Lecture 8: Univariate Processes with Unit Roots*

1 Introduction

1.1 Stationary AR(1) vs. Random Walk

In this lecture, we will discuss a very important type of processes: unit root processes. For an AR(1) process

$$x_t = \phi x_{t-1} + u_t \tag{1}$$

to be stationary, we require that $|\phi| < 1$. Or in an AR(p) process, we require that all the roots of

$$1 - \phi_1 z - \ldots - \phi_p z^p = 0$$

lie out of the unit circle.

If one of the roots turns out to be one, then this process is called unit root process. In an AR(1) process, we have $\phi = 1$,

$$x_t = x_{t-1} + u_t. \tag{2}$$

It turns out that two processes with ($|\phi| < 1$ and $\phi = 1$) behave in very different manners. For simplicity, we assume that the innovations u_t follows a i.i.d. Gaussian distribution with mean zero and variance σ^2.

First, I plot the following two graphs in figure 1. In the left graph, I set $\phi = 0.9$ and in the right graph, I draw the random walk process, $\phi = 1$. From the left graph, we see that x_t moves around zero and never gets out of the $[-6, 6]$ region. There seems to be some force there which pulls the process to its mean (zero). But in the right graph, we did not see a fixed mean, instead, x_t moves 'freely' and in this case, it goes to as high as about 72. If we repeat generating the above two process, we would see that the $\phi = 0.9$ processes look pretty much the same; but the random walk processes are very different from each other. For instance, in a second simulation, it may go down to -80, say.

Above is some graphical illustration. Second, consider some moments of the process x_t when $|\phi| < 1$ and $\phi = 1$. When $|\phi| < 1$, we have that

$$E(x_t) = 0 \quad \text{and} \quad E(x^2) = \sigma^2/(1-\phi^2),$$

when $\phi = 1$, we no longer have constant unconditional moments, the first two conditional moments are

$$E(x_t|\mathcal{F}_{t-1}) = x_{t-1} \quad \text{and} \quad E(x_t^2|\mathcal{F}_{t-1}) = \sigma^2.$$

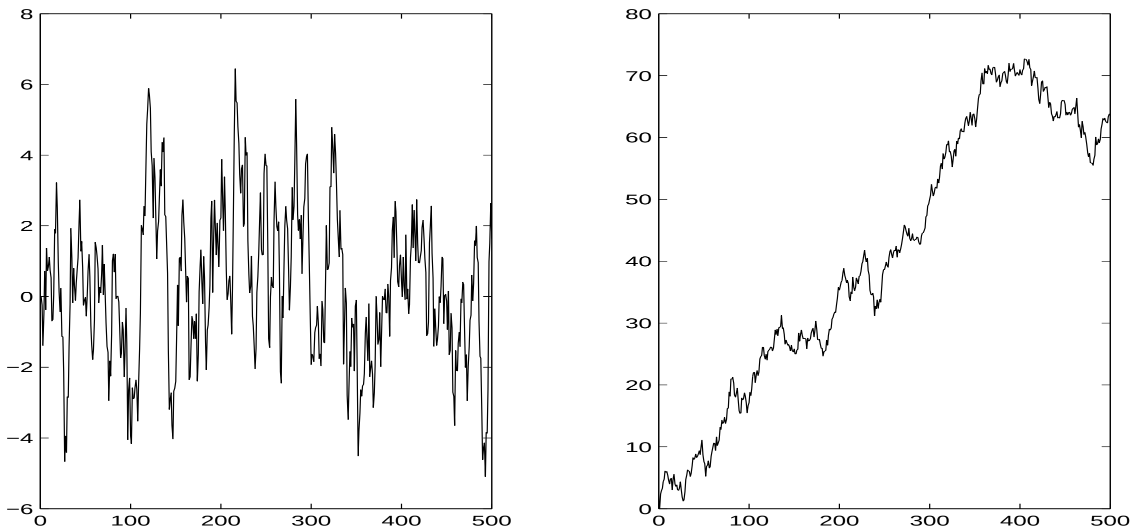

Figure 1: Simulated Autoregressive Processes with Coefficient $\phi = 0.9$ and $\phi = 1$

When we do k-period ahead forecasting, when $|\phi| < 1$,

$$E(x_{t+k}|\mathcal{F}_t) = \phi^k x_t.$$

Since $|\phi| < 1$, $\phi^k \to 0 = E(x_t)$ when $k \to \infty$. So as the forecasting horizon increases, the current value of x_t matters less and less since the conditional expectation converges to the unconditional expectation.

