = \mathbf{I} - \sum_{i=1}^{p-1} \boldsymbol{\xi}_i L^i. \tag{23-18}$$

Note that

$$\begin{aligned} -\boldsymbol{\xi}_0 \mathbf{y}_t + \boldsymbol{C}(L)\triangle \mathbf{y}_t &= -\boldsymbol{\xi}_0 \mathbf{y}_t + \boldsymbol{\xi}(L)\mathbf{y}_t - \boldsymbol{\xi}(1)\mathbf{y}_t \\ &= -\boldsymbol{\xi}_0 \mathbf{y}_t + \boldsymbol{\xi}(L)\mathbf{y}_t + \boldsymbol{\xi}_0 \mathbf{y}_t \\ &= \boldsymbol{\xi}(L)\mathbf{y}_t \end{aligned}$$

from the fact in (23-17) that $\boldsymbol{\xi}(1) = -\boldsymbol{\xi}_0$.

Johansen (1991) provide the following fundamental result about error correction models of order 1 and their structure. The basic results is due to Granger (1983) and Engle and Granger (1987). In addition he provide an explicit condition for the process to be integrated of order 1 and he clarify the role of the constant term.

Theorem 2 (Granger's Representation Theorem):
Let the process $\mathbf{y}_t$ satisfy the equation (17) for $t = 1, 2, ...,$ and let

$$\boldsymbol{\xi}_0 = -\mathbf{B}\mathbf{A}'$$

for $\mathbf{A}$ and $\mathbf{B}$ of dimension $k \times h$ and rank h,[7] and let

$$\mathbf{B}'_{\perp}\boldsymbol{C}(1)\mathbf{A}_{\perp}$$

have full rank $k-h$. We define

$$\boldsymbol{\Psi} = \mathbf{A}_{\perp}(\mathbf{B}'_{\perp}\boldsymbol{C}(1)\mathbf{A}_{\perp})^{-1}\mathbf{B}'_{\perp}.$$

Then $\triangle\mathbf{y}_t$ and $\mathbf{A}'\mathbf{y}_t$ can be given initial distributions, such that
(a) $\triangle\mathbf{y}_t$ is stationary,
(b) $\mathbf{A}'\mathbf{y}_t$ is stationary,
(c) $\mathbf{y}_t$ is nonstationary, with linear trend $\boldsymbol{\tau}t = \boldsymbol{\Psi}\mathbf{c}t$.

Further
(d) $E(\mathbf{A}'\mathbf{y}_t) = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{c} + (\mathbf{B}'\mathbf{B})^{-1}(\mathbf{B}'\boldsymbol{C}(1)\mathbf{A}_{\perp})(\mathbf{B}'_{\perp}\boldsymbol{C}(1)\mathbf{A}_{\perp})^{-1}\mathbf{B}'_{\perp}\mathbf{c}$,
(e) $\mathrm{E}(\triangle\mathbf{y}_t) = \boldsymbol{\tau}$.

If $\mathbf{B}'_{\perp}\mathbf{c} = \mathbf{0}$, then $\boldsymbol{\tau} = \mathbf{0}$ and the linear trend disappears. However, the cointegrating vector still contain a constant term, i.e. $E(\mathbf{A}'\mathbf{y}_t) = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{c}$, when $\mathbf{B}'_{\perp}\mathbf{c} = \mathbf{0}$.

(f) $\triangle\mathbf{y}_t = \boldsymbol{\Psi}(L)(\boldsymbol{\varepsilon}_t + \mathbf{c})$
with $\boldsymbol{\Psi}(1) = \boldsymbol{\Psi}$. For $\boldsymbol{\Psi}_1(L) = (\boldsymbol{\Psi}(L) - \boldsymbol{\Psi}(1))/(1-L)$ so that $\boldsymbol{\Psi}(L) = \boldsymbol{\Psi}(1) + (1-L)\boldsymbol{\Psi}_1(L)$, the process has the representation

(g) $\mathbf{y}_t = \mathbf{y}_0 + \boldsymbol{\Psi}\sum_{i=1}^{T}\boldsymbol{\varepsilon}_i + \boldsymbol{\tau}t + \mathbf{S}_t - \mathbf{S}_0$,
where $\mathbf{S}_t = \boldsymbol{\Psi}_1(L)\boldsymbol{\varepsilon}_t$.

Proof:
See Johansen (1991), p.1559.

$$\boldsymbol{\Delta}\mathbf{y_t} = \xi_{\mathbf{1}} \triangle \mathbf{y_{t-1}} + \xi_{\mathbf{2}} \triangle \mathbf{y_{t-2}} + \ldots + \xi_{\mathbf{p-1}}\mathbf{y_{t-p+1}} + \mathbf{c} + \xi_{\mathbf{0}}\mathbf{y_{t-1}} + \varepsilon_{\mathbf{t}}$$

The model is rewritten as

$$\xi(\mathbf{L})\mathbf{y_t} \;=\; -\xi_{\mathbf{0}}\mathbf{y_t} + \mathbf{C}(\mathbf{L})\triangle\mathbf{y_t} \;=\; \mathbf{c} + \varepsilon_{\mathbf{t}}$$

[7] Define the orthogonal complements $\mathbf{P}_{\perp}$ of any matrix $\mathbf{P}$ of rank q and dimension $n \times q$ as follows ($0 < q < n$):
(a) $\mathbf{P}_{\perp}$ is of dimension $n \times (n-q)$;
(b) $\mathbf{P}'_{\perp}\mathbf{P} = \mathbf{0}_{(n-q)\times q}$, $\mathbf{P}'\mathbf{P}_{\perp} = \mathbf{0}_{q\times(n-q)}$;
(c) $\mathbf{P}_{\perp}$ has rank $n-q$, and its column space lies in the null space of $\mathbf{P}'$.

where

$$\xi(\mathbf{L}) = (\mathbf{1}-\mathbf{L})\mathbf{I} - \sum_{\mathbf{i=1}}^{\mathbf{p-1}} \xi_{\mathbf{i}}(\mathbf{1}-\mathbf{L})\mathbf{L}^{\mathbf{i}} - \xi_{\mathbf{0}}\mathbf{L}^{\mathbf{1}}$$

$$(\xi(\mathbf{1}) = -\xi_{\mathbf{0}})$$

and

$$\mathbf{C}(\mathbf{L}) = (\xi(\mathbf{L}) - \xi(\mathbf{1}))/(\mathbf{1}-\mathbf{L}) = \mathbf{I} - \sum_{\mathbf{i=1}}^{\mathbf{p-1}} \xi_{\mathbf{i}}\mathbf{L}^{\mathbf{i}}$$

Theorem 2 (Granger's Representation Theorem)
Let the process y_t satisfy the equation

$$\triangle\mathbf{y_t} = \xi_{\mathbf{1}} \triangle \mathbf{y_{t-1}} + \xi_{\mathbf{2}} \triangle \mathbf{y_{t-2}} + \ldots + \xi_{\mathbf{p-1}}\mathbf{y_{t-p+1}} + \mathbf{c} + \xi_{\mathbf{0}}\mathbf{y_{t-1}} + \varepsilon_{\mathbf{t}}$$

and let

$$\xi_{\mathbf{0}} = -\mathbf{B}\mathbf{A}'$$

and let

$$\mathbf{B}'_{\perp}\mathbf{C}(\mathbf{1})\mathbf{A}_{\perp} \quad \textbf{have full rank}(\mathbf{k}-\mathbf{h})$$

and define

$$\mathbf{\Psi} = \mathbf{A}_{\perp}(\mathbf{B}'_{\perp}\mathbf{C}(\mathbf{1})\mathbf{A}_{\perp})^{-\mathbf{1}}\mathbf{B}'_{\perp}$$

