Lecture 9: Multivariate Unit Root Processes and Cointegration

1 Multivariate Unit Root Processes

From univariate unit root processes to multivariate unit root processes, we need to extend the scalar Brownian motion to the vector Brownian motion.

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

(a) $\mathbf{W}(0) = \mathbf{0}$*;*

(b) For any dates $0 \le r_1 < r_2 < \ldots < r_k \le 1$, the changes $[\mathbf{W}(r_2) - \mathbf{W}(r_1)]$, $[\mathbf{W}(r_3) - \mathbf{W}(r_2)]$, $\ldots, [\mathbf{W}(r_k) - \mathbf{W}(r_{k-1})]$, are independent multivariate Gaussian with $[\mathbf{W}(s) - \mathbf{W}(r)] \sim N(\mathbf{0}, (s-r)I_k)$;

(c) For any given realization, $\mathbf{W}(r)$ is continuous in r with probability 1.

Let $\mathbf{v}_t$ be a k-dimensional *i.i.d.* vector process with $E(\mathbf{v}_t) = 0$ and $E(\mathbf{v}_t\mathbf{v}_t') = I_k$. Define that $\tilde{\mathbf{X}}_n(r) = n^{-1}(\mathbf{v}_1 + \ldots + \mathbf{v}_{[nr]})$, then the vector version FCLT is given by

$$\sqrt{n}\tilde{\mathbf{X}}_n(\cdot) \to_d \mathbf{W}(\cdot), \tag{1}$$

Let $\boldsymbol{\epsilon}_t$ be a k-dimensional *i.i.d.* vector process with $E(\boldsymbol{\epsilon}_t) = 0$ and $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \Omega$. Cholesky decomposition of Ω gives

$$\Omega = PP'$$

Or we can write $\boldsymbol{\epsilon}_t = P\mathbf{v}_t$. Define $\mathbf{X}_n^*(r) = n^{-1}(\boldsymbol{\epsilon}_1 + \ldots + \boldsymbol{\epsilon}_{[nr]})$, Then (1) and CMT gives that

$$\sqrt{n}\mathbf{X}_n^*(\cdot) \to P\mathbf{W}(\cdot). \tag{2}$$

Finally consider serially correlated errors $\mathbf{u}_t = \sum_{s=0}^{\infty} C_s\boldsymbol{\epsilon}_{t-s}$, where if C_{ij}^s denotes the ijth element of $\mathbf{C}_s$,

$$\sum_{s=0}^{\infty} s|C_{ij}^s| < \infty.$$

Apply the BN decomposition,

$$\sum_{s=1}^{t} \mathbf{u}_s = \mathbf{C}(1)\sum_{s=1}^{t} \boldsymbol{\epsilon}_s + \boldsymbol{\eta_t} - \boldsymbol{\eta_0},$$

where $\mathbf{C}(1) = (\mathbf{C}_0 + \mathbf{C}_1 + \ldots)$, $\boldsymbol{\eta_t} = \sum_{s=0}^{\infty} \boldsymbol{\alpha}_s \boldsymbol{\epsilon}_{t-s}$ for $\boldsymbol{\alpha}_s = -(\mathbf{C}_{s+1} + \mathbf{C}_{s+2} + \ldots)$ and $\boldsymbol{\alpha}_s$ is absolutely summable. Now define $\mathbf{X}_n(r) = (1/n)(\mathbf{u}_1 + \ldots \mathbf{u}_{[nr]})$, then

$$\sqrt{n}\mathbf{X}_n(\cdot) \to \mathbf{C}(1)PW(\cdot). \tag{3}$$

The proposition 18.1 in Hamilton (p. 547) (we will use P 18.1 for short in this lecture) summarized many useful asymptotic results for vector unit root processes. Most of them are analogous to the univariate cases: you replace γ with Γ, replace λ with Λ, etc.

2 Spurious Regression

Consider two independent I(1) variables, x_1 and x_2. If we regress x_1 on x_2, despite the fact that they are actually independent, the OLS estimates of the coefficient may be significant. This phenomenon is called *spurious regression* (Granger and Newbold (1974), Phillips (1986)). Proposition 18.2 in Hamilton (1994) gives the results that have been developed by Phillips (1986). We will reproduce a two-variable version for simplicity (some degree) of representation. I think that this will be easier to read, but you are still encouraged to read the original propositions and proofs.