The variance of the forecasting

$$Var(x_{t+k}|\mathcal{F}_t) = (1 + \phi^2 + \ldots + \phi^{2k})\sigma^2 = \frac{1 - \phi^{2k+2}}{1 - \phi^2}\sigma^2,$$

which converges to $\sigma^2/(1 - \phi^2)$ as $k \to \infty$.

Next, consider the case when $\phi = 1$.

$$E(x_{t+k}|\mathcal{F}_t) = x_t,$$

which means that the current value does matter (actually it is the only thing that matters) even as $k \to \infty$!

The variance of the forecasting is

$$Var(x_{t+k}|\mathcal{F}_t) = k\sigma^2 \to \infty \quad \text{as} \quad k \to \infty.$$

If we let $x_0 = 1$, $\sigma^2 = 1$, we could draw the forecasting of x_{t+k} when $\phi = 0.9$ and $\phi = 1$ in Figure 2.

The upper graph in Figure 2 plots the forecasting for x_k when $\phi = 0.9$. The expectation for x_k conditional on $x_0 = 1$ drops to zero as k increases, and the forecasting error converges to the unconditional standard error quickly. The lower graph in Figure 2 plots the forecasting for x_k when $\phi = 1$. Obviously, the forecasting interval diverges as k increases.

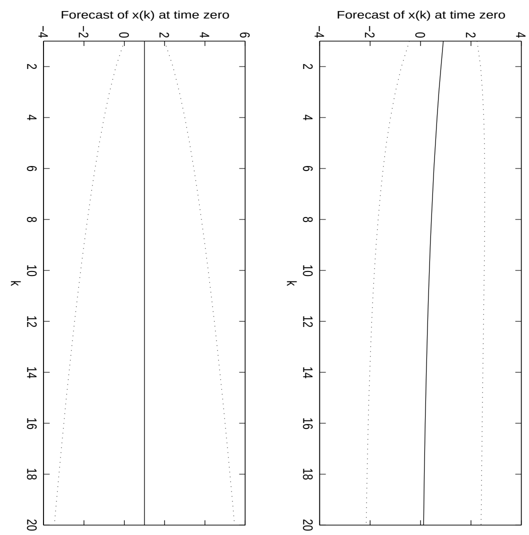

Figure 2: Forecasting of x_k at time zero when $\phi = 0.9$ and $\phi = 1$ ($x_0 = 1$, $\sigma^2 = 1$)

A third way to compare a stationary and a nonstationary autoregressive process is to compare their impulse-response functions. We can invert (1) to

$$\begin{aligned} x_t &= u_t + \phi u_{t-1} + \phi^2 u_{t-2} + \ldots + \phi^{t-1} u_1 \\ &= \sum_{k=0}^{t-1} \phi^k u_{t-k} \end{aligned}$$

So the effect of a shock u_t to x_{t+h} is ϕ^h, which dies out as h increases. In the unit root case,

$$x_t = \sum_{k=1}^{t} u_k,$$

The effect of u_t on x_{t+h} is one, which is independent of h. So if a process is a random walk, the effects of all shocks on the level of $\{x\}$ are *permanent*. Or the impulse-response function is flat at one.

Finally, we can compare the asymptotic distribution of the coefficient estimator of a stationary and a nonstationary autoregressive process.

For an AR(1) process, $x_t = \phi x_{t-1} + u_t$ where $|\phi| < 1$ and $u_t \sim i.i.d.N(0, \sigma^2)$, we have shown in lecture note 6 that the MLE estimator of ϕ is asymptotically normal,

$$n^{1/2}(\hat{\phi}_n - \phi_0) \sim N\left(0, \sigma^2 \left(n^{-1} \sum_{t=1}^{n} x_t^2\right)^{-1}\right).$$

However, if $\phi_0 = 1$, then $(n^{-1}\sum_{t=1}^{n} x_t^2)^{-1}$ goes to zero as $n \to \infty$. This implies that if $\phi_0 = 1$, then $\hat{\phi}_n$ converges at a order higher than $n^{1/2}$.

Above we have only considered AR(1) process. In a general AR(p) process, if there is one unit root, then the process is a nonstationary unit root process. Consider an AR(2) example. let $\lambda_1 = 1$, and $\lambda_2 = 0.5$, then

$$(1 - L)(1 - 0.5L)x_t = \epsilon_t, \quad \epsilon_t \sim i.i.d.(0, \sigma^2).$$

Then

$$(1 - L)x_t = (1 - 0.5L)^{-1}\epsilon_t = \theta(L)\epsilon_t \equiv u_t.$$

So the difference of x_t, $\Delta x_t = (1 - L)x_t$ is a stationary process, and

$$x_t = x_{t-1} + u_t,$$

is a unit root process with serially correlated errors.