Then $\triangle y_t$ and $\mathbf{A}'\mathbf{y_t}$ can be gien initial distributions, such that

(a). $\triangle\mathbf{y_t}$ is stationary

(b). $\mathbf{A}'\mathbf{y_t}$ is stationary

(c). $\mathbf{y_t}$ is nonstationary, with linear trend $\tau\mathbf{t} = \mathbf{\Psi}\mathbf{c}\mathbf{t}$

(d). $\mathrm{E}(\mathbf{A}'\mathbf{y_t}) = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}\mathbf{B}'\mathbf{c} + (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}(\mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_{\perp})(\mathbf{B}'_{\perp}\mathbf{C}(\mathbf{1})\mathbf{A}_{\perp})^{-\mathbf{1}}\mathbf{B}'_{\perp}\mathbf{c}$

(e). $\mathrm{E}(\triangle\mathbf{y_t}) = \tau$

(f). $\triangle\mathbf{y_t} = \mathbf{\Psi}(\mathbf{L})(\varepsilon_{\mathbf{t}} + \mathbf{c})$

(g). $\mathbf{y_t} = \mathbf{y_0} + \mathbf{\Psi}\sum_{\mathbf{i=1}}^{\mathbf{T}} \varepsilon_{\mathbf{i}} + \tau\mathbf{t} + \mathbf{S_t} - \mathbf{S_0}$

Proof:
Rewritten the equation

$$\xi(\mathbf{L})\mathbf{y_t} \;=\; \mathbf{BA'y_t} + \mathbf{C(L)}\triangle\mathbf{y_t} \;=\; \mathbf{c} + \varepsilon_\mathbf{t}$$

multiplier $\mathbf{B'}$ and $\mathbf{B'_\perp}$,we

$$\mathbf{B'BA'y_t} + \mathbf{B'C(L)}\triangle\mathbf{y_t} = \mathbf{B'}(\mathbf{c} + \varepsilon_\mathbf{t}) \qquad (\mathbf{a})$$

$$\mathbf{B'_\perp BA'y_t} + \mathbf{B'_\perp C(L)}\triangle\mathbf{y_t} = \mathbf{B'_\perp}(\mathbf{c} + \varepsilon_\mathbf{t}) \qquad (\mathbf{b})$$

because $\mathbf{B'_\perp B} = \mathbf{0}$, (b) will become $\mathbf{B'_\perp C(L)}\triangle\mathbf{y_t} = \mathbf{B'_\perp}(\mathbf{c} + \varepsilon_\mathbf{t})$

However, ξ is a non-inverse singular matrix

Here, we introduce two new variables:

$$\mathbf{z_t} = (\mathbf{A'A})^{-1}\mathbf{A'y_t}$$

$$\mathbf{x_t} = (\mathbf{A'_\perp A_\perp})^{-1}\mathbf{A'_\perp}\triangle\mathbf{y_t}$$

The process $\triangle\mathbf{y_t}$ can be recovered from $\mathbf{z_t}$ and $\mathbf{x_t}$

$$\triangle\mathbf{y_t} = (\mathbf{A_\perp \bar{A}_\perp}' + \mathbf{A\bar{A}'})\triangle\mathbf{y_t} = \mathbf{A_\perp x_t} + \mathbf{A\Delta z_t}$$

we use the result into (a) and (b)

$$\mathbf{B'BA'Az_t} + \mathbf{BC(L)A_\perp x_t} + \mathbf{B'C(L)A\Delta z_t} = \mathbf{B'}(\mathbf{c} + \varepsilon_\mathbf{t})$$

$$\mathbf{B'_\perp C(L)A_\perp x_t} + \mathbf{B'_\perp C(L)A\Delta z_t} = \mathbf{B'_\perp}(\mathbf{c} + \varepsilon_\mathbf{t})$$

Rewritten

$$\tilde{\mathbf{H}}(\mathbf{L})(\mathbf{z'_t}, \mathbf{x'_t})' = (\mathbf{B}, \mathbf{B_\perp})'(\varepsilon_\mathbf{t} + \mathbf{c})$$

with

$$\tilde{\mathbf{H}}(\mathbf{z}) = \begin{bmatrix} \mathbf{B'BA'A} + \mathbf{B'C(z)A(1-z)} & \mathbf{B'C(z)A_\perp} \\ \mathbf{B'_\perp C(z)A(1-z)} & \mathbf{B'_\perp C(z)A_\perp} \end{bmatrix}$$

When z=1,

$$|\tilde{\mathbf{H}}(\mathbf{1})| = |\mathbf{B'B}||\mathbf{A'A}||\mathbf{B'_\perp C(L)A_\perp}| \neq \mathbf{0}$$

Hence z=1 is not a root

Proof:(a)&(b) When z≠1,

$$\tilde{\mathbf{H}}(\mathbf{z}) = (\mathbf{B}, \mathbf{B_\perp})'\xi(\mathbf{z})(\mathbf{A}, \mathbf{A_\perp}(\mathbf{1} - \mathbf{z})^{-1})$$

The determinant is

$$|\tilde{\mathbf{H}}(\mathbf{z})| = |(\mathbf{B}, \mathbf{B}_\perp)||\xi(\mathbf{z})||\mathbf{A}, \mathbf{A}_\perp|(\mathbf{1}-\mathbf{z})^{-(\rho-\mathbf{r})})$$

This shows that all roots of $|\tilde{\mathbf{H}}(\mathbf{z})| = \mathbf{0}$ are outside the unit cycle.
The following equations can be inverse:

$$\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{A}\mathbf{z_t} + \mathbf{B}\mathbf{C}(\mathbf{L})\mathbf{A}_\perp\mathbf{x_t} + \mathbf{B}'\mathbf{C}(\mathbf{L})\mathbf{A}\boldsymbol{\Delta}\mathbf{z_t} = \mathbf{B}'(\mathbf{c}+\varepsilon_\mathbf{t})$$

$$\mathbf{B}'_\perp\mathbf{C}(\mathbf{L})\mathbf{A}_\perp\mathbf{x_t} + \mathbf{B}'_\perp\mathbf{C}(\mathbf{L})\mathbf{A}\boldsymbol{\Delta}\mathbf{z_t} = \mathbf{B}'_\perp(\mathbf{c}+\varepsilon_\mathbf{t})$$

Therefore, z_t and x_t become stationary,and $\Delta y_t = A_\perp x_t + A\Delta z_t$ is stationary; this implies $\mathbf{A}'\mathbf{y_t}$ is stationary

Proof:(d)
we still use the same equations:

$$\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{y_t} + \mathbf{B}'\mathbf{C}(\mathbf{L})\triangle\mathbf{y_t} = \mathbf{B}'(\mathbf{c}+\varepsilon_\mathbf{t}) \qquad (\mathbf{a})$$

$$\mathbf{A}'\mathbf{y_t} = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}[\mathbf{B}'(\varepsilon_\mathbf{t}+\mathbf{c}) - \mathbf{B}'\mathbf{C}(\mathbf{L})\boldsymbol{\Delta}\mathbf{y_t}]$$

let $\triangle\mathbf{y_t} = \mathbf{A}_\perp\mathbf{x_t} + \mathbf{A}\boldsymbol{\Delta}\mathbf{z_t}$

$$\mathbf{A}'\mathbf{y_t} = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}[\mathbf{B}'(\varepsilon_\mathbf{t}+\mathbf{c}) - \mathbf{B}'\mathbf{C}(\mathbf{L})(\mathbf{A}_\perp\mathbf{x_t} + \mathbf{A}\boldsymbol{\Delta}\mathbf{z_t})]$$

let L=1

$$\mathbf{A}'\mathbf{y_t} = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}[\mathbf{B}'(\varepsilon_\mathbf{t}+\mathbf{c}) - \mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp\mathbf{x_t}]$$

$$\begin{bmatrix} \mathbf{z_t}' \\ \mathbf{x_t}' \end{bmatrix} = \tilde{\mathbf{H}}(\mathbf{1})^{-\mathbf{1}} \begin{bmatrix} \mathbf{B} \\ \mathbf{B}_\perp \end{bmatrix} (\mathbf{c}+\varepsilon)$$
$$= \begin{bmatrix} -\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{A} & \mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp \\ \mathbf{0} & \mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B} \\ \mathbf{B}_\perp \end{bmatrix} (\mathbf{c}+\varepsilon)$$