Let $y_t = (x_{1t}, x_{2t})'$, and it is generated by

$$\Delta y_t = \Psi(L)\boldsymbol{\epsilon}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \boldsymbol{\epsilon}_t,$$

where the error $\boldsymbol{\epsilon}_t$ satisfies our standard assumption: mean zero and finite fourth moment and $s\mathbf{C}_s$ is absolutely summable.

Consider the regression

$$x_{1t} = \alpha + \gamma' x_{2t} + u_t. \tag{4}$$

The OLS coefficient estimates for a sample of size n are given by

$$\begin{bmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{bmatrix} = \begin{bmatrix} n & \sum x_{2t} \\ \sum x_{2t} & \sum x_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_{1t} \\ \sum x_{2t}x_{1t} \end{bmatrix}.$$

Consider a null hypothesis $H_0 : R\gamma = r$, where R is $m \times 2$ matrix representing m separate hypothesis involving the coefficients. the OLS F-statistics is

$$F_n = (R\hat{\gamma}_n - r)' \left\{ s_n^2 \begin{bmatrix} \mathbf{0} & R \end{bmatrix} \begin{bmatrix} n & \sum x_{2t} \\ \sum x_{2t} & \sum x_{2t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}' \\ R' \end{bmatrix} \right\}^{-1} (R\hat{\gamma}_n - r)/m. \tag{5}$$

where $s_n^2 = (n-2)^{-1} \sum_{t=1}^{n} \hat{u}_t^2$. To derive the asymptotics for the estimates and test statistics, we will do some transformation. Let $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = PP'$, and $\Lambda = \Psi(1)P$. Partition $\Lambda\Lambda'$ as

$$\Lambda\Lambda' = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

Suppose that $\Lambda\Lambda'$ is nonsingular, and define

$$(\sigma^*)^2 = \sigma_1^2 - \sigma_{12}^2/\sigma_2^2$$

.

To further simplify the problem we assume that ϵ_1 and ϵ_2 are independent, then $\sigma_{12} = \sigma_{21} = 0$ and $\sigma^* = \sigma_1$. And the L_{22} matrix in proposition 18.2 is just σ_2^{-1}. Then the three part of proposition 18.2 in this problem become

(a) The OLS estimates $\hat{\alpha}_n$ and $\hat{\gamma}_n$ are characterized by

$$\begin{bmatrix} n^{-1/2}\hat{\alpha}_n \\ \hat{\gamma}_n \end{bmatrix} \to_d \begin{bmatrix} \sigma_1 h_1 \\ (\sigma_1/\sigma_2)h_2 \end{bmatrix}, \tag{6}$$

where

$$\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \equiv \begin{bmatrix} 1 & \int [W_2(r)]' dr \\ \int W_2(r) dr & \int [W_2(r)][W_2(r)]' dr \end{bmatrix}^{-1} \begin{bmatrix} \int W_1(r) dr \\ \int W_1(r) W_2(r) dr \end{bmatrix} \equiv Q^{-1}F,$$

where $W_1(r)$ and $W_2(r)$ are independent standard Brownian motion and Q and F are defined to be the matrix and vector.

(b) The sum of squared residual RSS_n from OLS estimation satisfy

$$n^{-2}RSS_n \to \sigma_1^2 H$$

where

$$H = \int W_1(r)^2 dr - F'QF.$$

(c) The OLS F test satisfies

$$n^{-1}F_n \to [(\sigma_1/\sigma_2)R - r]\left[(\sigma_1/\sigma_2)^2 H \begin{bmatrix} \mathbf{0} & R \end{bmatrix} Q^{-1} \begin{bmatrix} \mathbf{0}' \\ R' \end{bmatrix}\right]^{-1} [(\sigma_1/\sigma_2)R - r]/m. \tag{7}$$

The proof of the above results for the general case can be found on page P564-548 in Hamilton. In our simple case, (6) tells neither of the estimates $(\hat{\alpha}, \hat{\gamma})$ is consistent. Recall that if $\hat{\beta}$ is a consistent estimate for β, then $\hat{\beta} - \beta \to 0$ we we have to scale it with n^r where $r > 0$ to obtain a nondegenerate limit distribution. Actually in this problem the OLS estimate of α diverges with sample size n, since we have to scale it with $n^{-1/2}$ to obtain a limit distribution.