1.2 Stochastic Trend v.s. Deterministic Trend

In a unit root process,

$$x_t = x_{t+1} + u_t,$$

where u_t is a stationary process, then x_t is said to be *integrated of order one*, denoted by $I(1)$. An $I(1)$ process is also said to be *difference stationary*, compared to *trend stationary* as has been discussed in the previous lecture. If we need to take difference twice to get a stationary process, then this process is said to be integrated of order two, denoted by $I(2)$, and so force. A stationary process can be denoted by $I(0)$. If a process is a stationary ARMA(p, q) after taking kth differences, then the original process is called ARIMA(p, k, q). The 'I' here denotes integrated.

Recall that we learned about spectrum in lecture 3. The implications of these 'integrate' and 'difference' operators to spectral analysis is that when you do 'integration' or sum, you filter out the high frequency components and what remains are low frequencies; which is a feature of unit root process. Recall that the spectrum of a stationary AR(1) process is

$$S(\omega) = \frac{1}{2\pi}(1 + \phi^2 - 2\phi\cos\,\omega)^{-1}.$$

When $\phi \to 1$, we have $S(\omega) = 1/[4\pi(1 - \cos\,\omega)]$. Then when $\omega \to 0$, $S(\omega) \to \infty$. So processes with stochastic trend have infinite spectrum at the origin. Recall that $S(\omega)$ decomposes the variance of a process into components contributed by each frequencies. So the variance of a unit root process are largely contributed by low frequencies. On the other hand, when we do 'difference', we filter out the low frequencies and what remains are the high frequencies.

In the previous lecture, we discussed the processes with deterministic trend. We can compare a process with deterministic trend (DT) and a process with stochastic trend (ST) from two perspectives. First, when we do k-period ahead forecasting, as $k \to \infty$, the forecasting error for DT converges to the variance of its stationary components, which is bounded. But as we see from the previous section, the forecasting error for ST diverges as $n \to \infty$. Second, the impulse-response function for a DT is the same as in the stationary case: the effect of a shock dies out quickly. While the impulse-function for ST is flat at one: the effect of all shocks on the level are permanent.

However, note that in Figure 1, we plot a simulated random walk, but part of its path looks like to have a upward time trend. This turns out to be a quite general problem: over a short time period, it is very hard to judge whether a process has a stochastic trend, or deterministic trend.

2 Brownian Motion and Functional Central Limit Theorem

2.1 Brownian Motion

To derive statistical inference of a unit root process, we need to make use of a very important stochastic process – *Brownian motion* (also called *Wiener process*). To understand a Brownian motion, consider a random walk

$$y_t = y_{t-1} + u_t, \quad y_0 = 0, \quad u_t \sim i.i.d.N(0,1). \tag{3}$$

We can then write

$$y_t = \sum_{s=1}^{t} u_s \sim N(0,t),$$

and the change in the value of x between dates t and s,

$$y_t - y_s = u_{s+1} + u_{s+1} + \ldots + u_t = \sum_{i=s+1}^{t} u_i \sim N(0, t-s)$$

and it is independent of the change between dates r and q for $s < t < r < q$.

Next, consider the change $y_t - y_{t-1} = u_t \sim i.i.d.N(0,1)$. If we view u_t as the sum of two independent Gaussian variables,

$$u_t = \epsilon_{1t} + \epsilon_{2t}, \quad \epsilon_{it} \sim i.i.d.N\left(0, \frac{1}{2}\right).$$

Then we can associate ϵ_{1t} with the change between y_{t-1} and the value of y at some interim point (say, $y_{t-(1/2)}$), and ϵ_{2t} with the change between $y_{t-(1/2)}$ and y_t:

$$y_{t-(1/2)} - y_{t-1} = \epsilon_{1t} \quad y_t - y_{t-(1/2)} = \epsilon_{2t}. \tag{4}$$

Sampled at integer dates $t = 1, 2, \ldots$, the process of (4) has the same properties as (3), since

$$y_t - y_{t-1} = \epsilon_{1t} + \epsilon_{2t} \sim i.i.d.N(0,1).$$

In addition, the process of (4) is defined also at the non-integer dates and remains the property for both integer and non-integer dates that $y_t - y_s \sim N(0, t-s)$ with $y_t - y_s$ independent of the change over any other nonoverlapping interval. Using the same reasoning, we could partition the change between $t-1$ and t into N separate subperiods:

$$y_t - y_{t-1} = \sum_{i=1}^{n} \epsilon_{it}, \quad \epsilon_{it} \sim i.i.d.N\left(0, \frac{1}{n}\right).$$

when $n \to \infty$, the limit process is known as *Brownian motion*. The value of this process at date t is denoted by $W(t)$. A realization of a continuous time process can be viewed as a stochastic function $W(\cdot)$. In particular, we will be interested in Brownian motion over the interval $t \in [0,1]$.