Hint:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1}(I + A_{12}F_2A_{21}A_{11}^{-1}) & -A_{11}^{-1}A_{12}F_2 \\ -F_2A_{21}A_{11}^{-1} & F_2 \end{bmatrix}$$

$$F_2 = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$$

$$\begin{bmatrix} \mathbf{z_t}' \\ \mathbf{x_t}' \end{bmatrix} = \tilde{\mathbf{H}}(\mathbf{1})^{-1} \begin{bmatrix} \mathbf{B} \\ \mathbf{B}_\perp \end{bmatrix} (\mathbf{c} + \varepsilon)$$
$$= \begin{bmatrix} -\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{A} & \mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp \\ \mathbf{0} & \mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{B} \\ \mathbf{B}_\perp \end{bmatrix} (\mathbf{c} + \varepsilon)$$
$$= \begin{bmatrix} (-\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{A})^{-\mathbf{1}} & W \\ \mathbf{0} & (\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}} \end{bmatrix} \begin{bmatrix} -\mathbf{B} \\ \mathbf{B}_\perp \end{bmatrix} (\mathbf{c} + \varepsilon)$$
$$(\quad W = -(-\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{A})^{-\mathbf{1}}\mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp(\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}} \quad)$$

We will get

$$\mathbf{x_t} = (\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'(\mathbf{c} + \varepsilon)$$

$$\mathbf{A}'\mathbf{y_t} = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}[\mathbf{B}'(\varepsilon_\mathbf{t} + \mathbf{c}) - \mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp\mathbf{x_t}]$$
$$\mathbf{x_t} = (\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'(\mathbf{c} + \varepsilon)$$
$$\mathbf{A}'\mathbf{y_t} = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}[\mathbf{B}'(\varepsilon_\mathbf{t} + \mathbf{c}) - \mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp(\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'(\mathbf{c} + \varepsilon)]$$
$$E(\mathbf{A}'\mathbf{y_t}) = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}[\mathbf{B}'\mathbf{c} - \mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp(\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'\mathbf{c}]$$
$$E(\mathbf{A}'\mathbf{y_t}) = (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}\mathbf{B}'\mathbf{c} - (\mathbf{B}'\mathbf{B})^{-\mathbf{1}}(\mathbf{B}'\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)(\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'\mathbf{c}$$

Proof: (e)

let $\triangle\mathbf{y_t} = \mathbf{A}_\perp\mathbf{x_t} + \mathbf{A}\boldsymbol{\Delta}\mathbf{z_t}$

$$E(\triangle\mathbf{y_t}) = \mathbf{A}_\perp\mathbf{E}(\mathbf{x_t}) + \mathbf{A}\mathbf{E}(\boldsymbol{\Delta}\mathbf{z_t})$$

let L=1

$$E(\triangle\mathbf{y_t}) = \mathbf{A}_\perp\mathbf{E}(\mathbf{x_t})$$
$$\mathbf{x_t} = (\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'(\mathbf{c} + \varepsilon)$$
$$E(\triangle\mathbf{y_t}) = \mathbf{A}_\perp(\mathbf{B}'_\perp\mathbf{C}(\mathbf{1})\mathbf{A}_\perp)^{-\mathbf{1}}\mathbf{B}_\perp{}'\mathbf{c}$$

$<$ *Question* $>$ Why we can't solve the equtions by $A'y_t$ & Δy_t directly?

$$\mathbf{B}'\mathbf{B}\mathbf{A}'\mathbf{y_t} + \mathbf{B}'\mathbf{C}(\mathbf{L})\triangle\mathbf{y_t} = \mathbf{B}'(\mathbf{c} + \varepsilon_\mathbf{t}) \qquad (\mathbf{a})$$

$$\mathbf{B}'_{\perp}\mathbf{B}\mathbf{A}'\mathbf{y_t} + \mathbf{B}'_{\perp}\mathbf{C}(\mathbf{L})\triangle\mathbf{y_t} = \mathbf{B}'_{\perp}(\mathbf{c}+\varepsilon_{\mathbf{t}}) \qquad (\mathbf{b})$$

$$\left[\begin{array}{cc} \mathbf{B}'\mathbf{B} & \mathbf{B}'\mathbf{C}(\mathbf{L}) \\ \mathbf{B}'_{\perp}\mathbf{B} & \mathbf{B}'_{\perp}\mathbf{C}(\mathbf{L}) \end{array}\right]\left[\begin{array}{c} \mathbf{A}'\mathbf{y_t} \\ \mathbf{\Delta y_t} \end{array}\right] = \left[\begin{array}{c} \mathbf{B}'(\mathbf{c}+\varepsilon) \\ \mathbf{B}'_{\perp}(\mathbf{c}+\varepsilon) \end{array}\right]$$

$$\mathbf{B}' \ \textbf{is} \ (\mathbf{h}\times\mathbf{k}) \qquad \mathbf{B}'_{\perp} \ \textbf{is} \ (\mathbf{k}-\mathbf{h})\times\mathbf{k} \qquad \mathbf{C}(\mathbf{L}) \ \textbf{is} \ (\mathbf{k}\times\mathbf{k})$$

So, we can't get the solution by this way. From the above, $\mathbf{B}'\mathbf{C}(\mathbf{L})$ and $\mathbf{B}'_{\perp}\mathbf{B}$ and $\mathbf{B}'_{\perp}\mathbf{C}(\mathbf{L})$ are all not square matrix; those are singular matrix, and the inverse matrix of them are non-existence.

4 Other representations for Cointegration

4.1 Phillips's Triangular Representation

Another convenient representation for a cointegrated system was introduced by Phillips (1991). Suppose that the rows of the $(h \times k)$ matrix $\mathbf{A}'$ form a basis for the space of the cointegrating vectors. By reordering and normalizing the cointegrating relations can be represented of the form

$$\mathbf{A}' = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{a}_h' \end{bmatrix} = \begin{bmatrix} 1 & 0 & . & . & . & 0 & -\gamma_{1,h+1} & -\gamma_{1,h+2} & . & . & . & -\gamma_{1,k} \\ 0 & 1 & . & . & . & 0 & -\gamma_{2,h+1} & -\gamma_{2,h+2} & . & . & . & -\gamma_{2,k} \\ . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 & -\gamma_{h,h+1} & -\gamma_{h,h+2} & . & . & . & -\gamma_{h,k} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{I}_h & -\mathbf{\Gamma}' \end{bmatrix},$$

where $\mathbf{\Gamma}'$ is an $(h \times g)$ matrix of coefficients for $g \equiv k - h$.