Result (b) then tells that the OLS estimate of the variance of u_t also diverges:

$$s_n^2 = (n-k)^{-1}RSS_n \to \infty.$$

This is because in a spurious regression, the residual $\hat{u}_t$ is an I(1) nonstationary process. To see this,

$$\hat{u}_t = x_{1t} - \hat{\alpha}_n - \hat{\gamma}_n x_{2t},$$

taking difference,

$$\Delta\hat{u}_t = \Delta x_{1t} - \hat{\gamma}_n \Delta x_{2t} = \begin{bmatrix} 1 & -\hat{\gamma}_n \end{bmatrix} \begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} \to \begin{bmatrix} 1 & -(\sigma_1/\sigma_2)h_2 \end{bmatrix} \Delta y_t,$$

which is a random vector times an I(0) variable, so $\Delta\hat{u}_t$ is I(0), and $\hat{u}_t$ is I(1). Hence we $s_n^2 = n^{-1}\sum_{t=1}^n \hat{u}_t^2$ diverges, and $n^{-2}\sum_{t=1}^n \hat{u}_t^2$ converges.

Result (c) tells that any OLS t or F statistics based on spurious regression also diverges. The usual t statistics has be divided by $n^{1/2}$ to converge, and the usual F statistics has to be divided by n to converge. If we draw inference based on the usual test statistics, we tend to accept that x_{1t} and x_{2t} are significantly related even when they are independent.

There are three ways to cure for spurious regression. First, we may include the lags of both the independent and dependent variables, say

$$x_{1t} = \alpha + \phi x_{1,t-1} + \gamma x_{2t} + \delta x_{2,t-1} + u_t.$$

Now the value $\phi = 1, \gamma = \delta = 0$ will make u_t I(0), therefore most of usual OLS inferences are valid now, although tests of some hypothesis will still involve non-standard distribution.

The second cure is to difference the data before regression,

$$\Delta x_{1t} = \alpha + \gamma \Delta x_{2t} + u_t. \tag{8}$$

Now u_t is again I(0) and all usual OLS inferences are valid. However, in differencing the data, we may lose some information in the data.

Finally, we could apply GLS to estimate the system. We first apply an AR(1) regression on the residual $\hat{u}_t$, $\hat{u}_t = \hat{\rho}\hat{u}_{t-1} + e_t$, then define $\tilde{x}_{1t} = x_{1t} - \hat{\rho} x_{1,t-1}$, $\tilde{x}_{2t} = x_{2t} - \hat{\rho} x_{2,t-1}$, and then regress $\tilde{x}_{1t}$ on $\tilde{x}_{2t}$. Since $\hat{u}_t$ is a unit root process, $\hat{\rho} \to 1$. Therefore, this Cochrane-Orcutt GLS regression is asymptotically equivalent to running OLS on the differenced data (8).

3 Cointegration

3.1 Introduction

In the previous section, we showed that when we regress one I(1) variable on another I(1) variable, and when the residuals of the regression is also I(1), then it is a spurious regression. Even when these two variables are independent, usual OLS inference may imply that they are significantly related. Now you may wonder when it is valid to run an OLS regression between I(1) variables? It turns out the regression is valid only when the residual is stationary, and in this case, we say that those I(1) variables are *cointegrated*.

There are two facts about cointegration. First, a cointegration is a relationship that applies only to I(1) series. Second, although each individual series, say $x_{1t}, x_{2t}, \ldots x_{kt}$, are $I(1)$, and let $y_t = (x_{1t}, x_{2t}, \ldots x_{kt})'$, there exist a nonzero k by 1 vector γ, such that the series γy_t is $I(0)$. There are many examples of cointegration in economic applications. For instance, both income and consumption maybe nonstationary, but they seem to keep a stable relation with each other. Or if we look at some data of short rate and 3-month forward rate, they also tend to have a stable relationship over time, although they all wander around.

Example: consider the following system of processes

$$\begin{aligned} x_{1t} &= \beta_1 x_{2t} + \beta_2 x_{3t} + u_{1t} \\ x_{2t} &= \beta_3 x_{3t} + u_{2t} \\ x_{3t} &= x_{3,t-1} + u_{3t} \end{aligned}$$

where the three error terms are uncorrelated white noise processes. Clearly, all those three processes are individually $I(1)$. Let $y_t = (x_{1t}, x_{2t}, x_{3t})'$ and $\gamma = (1, -\beta_1, -\beta_2)$, then $\gamma y_t = u_{1t}$ which is a $I(0)$ process. Another cointegrating relationship is between x_{2t} and x_{3t}. So we can let $\gamma^* = (0, 1, -\beta_3)$, then $\gamma^* y_t = u_{2t}$ is also I(0).