Definition 1 *(Brownian Motion) A standard Brownian motion* $W(t)$, $t \in [0,1]$, *is a continuous time stochastic process such that*

(a) $W(0) = 0$

(b) *For any time* $0 < s < t < 1$, $W(t) - W(s) \sim N(0, t-s)$. *And the differences* $W(t_2) - W(t_1)$ *and* $W(t_4) - W(t_3)$, *for any* $0 \leq t_1 < t_2 < t_3 < t_4 \leq 1$, *are independent.*

(c) *W(t) is continuous in time t with probability 1.*

So given a standard Brownian motion $W(t)$, we have $E(W(t)) = 0$, $Var(W(t)) = t$, and $Cov(W(t), W(s)) = \min(t, s)$. Other Brownian motions can be generated from a standard Brownian motion. For example, the process $Z(t) = \sigma \cdot W(t)$ has independent increments and is distributed $N(0, \sigma^2 t)$. Such a process is described as Brownian motion with variance σ^2.

An important feature of Brownian motion is that although it is continuous in t, it is not differentiable using standard calculus. The direction of change at t is likely to be completely different from that at $t + \Delta$, no matter how small is Δ. Even some parts of the realization of a Brownian motion looks 'smooth', if we see it with a 'microscope', we will see many zig-zags.

There are several concepts of smoothness of a function and continuity is the weakest one. Differentiability is another concept of smoothness. When the domain of the function is an interval, we have another smoothness condtion. A function $f : [a, b] \mapsto \mathbb{R}$ is of bounded variation if $\exists\, M < \infty$ such that for every partition of $[a, b]$ by finite collections of points $a = x_0 < x_1 < x_2 < \ldots < x_n = b$,

$$\sum_{k=1}^{n} |f(x_i) - f(x_{i-1})| \leq M.$$

Brownian motion is not of bounded variation.

Later in this lecture, we will also see integrals of Brownian motion ($\int_0^1 W(r)dr$) and a stochastic integral ($\int_0^1 W(r)dW(r)$, for $r \in [0, 1]$. First, note that $W(r)$ is a Gaussian process, hence $\int_0^1 W(r)dr$ is also a Gaussian process. It is easy to see that $E[\int_0^1 W(r)dr] = 0$. To compute its variance, let $s \leq t$ and write

$$\begin{aligned} E\left[\int_0^1 W(r)dr\right]^2 &= 2\int_0^1 \int_0^r E[W(r)W(s)]dsdr \\ &= 2\int_0^1 \int_0^r sdsdr \\ &= \int_0^1 r^2 dr = \frac{1}{3} \end{aligned}$$

Therefore, $\int_0^1 W(r)dr \sim N(0, 1/3)$. As another exercise, consider the distribution of $W(1) - \int_0^1 W(r)dr$. Again, it is a Gaussian process with zero mean. To compute its variance,

$$\begin{aligned} E\left[W(1) - \int_0^1 W(r)dr\right]^2 &= 1 + \frac{1}{3} - 2E\left[W(1)\int_0^1 W(r)dr\right] \\ &= \frac{4}{3} - 2\int_0^1 rdr = \frac{1}{3}. \end{aligned}$$

To study the stochastic integral, we need a fundamental theorem in stochastic calculus:

Definition 2 *(*Ito's Lemma*) Let X_r be a process given by*

$$dX(r) = udr + vdW(r).$$

Let $g(r, x)$ be a twice continuously differentiable real function. Let

$$Y(r) = g(r, X_r),$$

then

$$dY(r) = \frac{\partial g}{\partial r}(r,x)dr + \frac{\partial g}{\partial x}(r,x)dX(r) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(r,x)\cdot(dX(r))^2,$$

where $(dX(r))^2 = (dX(r))\cdot(dX(r))$ is computed according to the rules

$$dr \cdot dr = dr \cdot dW(r) = dW(r)\cdot dr = 0, \quad dW(r)\cdot dW(r) = dr.$$