Let $\mathbf{z}_t$ denote the errors associated with the set of cointegrating relations:

$$\mathbf{z}_t \equiv \mathbf{A}'\mathbf{y}_t.$$

Since $\mathbf{z}_t$ is $I(0)$, then the mean $\boldsymbol{\mu}_1^* \equiv E(\mathbf{z}_t)$ exists, and we can define

$$\mathbf{z}_t^* \equiv \mathbf{z}_t - \boldsymbol{\mu}_1^*.$$

Partition $\mathbf{y}_t$ as

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1t(h\times 1)} \\ \mathbf{y}_{2t(g\times 1)} \end{bmatrix}. \quad (23\text{-}19)$$

Then

$$\mathbf{z}_t = \mathbf{z}_t^* + \boldsymbol{\mu}_1^* = \begin{bmatrix} \mathbf{I}_h & -\mathbf{\Gamma}' \end{bmatrix} \begin{bmatrix} \mathbf{y}_{1t(h\times 1)} \\ \mathbf{y}_{2t(g\times 1)} \end{bmatrix}$$

or

$$\mathbf{y}_{1t(h\times 1)} = \mathbf{\Gamma}'_{(h\times g)} \cdot \mathbf{y}_{2t(g\times 1)} + \mathbf{z}^*_{t\,(h\times 1)} + \boldsymbol{\mu}^*_{1(h\times 1)}. \quad (23\text{-}20)$$

A representation for $\mathbf{y}_{2t}$ is given by the last g rows of (23-1):

$$\triangle \mathbf{y}_{2t(g\times 1)} = \boldsymbol{\delta}_{2(g\times 1)} + \mathbf{u}_{2t(g\times 1)}, \quad (23\text{-}21)$$

where $\boldsymbol{\delta}_2$ and $\mathbf{u}_{2t}$ represent the last g elements of the $(k \times 1)$ vector $\boldsymbol{\delta}$ and $\mathbf{u}_t$ in (23-1), respectively. (23-20) and (23-21) constitute Phillips's (1991) triangular representation of a system with exactly h cointegrating relations. Note that $\mathbf{z}_t^*$ and $\mathbf{u}_{2t}$ represent zero-mean stationary disturbance in this representation.

Example.

Let the individual elements $(p_t \; s_t \; p_t^*)'$ are all $I(1)$ and have a single cointegrating vector $\mathbf{a} = (1 \;\; -1 \;\; -1)'$ among them. The triangular representation of these three variables are: given

$$\mathbf{A}' = \mathbf{a}' = [1 \;\; -\gamma_1 \;\; -\gamma_2],$$

then

$$\begin{aligned} p_t &= \gamma_1 s_t + \gamma_2 p_t^* + \mu_1^* + z_t^* \\ \triangle s_t &= \delta_s + u_{st} \\ \triangle p_t^* &= \delta_{p*} + u_{p*,t}, \end{aligned}$$

where the hypothesized values are $\gamma_1 = \gamma_2 = 1$. ■

4.2 The Stock-Watson's Common Trends Representation

Another useful representation for any cointegrated system was proposed by Stock and Watson (1988). Suppose that an $(k \times 1)$ vector $\mathbf{y}_t$ is characterized by exact h cointegrating relations with $g = k - h$. We have seen that it is possible to order the element of $\mathbf{y}_t$ in such a way that a triangular representation of the form of (23-20) and (23-21) exists with $(\mathbf{z}'^*_t, \mathbf{u}'_{2t})'$ a $I(0)$ $(k \times 1)$ vector with zero mean. Suppose that

$$\begin{bmatrix} \mathbf{z}'^*_t \\ \mathbf{u}'_{2t} \end{bmatrix} = \sum_{s=0}^{\infty} \begin{bmatrix} \mathbf{H}_s \boldsymbol{\varepsilon}_{t-s} \\ \mathbf{J}_s \boldsymbol{\varepsilon}_{t-s} \end{bmatrix}$$

for $\boldsymbol{\varepsilon}_t$ an $(k \times 1)$ white noise process with $\{s\mathbf{H}_s\}_{s=0}^{\infty}$ and $\{s\mathbf{J}_s\}_{s=0}^{\infty}$ absolutely summable sequences of $(h \times k)$ and $(g \times k)$ matrices, respectively. From B-N decomposition we have

$$\begin{aligned} \mathbf{y}_{2t} &= \mathbf{y}_{2,0} + \boldsymbol{\delta}_2 \cdot t + \sum_{s=1}^{t} \mathbf{u}_{2s} \\ &= \mathbf{y}_{2,0} + \boldsymbol{\delta}_2 \cdot t + \mathbf{J}(1) \cdot (\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + ... + \boldsymbol{\varepsilon}_t) + \boldsymbol{\eta}_{2t} - \boldsymbol{\eta}_{2,0}, \end{aligned} \quad (23\text{-}22)$$

where $\mathbf{J}(1) \equiv (\mathbf{J}_0 + \mathbf{J}_1 + \mathbf{J}_2 + ...)$, $\boldsymbol{\eta}_{2t} \equiv \sum_{s=0}^{\infty} \boldsymbol{\alpha}_{2s}\boldsymbol{\varepsilon}_{t-s}$, and $\boldsymbol{\alpha}_{2s} \equiv -(\mathbf{J}_{s+1} + \mathbf{J}_{s+2} + \mathbf{J}_{s+3} + ...)$.

Since the $(k \times 1)$ vector $\boldsymbol{\varepsilon}_t$ is white noise, the $(g \times 1)$ vector $\mathbf{J}(1)\boldsymbol{\varepsilon}_t$ is also white noise, implying that each element of the $(g \times 1)$ vector $\boldsymbol{\xi}_{2t}$ defined by

$$\boldsymbol{\xi}_{2t} = \mathbf{J}(1) \cdot (\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + ... + \boldsymbol{\varepsilon}_t) \tag{23-23}$$

is described by a random walk. Substituting (23-23) into (23-22) results in

$$\mathbf{y}_{2t} = \tilde{\boldsymbol{\mu}}_2 + \boldsymbol{\delta}_2 \cdot t + \boldsymbol{\xi}_{2t} + \boldsymbol{\eta}_{2t} \tag{23-24}$$

for $\tilde{\boldsymbol{\mu}}_2 = (\mathbf{y}_{2,0} - \boldsymbol{\eta}_{2,0})$.

Substituting (23-24) into (23-20) produces

$$\mathbf{y}_{1t} = \tilde{\boldsymbol{\mu}}_1 + \boldsymbol{\Gamma}'(\boldsymbol{\delta}_2 \cdot t + \boldsymbol{\xi}_t) + \tilde{\boldsymbol{\eta}}_{1t} \tag{23-25}$$

for $\tilde{\boldsymbol{\mu}}_1 = \boldsymbol{\mu}_1^* + \boldsymbol{\Gamma}'\tilde{\boldsymbol{\mu}}_2$ and $\tilde{\boldsymbol{\eta}}_{1t} = \mathbf{z}_t^* + \boldsymbol{\Gamma}'\tilde{\boldsymbol{\eta}}_{2t}$.

Equations (23-24) and (23-25) give Stock and Watson's (1988) common trends representation. These equations show that the vector $\mathbf{y}_t$ can be described as a stationary component,

$$\begin{bmatrix} \tilde{\boldsymbol{\mu}}_1 \\ \tilde{\boldsymbol{\mu}}_2 \end{bmatrix} + \begin{bmatrix} \tilde{\boldsymbol{\eta}}_{1t} \\ \boldsymbol{\eta}_{2t} \end{bmatrix},$$

plus linear combinations of up to g common deterministic trends, as described by the $(g \times 1)$ vector $\boldsymbol{\delta}_2 \cdot t$, the linear combination of the g **common stochastic trend** as described by the $(g \times 1)$ vector $\boldsymbol{\xi}_{2t}$. Therefore, when we say that a $k \times 1$ vector $\mathbf{y}_t$ is characterized by exactly h cointegrations, it is equivalent to say that there are $g(= k - h)$ common trends among them.