3.1.1 Cointegrating Matrix

In the above example, we see that the cointegrating vector is not unique. Also, note that γ and γ^* are lineally independent. In general, if the cointegrating system has k $I(1)$ series, we can have h lineally independent cointegrating vectors, with $h < k$. In the example, we have 3 I(1) series and we have 2 cointegrating relations. Let γ_i, $i = 1, \ldots, h$ denote each of these vectors, then we could construct a h by k matrix

$$A' = \begin{pmatrix} \gamma_1' \\ \vdots \\ \gamma_h' \end{pmatrix}.$$

Then the vector $A'y_t$ is h-vector valued stationary time series. In our above example,

$$A' = \begin{pmatrix} 1 & -\beta_1 & -\beta_2 \\ 0 & 1 & -\beta_3 \end{pmatrix}$$

and

$$A'y_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Given a matrix A' $(h \times k)$ whose rows are linearly independent and $A'y_t$ is a stationary $(h \times 1)$ vector. Suppose further that if c' is any $(1 \times k)$ vector that is linearly independent of the rows of A', then $c'y_t$ is nonstationary. Then we say there are exactly h conintegrating relations among the elements of y_t and that the rows of A' $(\gamma_1, \ldots, \gamma_h)$ form a basis for the space of cointegrating vectors.

3.1.2 MA representations

In univariate time series analysis, sometimes we would like to use differenced data if the original data is $I(1)$. For example, if x_t is I(1), then we may take difference to get Δx_t and specify an AR(p) process for Δx_t. However, we cannot do this for a cointegrated system. Assume that Δy_t is $I(0)$, and let $\delta = E(\Delta y_t)$. Define

$$\mathbf{u}_t = \Delta y_t - \delta. \tag{9}$$

The $\mathbf{u}_t$ is a stationary process by assumption. Suppose that $\mathbf{u}_t$ has the Wold decomposition $\mathbf{u}_t = \Psi(L)\boldsymbol{\epsilon}_t$ where $\boldsymbol{\epsilon}_t$ is a vector white noise. Let $\Psi(1) = I_k + \Psi_1 + \Psi_2 + \ldots$.

The difference equation (9) implies that the BN-decomposition of y_t gives

$$\begin{aligned} y_t &= y_0 + \delta t + \mathbf{u}_1 + \mathbf{u}_2 + \ldots \\ &= y_0 + \delta t + \Psi(1)(\boldsymbol{\epsilon}_1 + \boldsymbol{\epsilon}_2 + \ldots + \boldsymbol{\epsilon}_t) + \eta_t - \eta_0 \end{aligned}$$

Premultiply by A',

$$A'y_t = A'y_0 + A'\delta t + A'\Psi(1)(\boldsymbol{\epsilon}_1 + \boldsymbol{\epsilon}_2 + \ldots + \boldsymbol{\epsilon}_t) + A'\eta_t - A'\eta_0.$$

To ensure that $A'y_t$ is stationary, the coefficients on the nonstationary components t and $\sum_{i=1}^t \boldsymbol{\epsilon}_i$ must be zero, i.e.

$$A'\Psi(1) = 0 \quad A'\delta = 0.$$

Note that $A'\Psi(1) = 0$ implies that the $|\Psi(z)| = 0$ at $z = 1$. This in turn means that $\Psi(L)$ is noninvertible. Thus a cointegrated system can never be represented by a finite order VAR in the differenced data Δy_t. Intuitively, this is because in the dynamics of the system, the *level* of the variables matters.

3.1.3 Phillips's Triangular Representation

Obviously the cointegrating matrix is not unique for a cointegrated system. Therefore researchers can choose to use a representation that is convenient for their problems.