Now, choose $X(r) = \sigma W(r)$, $g(r, x) = \frac{1}{2}x^2$. Then

$$Y(r) = g(r, X(r)) = \frac{1}{2}X(r)^2 = \frac{1}{2}\sigma^2 W(r)^2.$$

By Ito's lemma,

$$dY(r) = \frac{\partial g}{\partial r}dr + \frac{\partial g}{\partial x}dX(r) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\cdot(dX(r))^2 = \sigma^2 W(r)dW(r) + \frac{1}{2}(\sigma dW(r))^2 = \sigma^2 W(r)dW(r) + \frac{1}{2}\sigma^2 dr.$$

where the first derivative of g with respect to x gives $X_r = \sigma W(r)$, and $dX_r = \sigma dW(r)$. Hence,

$$d\left(\frac{1}{2}\sigma^2 W(r)^2\right) = \sigma^2 W(r)dW(r) + \frac{1}{2}\sigma^2 dr.$$

Integrate them from 0 to 1,

$$\frac{1}{2}\sigma^2 W(1)^2 = \sigma^2 \int_0^1 W(r)dW(r) + \frac{1}{2}\sigma^2,$$

therefore,

$$\sigma^2 \int_0^1 W(r)dW(r) = \frac{1}{2}\sigma^2(W(1)^2 - 1).$$

2.2 Functional Central Limit Theorem (FCLT)

2.2.1 Introduction

Recall that the central limit theorem (CLT) tells that the sample mean of a stationary process is asymptotically normal and centered at the the population mean. However, if x_t is a random walk, there is no such a thing as 'population mean'. Therefore, to draw inference for processes with unit root, we need a new tool which is called *functional central limit theorem'* (FCLT), the central limit theorem defined on the function spaces. FCLT is important to unit root limit theory just as CLT is important to stationary time series limit theory.

As usual, let n denotes the sample size, and we let $r = t/n$, so $r \in [0, 1]$. And we use the symbol $[nr]$ to denote the largest integer that is less than or equal to nr.

Consider a process $u_t \sim i.i.d.(0, \sigma^2)$ with its mean denoted by $\bar{u}_n$, then CLT tells that $n^{1/2}\bar{u}_n \to N(0, \sigma^2)$. Now consider that given a sample of size n, we calculate the mean of the first half of the sample and throw out the rest of the observations:

$$\bar{u}_{[n/2]} = \frac{1}{[n/2]} \sum_{t=1}^{[n/2]} u_t.$$

This estimator also satisfies the CLT:

$$\sqrt{[n/2]}\bar{u}_{[n/2]} \to N(0, \sigma^2).$$

Moreover, this estimator would be independent of an estimator tha uses only the second half of the sample. More generally, let's construct a new random variable $X_n(r)$ for $r \in [0, 1]$,

$$X_n(r) = (1/n) \sum_{t=1}^{[nr]} u_t.$$

or

$$X_n(r) = \begin{cases} 0 & \text{for} \quad r \in [0, 1/n) \\ u_1/n & \text{for} \quad r \in [1/n, 2/n) \\ (u_1 + u_2)/n & \text{for} \quad r \in [2/n, 3/n) \\ \vdots & \\ (u_1 + u_2 + \ldots + u_n)/n & \text{for} \quad r = 1. \end{cases}$$

It is easy to see that $n^{1/2}X_n(1) = n^{1/2}\bar{u}_n$ and CLT tells that it converges to $N(0, \sigma^2)$. But what $X_n(r)$ converges to as $n \to \infty$? Write

$$n^{1/2}X_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t = \frac{\sqrt{[nr]}}{\sqrt{n}} \frac{1}{\sqrt{[nr]}} \sum_{t=1}^{[nr]} u_t.$$

By CLT we have $[nr]^{-1/2} \sum_{t=1}^{[nr]} u_t \to N(0, \sigma^2)$, while $([nr]/n)^{1/2} \to r^{1/2}$, therefore we have

$$n^{1/2}X_n(r)/\sigma \to N(0, r). \tag{5}$$

Next if we consider the behavior of a sample mean based on observations $[nr_1]$ through $[nr_2]$ for $r_2 > r_1$, this is also asymptotically normal using similar approach,

$$n^{1/2}(X_n(r_2) - X_n(r_1))/\sigma \to N(0, r_2 - r_1)$$

and it is independent of the estimator in (5) for $r < r_1$. Therefore, the sequence of stochastic functions $\{\sqrt{n}X_n(\cdot)/\sigma\}_{n=1}^{\infty}$ has an asymptotic probability law:

$$n^{1/2}X_n(\cdot)/\sigma \to W(\cdot). \tag{6}$$

Note that here $X_n(\cdot)$ is a function, while in (5) $X_n(r)$ is a random variable. The asymptotic result (6) is known as the *functional central limit theorem* (FCLT). Later on, we may also write $n^{1/2}X_n(r) \to W(r)$, but note this does not mean the variable $X_n(r)$ converges to a variable which has $N(0,r)$ distribution, but that the function converges to a stochastic function: the standard Brownian motion.