5 Estimation and Testing of Cointegration from Single Equation

5.1 Testing for Cointegration When the Cointegrating Vector is Known

Often when theoretical considerations suggest that certain variables will be cointegrated, or that $\mathbf{a}'\mathbf{y}_t$ is stationary for some $(k \times 1)$ cointegrating vector $\mathbf{a}$, the theory is based on a particular known value for $\mathbf{a}$. In the purchasing power parity example, $\mathbf{a} = (1 \ \ -1 \ \ -1)'$. Given the null hypothesis of unit root can not be rejected from various unit root tests on the individual series p_t, s_t, and p_t^*, the next step is to test whether their particular linear combination $z_t = \mathbf{a}'\mathbf{y}_t = p_t - s_t - p_t^*$ is stationary from various unit root tests. See the example on p.585 of Hamilton.

5.2 Testing the Null Hypothesis of No Cointegration, Residual-Based Tests for Cointegration

If the theoretical model of the system dynamic does not suggest a particular value for the cointegrating vector $\mathbf{a}$, then one approach is first to estimate $\mathbf{a}$ by OLS.

5.2.1 Estimating The Cointegrating Vector

If it is known for certain that the cointegrating vector has a nonzero coefficient for the first element of $\mathbf{y}_t$ ($a_1 \neq 0$), then a particularly convenient normalization is to set $a_1 = 1$ and represent subsequent entries of $\mathbf{a}$ $(a_2, a_3, ..., a_k)$ as the negative s of a set of unknown parameters $(\gamma_2, \gamma_3, ..., \gamma_k)$:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \\ a_k \end{bmatrix} = \begin{bmatrix} 1 \\ -\gamma_2 \\ -\gamma_3 \\ . \\ . \\ . \\ -\gamma_k \end{bmatrix},$$

where $\gamma_i = -a_i/a_1, \forall i = 2, ..., k$. Then consistent estimation of $\mathbf{a}$ is achieved by an OLS regression of the first element of $\mathbf{y}_t$ on all of the other:

$$Y_{1t} = \gamma_2 Y_{2t} + \gamma_3 Y_{3t} + ... + \gamma_k Y_{kt} + u_t. \tag{23-26}$$

Consistent estimates of $\gamma_2, \gamma_3, ..., \gamma_k$ are also obtained when a constant term is included in (27), as in

$$Y_{1t} = \alpha + \gamma_2 Y_{2t} + \gamma_3 Y_{3t} + ... + \gamma_k Y_{kt} + u_t.$$

or

$$Y_{1t} = \alpha + \boldsymbol{\gamma}' \mathbf{y}_{2t} + u_t,$$

where $\boldsymbol{\gamma}' = (\gamma_2, \gamma_3, ..., \gamma_k)$ and $\mathbf{y}_{2t} = (Y_{2t}, Y_{3t}, ..., Y_{kt})'$.

Theorem (Stock, 1986):
Let Y_{1t} be a scalar $\mathbf{y}_{2t}$ be a $(g \times 1)$ vector. Let $k \equiv g + 1$, and suppose that the $(k \times 1)$ vector $(Y_{1t}, \mathbf{y}'_{2t})'$ is characterized by exactly one cointegrating relation $(h = 1)$ that has a nonzero coefficients on y_{1t}. Let the triangular representation for the system be

$$Y_{1t} = \alpha + \boldsymbol{\gamma}' \mathbf{y}_{2t} + z_t^* \quad (23\text{-}27)$$

$$\triangle \mathbf{y}_{2t} = \mathbf{u}_{2t}. \quad (23\text{-}28)$$

Suppose that

$$\begin{bmatrix} z_t^* \\ \mathbf{u}_{2t} \end{bmatrix} = \boldsymbol{\Psi}^*(L)\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t$ is an $(k \times 1)$ $i.i.d.$ vector with mean zero, finite fourth moments, and positive definite variance-covariance matrix $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) = \mathbf{PP}'$. Suppose further that the sequence of $(k \times k)$ matrices $\{s \cdot \boldsymbol{\Psi}_s^*\}_{s=0}^{\infty}$ is absolutely summable and that the rows of $\boldsymbol{\Psi}^*(1)$ are linearly independent. Let $\hat{\alpha}_T$ and $\hat{\boldsymbol{\gamma}}_T$ be the OLS estimators of (28). Partition $\boldsymbol{\Psi}^*(1) \cdot \mathbf{P}$ as

$$\boldsymbol{\Psi}^*(1) \cdot \mathbf{P} = \begin{bmatrix} \boldsymbol{\lambda}_{1\,(1\times n)}^{*\prime} \\ \boldsymbol{\Lambda}_{2(g\times n)}^{*} \end{bmatrix}.$$

Then

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma}) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \{\int [\mathbf{W}(r)]' dr\} \cdot \boldsymbol{\Lambda}_2^{*\prime} \\ \boldsymbol{\Lambda}_2^* \cdot \int \mathbf{W}(r) dr & \boldsymbol{\Lambda}_2^* \{\int [\mathbf{W}(r)] \cdot [\mathbf{W}(r)]' dr\} \cdot \boldsymbol{\Lambda}_2^{*\prime} \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix} \quad (23\text{-}29)$$

where $\mathbf{W}(r)$ is k-dimensional standard Brownian motion, the integral sign denotes integration over r from 0 to 1, and

$$h_1 \equiv \boldsymbol{\lambda}_1^{*\prime} \cdot \mathbf{W}(1)$$

$$\mathbf{h}_2 \equiv \boldsymbol{\Lambda}_2^* \left\{ \int [\mathbf{W}(r)] \cdot [\mathbf{W}(r)]' dr \right\} \cdot \boldsymbol{\lambda}_1^* + \sum_{v=0}^{\infty} E(\mathbf{u}_{2t} z_{t+v}^*).$$

This theorem shows that the OLS estimator of the cointegrating vector is consistent. However, it is noted that the correlation between the regressors $\mathbf{y}_{2t}$ and the error z_t^* is not to induce inconsistency of $\hat{\boldsymbol{\gamma}}_T$; instead, the asymptotic distribution exhibits a bias since the distribution of $T(\hat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})$ does not centered around zero.[8]

In the next chapter we will consider system estimation of cointegrating vector. Banerjee et al. (1993, p.214) examined one of main reasons for using such an estimation: the large finite-sample biases that can arise static OLS estimates of cointegrating vectors or parameters. While such estimator are super-consistent (T-consistent), Monte Carlo experiments nonetheless suggest that a large number of observations may be necessary before the biases become small.

Example.

The following is the code to generate the spurious regression. Let $X_t = X_{t-1} + v_t$, and $Y_t = 2 + 3X_t + u_t$ where

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \overset{i.i.d}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} \right).$$

Consider the sequence of an OLS regression of Y_t on X_t and a constant

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \ t = 1, 2, ..., T.$$

It can see that the OLS estimates of $\hat{\alpha}$ and $\hat{\beta}$ is super-consistent and the t-ratio to test the null hypothesis that $\alpha = 0$ and $\beta = 0$, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$, ??is increasing with sample. We always incorrectly reject the null hypothesis.

(a). Plot super-consistency of $\hat{\alpha}$ and $\hat{\beta}$.

(b). Plot the t ratio, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$.

[8] See similar results at Phillips-Perron Test of Chapter 21.