Phillips (1991) suggested that the h by k cointegrating matrix A be transformed as

$$A = \begin{pmatrix} I_h & -\Gamma \end{pmatrix},$$

where Γ is a matrix of size h by $g = k - h$. Define z_t as

$$z_t = A' y_t.$$

Correspondingly, rearrange $y_t = (y_{1t}, y_{2t})'$, then we could represent y_{1t} and y_{2t} separately,

$$y_{1t} = \Gamma y_{2t} + z_t$$

and

$$\Delta y_{2t} = \delta_2 + v_{2t}$$

where δ_2 and u_{2t} are the last g elements of δ and u_t. I will show how this works with our example.

In our example,

$$A' = \begin{pmatrix} 1 & -\beta_1 & -\beta_2 \\ 0 & 1 & -\beta_3 \end{pmatrix}$$

Transform A to the form

$$A' = \begin{pmatrix} 1 & 0 & -\beta_2 - \beta_1\beta_3 \\ 0 & 1 & -\beta_3 \end{pmatrix}$$

Then

$$z_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Therefore, we could represent y_t as

$$\begin{aligned} y_{1t} &= \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \beta_2 + \beta_1\beta_3 \\ \beta_3 \end{pmatrix} x_{3t} + z_t \\ \Delta y_{2t} &= \Delta x_{3t} = u_{3t} \end{aligned}$$

The cointegrating relationships is very clear from this representation. There are also other representations which are convenient in some problems (reading: page 578 - 582).

3.2 Cointegration tests

In our discussion of spurious regression, we learn that given two $I(1)$ processes, even if they are independent, OLS estimator may turn out to be significant according to regular test statistics, such as t and F test. Therefore, we should be cautious when running regressions between nonstationary time series variables. However, if we know that two or more $I(1)$ series are cointegrated, then it is valid to apply our linear regression techniques. Therefore, to test whether a system of nonstationary processes are cointegrated becomes critical in multivariate nonstationry time series studies. In the test for cointegration, we let the null be no integration among the elements of a $(k \times 1)$ vector y_t; rejection of the null is then taken as evidence of cointegration.

3.2.1 Cointegrating vector is known

If we already know the cointegrating vector, or a specific relationship is implied by economic theories, say, γ, then to test $y_t = (x_{1t}, \ldots, x_{kt})$ are cointegrated can be done in two steps. First, we test if x_{it} is $I(1)$ for $i = 1, \ldots, k$. Second, if all series in y_t are $I(1)$, then test if $\gamma y_t = z_t$ is $I(0)$. If it is, then the system is cointegrated. In both steps, we could use the unit root tests we have discussed in the previous lecture.

3.2.2 Estimating the cointegrating vector

If the cointegration vector is unknown, then we could first test if each series is $I(1)$, then estimate the cointegrating vector γ using OLS, and finally, we test the null hypothesis of cointegration, which is equivalent to test that the residual $\hat{u}_t$ is $I(0)$. With the OLS estimates, $\hat{\gamma}$, if $\hat{u}_t$ is I(0), then the vector is cointegrated; if $\hat{u}_t$ is I(1), then the regression is spurious. The following proposition (proposition 19.2) summarized the asymptotic results in this approach (for simplicity, we let y_2 be a scalar)

Proposition 1 *Suppose*

$$\begin{aligned} y_{1t} &= \alpha + \gamma y_{2t} + u_{1t} \\ \Delta y_{2t} &= u_{2t} \\ \mathbf{u}_t &= \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = C(L)\boldsymbol{\epsilon}_t, \end{aligned} \tag{10}$$

where $\boldsymbol{\epsilon}_t$ is an i.i.d. vector with mean zero, finite fourth moments, and positive definite variance-covariance matrix $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = PP'$. Further suppose that sC_s is absolutely summable and that the rows of $C(1)$ are linearly independent. Let $\hat{\alpha}$ and $\hat{\gamma}$ be OLS estimate

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} n & \sum y_{2t} \\ \sum y_{2t} & \sum y_{2t}^2 \end{bmatrix} \begin{bmatrix} \sum y_{1t} \\ \sum y_{2t} y_{1t} \end{bmatrix} \tag{11}$$