Evaluated at $r = 1$, the function $X_n(r)$ is just the sample mean. Thus when the function in (6) is evaluated at $t = 1$, we get the conventional CLT:

$$\sqrt{n}X_n(1)/\sigma = \frac{1}{\sigma\sqrt{n}}\sum_{t=1}^{n} u_t \to W(1) \sim N(0,1). \tag{7}$$

2.2.2 Convergence of a random function

In lecture 4, we discussed various convergence and continuous mapping theorem for a random variable. Now, let's define convergence of a random function, such as $X_n(r)$ we defined earlier.

We first define convergence in distribution for a random function. Let $S(\cdot)$ represent a continuous-time stochastic process with $S(r)$ representing its value at some date r for $r \in [0,1]$. Also suppose that for any given realization, $S(\cdot)$ is a continuous function of r with probability 1. For $\{S_n(\cdot)\}_{n=1}^{\infty}$ a sequence of such continuous functions, we say that $S_n(\cdot) \to_d S(\cdot)$ if all of the following hold:

(a) For any finite collection of k particular dates, $0 \le r_1 < r_2 < \ldots < r_k \le 1$, the sequence of k-dimensional random vectors $\{\mathbf{y}_n\}_{n=1}^{\infty}$ converges in distribution to the vector $\mathbf{y}$, where

$$\mathbf{y}_n \equiv \begin{bmatrix} S_n(r_1) \\ S_n(r_2) \\ \vdots \\ S_n(r_k) \end{bmatrix} \quad \mathbf{y} \equiv \begin{bmatrix} S(r_1) \\ S(r_2) \\ \vdots \\ S(r_k) \end{bmatrix};$$

(b) For each $\epsilon > 0$, the probability that $S_n(r_1)$ differs from $S_n(r_2)$ for any dates r_1 and r_2 within δ of each other goes to zero uniformly in n as $\delta \to 0$.

(c) $P(|S_n(0)| > \lambda) \to 0$ uniformly in n as $\lambda \to \infty$.

Next, we will extend convergence in probability for a random function. Let $\{S_n(\cdot)\}_{n=1}^{\infty}$ and $\{V_n(\cdot)\}_{n=1}^{\infty}$ denote sequences of random continuous functions with $S_n : r \in [0,1] \mapsto \mathbb{R}$ and $V_n : r \in [0,1] \mapsto \mathbb{R}$. Define Y_n as:

$$Y_n = \sup_{r\in[0,1]} |S_n(r) - V_n(r)|.$$

Then Y_n is a sequence of random variables. If $Y_n \to_p 0$ (this is the usual convergence in probability for a random variable), then we have that

$$S_n(\cdot) \to_p V_n(\cdot).$$

In other words, we define convergence in probability of a random function in terms of convergence of the upper bound of its distance from the limit function. Further, if $V_n(\cdot) \to_p S_n(\cdot)$ and $V_n(\cdot) \to_d V(\cdot)$ where $S(\cdot)$ is a continuous function, then $V_n(\cdot) \to_d S(\cdot)$.

Example 1 Let u_t be strictly stationary time series with finite fourth moment, and let $S_n(r) = n^{-1/2}u_{[nr]}$. Then $S_n(\cdot) \to_p 0$. Proof:

$$P\left(\sup_{r\in[0,1]} |S_n(r)| > \delta\right)$$

$$
\begin{aligned}
&= P\left\{[|n^{-1/2}u_1| > \delta], \text{ or } [|n^{-1/2}u_2| > \delta], \ldots \text{ or } [|n^{-1/2}u_n| > \delta]\right\} \\
&\leq nP(|n^{-1/2}u_t| > \delta) \\
&\leq n\frac{E(n^{-1/2}u_t)^4}{\delta^4} \\
&= \frac{E(u_t^4)}{n\delta^4} \\
&\to 0.
\end{aligned}
$$

So we ave $S_n(\cdot) \to_p 0$.