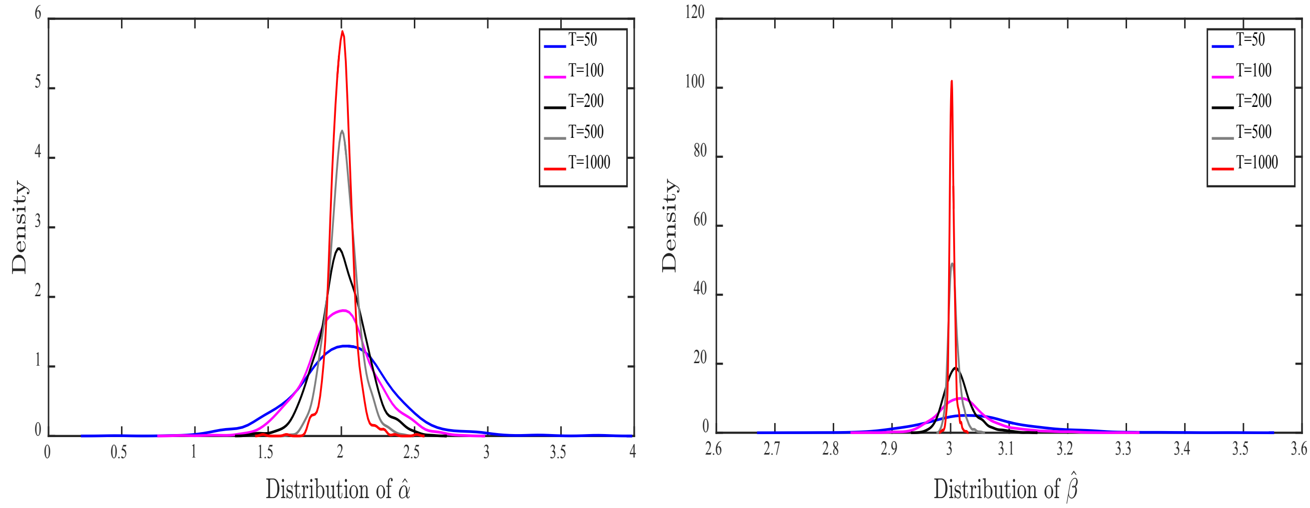


Figure (23-1a). Super-consistency of $\hat{\alpha}$ and $\hat{\beta}$ in a cointegrated regression.

■

5.3 The Role of Normalization

The OLS estimate of the cointegrating vector was obtained by normalizing the first element of the cointegrating vector $\mathbf{a}$ to be unity. The proposal was then to regress the first element of $\mathbf{y}_t$ on the others. For example, with $k = 2$, we would regression Y_{1t} on Y_{2t}:

$$Y_{1t} = \alpha + \gamma Y_{2t} + u_t.$$

Obviously, we might equally well have normalized $a_2 = 1$ and use the same argument to suggest a regression of Y_{2t} on Y_{1t}:

$$Y_{2t} = \theta + \xi Y_{1t} + v_t.$$

The OLS estimate of $\hat{\xi}$ is not simply the inverse of $\hat{\gamma}$, meaning that these two regressions will give different estimate of the cointegrating vector:

$$\begin{bmatrix} 1 \\ -\hat{\gamma} \end{bmatrix} \neq -\hat{\gamma} \begin{bmatrix} -\hat{\xi} \\ 1 \end{bmatrix}.$$

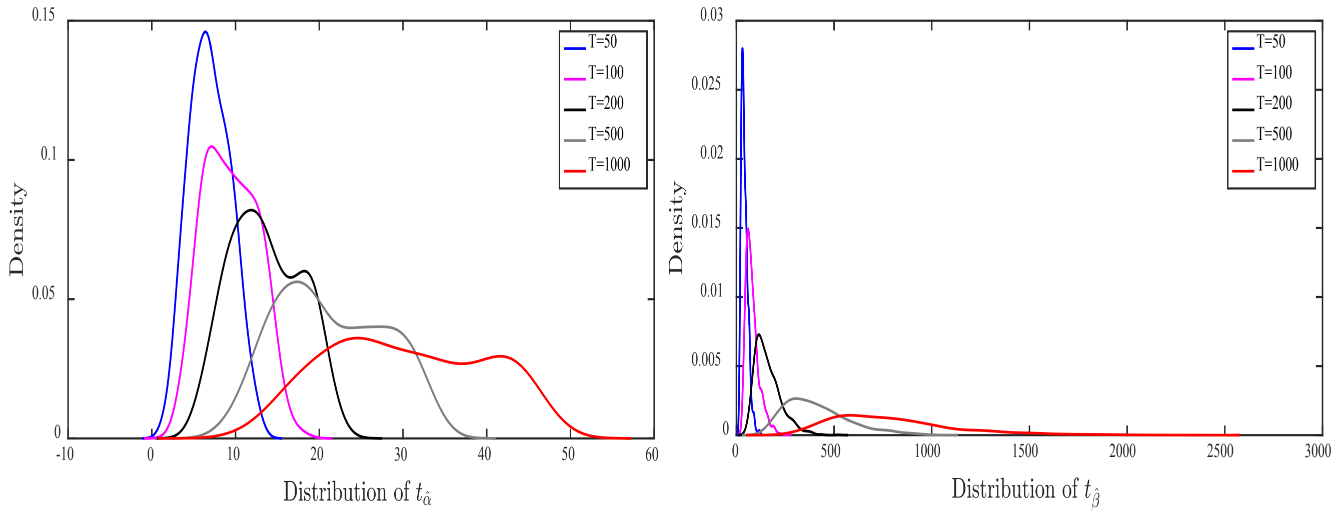


Figure (22-1b). The two t-ratios to test the null hypothesis that $\alpha = 0$ and $\beta = 0$, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$, are not increasing with sample size.

Thus, choosing which variable to call Y_{1t} and which to call Y_{2t} might end up making a material difference for the estimate of **a** as well as for the evidence one finds for cointegration among the series.

5.3.1 What Is the Regression Estimating When There Is More Than One Cointegrating Relation ?

The limiting distribution of the OLS estimation in Theorem 1 was derived under the assumption that there is only one cointegration ($h = 1$). In a more general case with $h > 1$, OLS estimate of (28) should still provide a consistent estimate of a cointegrating vector. But which cointegrating vector is it ? Wooldridge (1991) show that among the set of possible cointegrating relations, OLS estimation of (28) select the relation whose residuals are uncorrelated with any other $I(1)$ linear combination of $(Y_{2t}, Y_{3t}, ..., Y_{kt})$.

5.3.2 What Is the Regression Estimating When There Is No Cointegrating Relation ?

Let us now consider the properties of OLS estimation when there is no cointegrating relation. Then (28) is a regression of an $I(1)$ variables on a set of $(k-1)$ $I(1)$ variables for which no coefficients produce an $I(0)$ error term. The regression is therefore subject to the spurious regression problem described in Chapter 22. The coefficient $\hat{\alpha}$ and $\hat{\gamma}$ do not provide consistent estimate of any population parameter, and the OLS sample residual $\hat{u}_t$ will be non-stationary. However, this last property can be exploited to test for cointegration. If there is no cointegration, then a regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ should yield a unit coefficient. If there is cointegration, then a regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ should yield a coefficient that is less than one.

The proposal is thus to estimate (28) by OLS and then construct one of the standard unit root tests on the **estimated residuals**, such as the ADF t test or the PP's Z_β or Z_t test. Although theses test statistics are constructed in the same way as when they are applied to individual series y_t, when the test are applied to the residual $\hat{u}_t$ from a spurious regression, the critical values that are used to interpret the test statistics are different from those employed in Chapter 21.