Partition $C(1) \cdot P$ as

$$C(1) \cdot P = \begin{bmatrix} \boldsymbol{\lambda}_1' \\ \boldsymbol{\lambda}_2 \end{bmatrix}$$

then

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha) \\ n(\hat{\gamma} - \gamma) \end{bmatrix} \to_d \begin{bmatrix} 1 & \int \mathbf{W}(r)' dr \boldsymbol{\lambda}_2' \\ \boldsymbol{\lambda}_2 \int \mathbf{W}(r) dr & \boldsymbol{\lambda}_2 \int \mathbf{W}(r)\mathbf{W}(r)' dr \boldsymbol{\lambda}_2' \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

where $\mathbf{W}(r)$ is a 2 dimensional standard Brownian motion, and

$$\begin{aligned} h_1 &\equiv \boldsymbol{\lambda}_1' W(1) \\ h_2 &\equiv \boldsymbol{\lambda}_2 \left[\int_0^1 \mathbf{W}(r) d\mathbf{W}(r)' \right] \boldsymbol{\lambda}_1' + \sum_{v=0}^{\infty} E(u_{2t} u_{1,t+v}). \end{aligned}$$

To understand the results, consider the simple example that u_{1t} and u_{2s} are uncorrelated,

$$C(L) = I_2 + \begin{bmatrix} 0 & 0 \\ 0 & cL \end{bmatrix}, \quad E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

Then

$$\boldsymbol{\lambda}_1' = [\sigma_1 \quad 0] \quad \boldsymbol{\lambda}_2 = [0 \quad (1+c)\sigma_2].$$

Hence

$$\begin{aligned} h_1 &= \sigma_1 W_1(1) \\ h_2 &= (1+c)\sigma_1\sigma_2 \int_0^1 W_2(r)dW_1(r) \end{aligned}$$

The proof of this proposition can be found on page 618-619 in Hamilton. Basically the proof uses the results from proposition 18.1 on multivariate unit root processes.

Note that in this example we assume u_{1t} and u_{2s} are uncorrelated, so $\sum_{v=0}^{\infty} E(u_{2t}u_{1,t+v}) = 0$. In the general case, u_{1t} and u_{2s} could be correlated, therefore induce bias in the estimates, however this bias is in $\hat{\gamma}_n$ is $O_p(n^{-1})$. To correct this bias caused by correlations between u_1 and u_2, we can add leads and lags in the regression. Define $\tilde{u}_{1t}$ as the residual from a linear projection of u_{12} on $\{u_{2,t-p}, \ldots, u_{2,t-1}, u_{2t}, u_{2,t+1}, \ldots, u_{2,t+p}\}$,

$$u_{1t} = \sum_{s=-p}^{p} \beta_s' u_{2,t-s} + \tilde{u}_{1t},$$

then $\tilde{u}_{1t}$ is uncorrelated with u_{2t}. We can then rewrite the regression (10) can be written as

$$y_{1t} = \alpha + \gamma y_{2t} + \sum_{s=-p}^{p} \beta_s' u_{2,t-s} + \tilde{u}_{1t}. \tag{12}$$

Now the estimates are consistent.

3.2.3 Testing for cointegration among trending series

Still with our 2-variable model, suppose that there is a time trend in y_{2t},

$$\begin{aligned} y_{1t} &= \alpha + \gamma y_{2t} + u_{1t} \\ \Delta y_{2t} &= \delta_2 + u_{2t} \end{aligned} \tag{13}$$

with $\delta \neq 0$. Then the process

$$y_{2t} = y_{20} + \delta_2 t + \sum_{s=1}^{t} u_{2s}$$

is asymptotically dominated by the deterministic time trend $\delta_2 t$. Then the OLS estimates $\hat{\alpha}$ and $\hat{\gamma}$ in (13) have the same limit distribution as regressing an I(1) series on a constant and a time trend. If y_{1t} also contain a deterministic time trend:

$$\Delta y_{1t} = \delta_1 + u_{1t},$$

then $\hat{\gamma}_n$ in (13) converges to (δ_1/δ_2).

3.2.4 Phillips and Hansen's fully modified OLS estimates

3.3 Testing hypothesis of cointegration vector

Consider the system

$$y_{1t} = \alpha + \gamma y_{2t} + u_{1t}, \tag{14}$$

$$y_{2t} = y_{2,t-1} + u_{2t}, \tag{15}$$

where y_{1t}, y_{2t} are I(1) while u_{1t}, u_{2t} are i.i.d. normal sequence and they are independent of each other.

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right).$$

And if we consider a null hypothesis about the cointegrating vector, say,

$$R_1\alpha + R_2\gamma = r.$$

It turns out the correct approach is just to estimate (14) with OLS and use standard t or F statistics to test any hypothesis about the cointegrating vector. No special procedures or unusual critical values are needed.