In Lecture 4, we also reviewed that the continuous mapping theorem (CMT) tells that if $x_n \to x$, and $g(\cdot)$ is a continuous function, then we have $g(x_n) \to g(x)$. We have a similar results for the FCLT. If $S_n(\cdot) \to S(\cdot)$, and $g(\cdot)$ is a continuous functional, then $g(S_n(\cdot)) \to g(S(\cdot))$. For example, $\sqrt{n}X_n(\cdot)/\sigma \to_d W(\cdot)$ implies that

$$\sqrt{n}X_n(\cdot) \to_d \sigma W(\cdot) \sim N(0, \sigma^2 r). \tag{8}$$

As another example, let

$$S_n(r) \equiv [\sqrt{n}X_n(r)]^2. \tag{9}$$

Since $\sqrt{n}X_n(r) \to_d \sigma W(\cdot)$, it follows that

$$S_n(\cdot) \to_d \sigma^2[W(\cdot)]^2. \tag{10}$$

2.3 Applications to Unit Root Processes

The simplist case to illustrate how to use FCLT to compute the asymptotics is to consider a random walk y_t with $y_0 = 0$,

$$y_t = y_{t-1} + u_t = \sum_{i=1}^{t} u_i, \quad u_t \sim i.i.d. N(0, \sigma^2).$$

Define $X_n(\cdot)$ as:

$$X_n(r) = \begin{cases} 0 & \text{for} \quad r \in [0, 1/n) \\ y_1/n & \text{for} \quad r \in [1/n, 2/n) \\ y_2/n & \text{for} \quad r \in [2/n, 3/n) \\ \vdots & \\ y_n/n & \text{for} \quad r = 1. \end{cases} \tag{11}$$

If we integrate $X_n(r)$ over $r \in [0, 1]$, we have

$$
\begin{aligned}
\int_0^1 X_n(r)dr &= y_1/n^2 + y_2/n^2 + \ldots + y_{n-1}/n^2 \\
&= n^{-2}\sum_{t=1}^{n} y_{t-1}
\end{aligned}
$$

Multiplying its both sides by $\sqrt{n}$:

$$\int_0^1 \sqrt{n}X_n(r)dr = n^{-3/2}\sum_{t=1}^{n} y_{t-1}.$$

From (8) we know that $\sqrt{n}X_n(\cdot) \to_d \sigma W(\cdot)$, by CMT,

$$\int_0^1 \sqrt{n}X_n(r)dr \to_d \sigma \int_0^1 W(r)dr.$$

therefore, we got the limit for $n^{-3/2}\sum_{t=1}^{n-1} y_t$,

$$n^{-3/2}\sum_{t=1}^{n} y_{t-1} \to \sigma \int_0^1 W(r)dr. \tag{12}$$

Thus, when y_t is a driftless random walk, its sample mean $\frac{1}{n}\sum_{t=1}^n y_t$ diverges but $n^{-3/2}\sum_{t=1}^n y_t$ converges. An alternative way to find the limit distribution of $n^{-3/2}\sum_{t=1}^n y_t$ follows:

$$\begin{aligned} n^{-3/2}\sum_{t=1}^{n} y_{t-1} &= n^{-3/2}[u_1 + (u_1+u_2) + \ldots + (u_1+u_2+\ldots+u_{n-1})] \\ &= n^{-3/2}[(n-1)u_1 + (n-2)u_2 + \ldots + u_{n-1}] \\ &= n^{-3/2}\sum_{t=1}^{n}(n-t)u_t \\ &= n^{-1/2}\sum_{t=1}^{n} u_t - n^{-3/2}\sum_{t=1}^{n} tu_t \end{aligned}$$

while from the previous lecture, we know that

$$\begin{bmatrix} n^{-1/2}\sum_{t=1}^n u_t \\ n^{-3/2}\sum_{t=1}^n tu_t \end{bmatrix} \to_d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}\right). \tag{13}$$

Therefore $n^{-3/2}\sum_{t=1}^n y_t$ is asymptotically Gaussian with mean zero and variance equal to $\sigma^2[1-2(1/2)+1/3] = \sigma^2/3$. From this expression we also have

$$\begin{aligned} n^{-3/2}\sum_{t=1}^{n} tu_t &= n^{-1/2}\sum_{t=1}^{n} u_t - n^{-3/2}\sum_{t=1}^{n} y_{t-1} \\ &\to \sigma W(1) - \sigma\int_0^1 W(r)dr \end{aligned} \tag{14}$$

Using similar methods we could compute the asymptotic distribution of the sum of squares of a random walk. Define

$$S_n(r) = n[X_n(r)]^2.$$

and it can be written as

$$S_n(r) = \begin{cases} 0 & \text{for} \quad r \in [0, 1/n) \\ y_1^2/n & \text{for} \quad r \in [1/n, 2/n) \\ y_2^2/n & \text{for} \quad r \in [2/n, 3/n) \\ \vdots & \\ y_n^2/n & \text{for} \quad r = 1 \end{cases} \tag{15}$$