Theorem 2 (Residual-Based test for Cointegration, Test with No Cointegration as Null):
Consider an $(k \times 1)$ vector $\mathbf{y}_t$ such that

$$(1-L)\mathbf{y}_t = \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t = \sum_{s=0}^{\infty} \mathbf{\Psi}_s \boldsymbol{\varepsilon}_{t-s},$$

for $\boldsymbol{\varepsilon}_t$ an $i.i.d.$ vector with mean zero, variance $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{\Omega} = \mathbf{PP}'$, and finite fourth moment and where $\{s \cdot \mathbf{\Psi}_s\}_{s=0}^{\infty}$ is absolutely summable. Let $g = (k-1)$ and $\mathbf{\Lambda} = \mathbf{\Psi}(1)\mathbf{P}$. Suppose that the $(k \times k)$ matrix $\mathbf{\Lambda\Lambda}'$ is nonsingular, and let $\mathbf{L}$ denote the Cholesky decomposition of $(\mathbf{\Lambda\Lambda}')^{-1}$. Partition $\mathbf{y}_t$ as $\mathbf{y}_t = (Y_{1t}, \mathbf{y}_{2t}')'$ and consider the OLS regression:

$$Y_{1t} = \hat{\alpha}_T + \mathbf{y}_{2t}'\hat{\boldsymbol{\beta}}_T + \hat{u}_t. \tag{23-30}$$

The residual $\hat{u}_t$ can then be regression on its own lagged value $\hat{u}_{t-1}$ without a constant term (since the original regression (31) has contained a constant term, the disturbance term u_t is zero-mean):

$$\hat{u}_t = \rho\hat{u}_{t-1} + e_t, \tag{23-31}$$

yield the OLS estimate

$$\hat{\rho}_T = \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_{t-1}^2}. \tag{23-32}$$

We may form standard Dickey-Fuller and Phillips-Perron (Z_ρ, Z_t) from (32). Alternatively we can form a ADF test from

$$\hat{u}_t = \zeta_1\triangle\hat{u}_{t-1} + \zeta_2\triangle\hat{u}_{t-2} + ... + \zeta_{p-1}\triangle\hat{u}_{t-p+1} + \rho\hat{u}_{t-1} + e_t. \tag{23-33}$$

Then the following results hold.

(a) The statistics $\hat{\rho}$ defined in (33) satisfies (standard DF test)

$$\begin{aligned}(T-1)(\hat{\rho}-1) \quad \xrightarrow{L} \quad & \left\{\frac{1}{2}\left\{[1 \ \ -\mathbf{h}_2'] \cdot [\mathbf{w}^*(1)][\mathbf{w}^*(1)]' \begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix}\right\}\right. \\ & -h_1[\mathbf{w}^*(1)]' \begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix} \\ & \left. -\frac{1}{2}[1 \ \ -\mathbf{h}_2']\mathbf{L}'[E(\triangle \mathbf{y}_t)(\triangle \mathbf{y}_t')]\mathbf{L} \begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix}\right\} \div H_n. \qquad (23\text{-}34)\end{aligned}$$

Here, $\mathbf{w}^*$ denotes n-dimensional standard Brownian motion partitioned as

$$\mathbf{w}^* = \begin{bmatrix} W_1^*(r)_{(1\times 1)} \\ \mathbf{w}_2^*(r)_{(g\times 1)} \end{bmatrix};$$

h_1 is a scalar and $\mathbf{h}_2$ is a $(g \times 1)$ vector given by

$$\begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & \int_0^1 [\mathbf{w}_2^*(r)]' dr \\ \int_0^1 \mathbf{w}_2^*(r) dr & \int_0^1 [\mathbf{w}_2^*(r)][\mathbf{w}_2^*(r)]' dr \end{bmatrix}^{-1} \times \begin{bmatrix} \int_0^1 W_1^*(r) dr \\ \int_0^1 \mathbf{w}_2^*(r) W_1^*(r) dr \end{bmatrix},$$

and

$$H_n = \int_0^1 [W_1^*(r)]^2 dr - \left[\int_0^1 W_1^*(r) dr \quad \int_0^1 [W_1^*(r)][\mathbf{w}_2^*(r)]' dr\right] \begin{bmatrix} h_1 \\ \mathbf{h}_2 \end{bmatrix}.$$

(b) If the $l \to \infty$ (Newey-West truncated parameter) as $T \to \infty$ but $l/T \to 0$, then the Phillips-Perron statistics constructed from $\hat{u}_t$, Z_ρ satisfies

$$Z_\rho \xrightarrow{L} Z_n, \qquad (23\text{-}35)$$

where

$$\begin{aligned} Z_n \quad \equiv \quad & \left\{\frac{1}{2}\left\{[1 \ \ -\mathbf{h}_2'] \cdot [\mathbf{w}^*(1)][\mathbf{w}^*(1)]' \begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix}\right\}\right. \\ & \left. -h_1[\mathbf{w}^*(1)]' \begin{bmatrix} 1 \\ -\mathbf{h}_2 \end{bmatrix} - \frac{1}{2}(1+\mathbf{h}_2'\mathbf{h}_2)\right\} \div H_n. \qquad (23\text{-}36)\end{aligned}$$

(c) If the $l \to \infty$ as $T \to \infty$ but $l/T \to 0$, then the Phillips-Perron statistics constructed from $\hat{u}_t$, Z_t satisfies

$$Z_t \xrightarrow{L} Z_n \cdot \sqrt{H_n} \div (1+\mathbf{h}_2'\mathbf{h}_2)^{1/2}. \qquad (23\text{-}37)$$

(d) If, in addition to the preceding assumptions, $\triangle \mathbf{y}_t$ follows a zero-mean stationary vector *ARMA* process and if $p \to \infty$ as $t \to \infty$ but $p/T^{1/3} \to 0$, then the *ADF* t test statistics associated with (34) has the same limiting distribution as the test statistics Z_t described in (38).

Results (a) implies that $\hat{\rho} \xrightarrow{p} 1$. Note that although $W_1^*(r)$ and $\mathbf{w}_2^*(r)$ are standard Brownian motion, the distribution of the term h_1, $\mathbf{h}_2$, H_n and Z_n above depend only on the number of stochastic explanatory variables included in the cointegrating regression $(k-1)$ and on whether a **constant term** (original cointegration regression) appears in the regression, $(T-1)(\hat{\rho}-1)$ are affected by the variance, correlations and dynamics of $\triangle \mathbf{y}_t$.

In the special case when $\triangle \mathbf{y}_t$ is *i.i.d.*, then $\mathbf{\Psi}(L) = \mathbf{I}_n$, and the matrix $\mathbf{\Lambda\Lambda}' = E(\triangle \mathbf{y}_t \triangle \mathbf{y}_t')$. Since $\mathbf{LL}' = (\mathbf{\Lambda\Lambda}')^{-1}$, it follows that $(\mathbf{\Lambda\Lambda}') = (\mathbf{L}')^{-1}(\mathbf{L})^{-1}$. Hence,

$$\mathbf{L}'[E(\triangle \mathbf{y}_t)(\triangle \mathbf{y}_t')]\mathbf{L} = \mathbf{L}'(\mathbf{\Lambda\Lambda}')\mathbf{L} = \mathbf{L}'[(\mathbf{L}')^{-1}(\mathbf{L})^{-1}]\mathbf{L} = \mathbf{I}_n. \quad (23\text{-}38)$$

If (39) is substituted into (35), the results is that when $\triangle \mathbf{y}_t$ is *i.i.d.*,

$$T(\hat{\rho}-1) \xrightarrow{L} Z_n$$

for Z_n defined in (37).

In the more general case when $\triangle \mathbf{y}_t$ is serially correlated, the limiting distribution of $T(\hat{\rho}-1)$ depends on the nature of this correlation as captured by the elements $\mathbf{L}$. However, the corrections for autocorrelation implicit in Phillips's Z_ρ and Z_t statistics or the augmented Dickey-Fuller t test turn out to generate variables whose distribution do not depend on any nuisance parameters.

Although the distribution of Z_ρ Z_t and the ADF_t do not depend on nuisance parameters, the distribution when these statistics are calculated from the residuals $\hat{u}_t$ **are not the same** as the distribution these statistics would have if calculated from the raw data $\triangle \mathbf{y}_t$. Moreover, different values for $(k-1)$ (the number of stochastic explanatory variables in the cointegrating regression) imply different characterizations of the limiting statistics, h_1, $\mathbf{h}_2$, H_n, and Z_n, meaning that a different critical value must be used to interpret Z_ρ for each value of $(k-1)$. Similarly, the asymptotic distribution of $\mathbf{h}_2$, H_n, and Z_n are different depending on whether a constant term is included in the cointegrating regression.

Example:
See the purchasing power parity example on Hamilton's p.598.

Exercise:
Reproduce the values in case 2 of Table B.8 and B.9 on Hamilton's p.765-766.

5.4 Tests with Cointegration as Null

The test considered in the previous sections are for the null hypothesis of no cointegration. These are based on tests for a unit root hypothesis in the residuals for the cointegrating regression. In Chapter 21 we discussed unit root tests with stationarity as the null hypothesis (e.g. $KPSS$). Correspondingly these are tests with cointegration as the null. They are
(a) the Leybourne and McCabe test (1993) which is based on an unobserved components model;
(b) the Park and Choi test (1988,1990) which is based on testing the significance of superfluous regressors;
(c) the Shin (1994) test which is a residual-based test
(d) Harris and Inder (1994) test which use non-parametric correction procedure for estimation of cointegration regression.

5.5 Testing Hypothesis About the Cointegrating Vector

The previous section described some way to test whether a vector $\mathbf{y}_t$ is cointegrated. It was noted that if $\mathbf{y}_t$ is cointegrated, then a consistent estimate of the cointegrating vector can be obtained by OLS. However, a difficulties arise with nonstandard distribution for hypothesis test about the cointegrating vector due to the possibility of nonzero correlation between z_t^* and $\mathbf{u}_{2t}$. The nuisance parameters $\boldsymbol{\lambda}_1^*$ and $\boldsymbol{\Lambda}_2^*$ which appear in (30) also cause a problem. The basic approach to constructing hypothesis tests will therefore be to transform the regression or the estimate so as to eliminate the effects of this correlation. The first one is Stock and Watson (1993)'s dynamic OLS which corrects the correlation by adding leads and lags of $\triangle\mathbf{y}_{2t}$. The second one is

Phillips and Hansen (1990)'s fully modified OLS estimate. Modification of the *OLS* have been made in two points. See Hatanaka (1996) p.266 for details.

6 Simulation of Bivariate Cointegrated System

To illustrate the potential difference in size and power between various residual-based test for cointegration in finite sample, a monte Carlo experiment proposed by Cheung and Lai (1993), similar to that of Engle and Granger (1987), can be conducted. A bivariate system of x_t and y_t is modeled by

$$x_t + y_t = u_t \tag{23-39}$$

and

$$x_t + 2y_t = v_t \tag{23-40}$$

with $(1-L)u_t = \varepsilon_t$, and v_t is generated as an $AR(1)$ process

$$(1-\phi L)v_t = \eta_t. \tag{23-41}$$

The innovation ε_t and η_t are generated as independent standard normal variates. When v_t is given by (42) with $|\phi| < 1$, x_t and y_t are cointegrated and (41) is their cointegrating relationship. However, when $|\phi| = 1$, the two series are not cointegrated.

Exercise:

Use the simulation based on 10000 replication in a sample of size 500 to compare the performance of size and power ($\phi = 0.85$) of residual-based ADF and PP test for cointegration on a nominal size 5%. Truncated number is chosen as $p = l = 4$.

Example.

The following is the code to cointegration regression. Let $X_t = X_{t-1} + v_t$ and $Y_t = 2 + 3X_t + u_t$, where

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \overset{i.i.d}{\sim} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}\right).$$

Consider the sequence of an OLS regression of Y_t on X_t and a constant

$$Y_t = \alpha + \beta X_t + \varepsilon_t,\ t = 1, 2, ..., T.$$

It can see that the OLS estimates of $\hat{\alpha}$ and $\hat{\beta}$ is super-consistent and the t-ratio to test the null hypothesis that $\alpha = 0$ and $\beta = 0$, $t_{\hat{\alpha}}$ and $t_{\hat{\beta}}$.

[Appendix]

$$\tilde{\mathbf{H}}(\mathbf{L}) = \begin{bmatrix} \mathbf{B'BA'A + B'C(L)A(1-L)} & \mathbf{B'C(L)A_\perp} \\ \mathbf{B'_\perp C(L)A(1-L)} & \mathbf{B'_\perp C(L)A_\perp} \end{bmatrix}$$

$$C(L) = (\xi(\mathbf{L}) - \xi(\mathbf{1}))/(\mathbf{1-L})$$

$$\tilde{\mathbf{H}}(\mathbf{L}) = \begin{bmatrix} \mathbf{B'BA'A + B'}(\xi(\mathbf{L}) - \xi(\mathbf{1}))\mathbf{A} & \mathbf{B'}(\xi(\mathbf{L}) - \xi(\mathbf{1}))\mathbf{A_\perp(1-L)^{-1}} \\ \mathbf{B'_\perp}(\xi(\mathbf{L}) - \xi(\mathbf{1}))\mathbf{A} & \mathbf{B'_\perp}(\xi(\mathbf{L}) - \xi(\mathbf{1}))\mathbf{A_\perp(1-L)^{-1}} \end{bmatrix}$$

$$\xi(\mathbf{1}) = -\xi_{\mathbf{0}} = \mathbf{BA'}$$

$$\tilde{\mathbf{H}}(\mathbf{L}) = \begin{bmatrix} \mathbf{B'BA'A + B'}(\xi(\mathbf{L}) - \mathbf{BA'})\mathbf{A} & \mathbf{B'}(\xi(\mathbf{L}) - \mathbf{BA'})\mathbf{A_\perp(1-L)^{-1}} \\ \mathbf{B'_\perp}(\xi(\mathbf{L}) - \mathbf{BA'})\mathbf{A} & \mathbf{B'_\perp}(\xi(\mathbf{L}) - \mathbf{BA'})\mathbf{A_\perp(1-L)^{-1}} \end{bmatrix}$$

$$\tilde{\mathbf{H}}(\mathbf{L}) = \begin{bmatrix} \mathbf{B'}\xi(\mathbf{L})\mathbf{A} & \mathbf{B'}\xi(\mathbf{L})\mathbf{A_\perp(1-L)^{-1}} \\ \mathbf{B'_\perp}\xi(\mathbf{L})\mathbf{A} & \mathbf{B'_\perp}\xi(\mathbf{L})\mathbf{A_\perp(1-L)^{-1}} \end{bmatrix}$$

$$\tilde{\mathbf{H}}(\mathbf{L}) = (\mathbf{B}, \mathbf{B_\perp})'\xi(\mathbf{L})(\mathbf{A}, \mathbf{A_\perp(1-L)^{-1}})$$

by the determinant

$$|AB| = |A||B| \& |B'| = |B|$$

$$|\tilde{\mathbf{H}}(\mathbf{L})| = |(\mathbf{B}, \mathbf{B_\perp})||\xi(\mathbf{L})||(\mathbf{A}, \mathbf{A_\perp})|(\mathbf{1-L})^{-(\mathbf{k-h})}$$