Again we compute the sum

$$\int_0^1 S_n(r)dr = y_1^2/n^2 + y_2^2/n^2 + \ldots + y_{n-1}^2/n^2.$$

Since we have that $S_n(r) \to \sigma^2 W(r)^2$, by CMT,

$$n^{-2}\sum_{t=1}^{n} y_{t-1}^2 \to \sigma^2 \int_0^1 [W(r)]^2 dr. \tag{16}$$

If we make use of $n^{-3/2}\sum_{t=1}^{n} y_{t-1} \to_d \sigma \int_0^1 W(r)dr$ and for $r = t/n$, we also have

$$n^{5/2}\sum_{t=1}^{n} ty_{t-1} = n^{-3/2}\sum_{t=1}^{n}(t/n)y_{t-1} \to_d \sigma \int_0^1 rW(r)dr. \tag{17}$$

Similarly, for $r = t/n$ and we use (16) to get

$$n^{-3}\sum_{t=1}^{n} ty_{t-1}^2 = n^{-2}\sum_{t=1}^{n}(t/n)y_{t-1}^2 \to_d \sigma^2 \int_0^1 r[W(r)]^2 dr. \tag{18}$$

Another useful result is

$$n^{-1}\sum_{t=1}^{n} y_{t-1}u_t \to (1/2)\sigma^2[W(1)^2 - 1].$$

Proof: first

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2,$$

so

$$\begin{aligned} & n^{-1}\sum_{t=1}^{n} y_{t-1}u_t \\ = \; & n^{-1}(1/2)\sum_{t=1}^{n}(y_t^2 - y_{t-1}^2) - n^{-1}(1/2)\sum_{t=1}^{n} u_t^2 \\ = \; & n^{-1}(1/2)y_n^2 - n^{-1}(1/2)\sum_{t=1}^{n} u_t^2 \end{aligned}$$

By (6), we have $n^{-1}y_n \to \sigma W(1)$, by CMT, we have

$$n^{-1}(1/2)y_n^2 \to (1/2)\sigma^2 W(1)^2.$$

By LLN,

$$n^{-1}(1/2)\sum_{t=1}^{n} u_t^2 \to (1/2)\sigma^2,$$

Therefore,

$$n^{-1}\sum_{t=1}^{n} y_{t-1}u_t \to (1/2)\sigma^2[W(1)^2 - 1] \tag{19}$$

3 Unit Root Tests

3.1 Unit Root Tests with $i.i.d$ Error

The asymptotics of a a random walk with *i.i.d.* shocks is summarized in the following proposition. The number in bracket shows where the result is first introduced and proved.

Proposition 1 *Suppose that ξ_t follows a random walk without drift,*

$$\xi_t = \xi_{t-1} + u_t, \quad \xi_0 = 0, \quad u_t \sim i.i.d(0, \sigma^2).$$

Then

(a) $n^{-1/2} \sum_{t=1}^n u_t \to_d \sigma W(1)$ [7];

(b) $n^{-1} \sum_{t=1}^n \xi_{t-1} u_t \to_d \frac{1}{2}\sigma^2 [W(1)^2 - 1]$ [19];

(c) $n^{-3/2} \sum_{t=1}^n t u_t \to_d \sigma W(1) - \sigma \int_0^1 W(r)dr$ [14];

(d) $n^{-3/2} \sum_{t=1}^n \xi_{t-1} \to_d \sigma \int_0^1 W(r)dr$ [12];

(e) $n^{-2} \sum_{t=1}^n \xi_{t-1}^2 \to_d \sigma^2 \int_0^1 W(r)^2 dr$ [16];

(f) $n^{-5/2} \sum_{t=1}^n t\xi_{t-1} \to_d \sigma \int_0^1 rW(r)dr$ [17];

(g) $n^{-3} \sum_{t=1}^n t\xi_{t-1}^2 \to_d \sigma^2 \int_0^1 rW(r)^2 dr$ [18].

(h) $n^{-v+1} \sum_{t=1}^n t^v \to 1/(v+1)$ *for $v = 0, 1, 2, \ldots$ [lecture 7]*

Note that all those $W(\cdot)$ is the same Brownian motion, so all those results are correlated. If we are not interested in their correlations, we can find simpler expressions for them. For example, (a) is just $N(0, \sigma^2)$, (b) is $(1/2)\sigma^2[\chi^2(1) - 1]$, (c) and (d) are $N(0, \sigma^2/3)$.

In general, the correspondence between the finite sample and their limits are like $\sum_{t=1}^n \to \int_0^1$, $(t/n) \to r$, $(1/n) \to dr$, $n^{-1/2}u_t \to dW$, etc. Take (h) as an example, and let $v = 2$. From previous lecture we know that $n^{-3} \sum_{t=1}^n t^2 \to 1/3$. Using the correspondence here, we have

$$n^{-3} \sum_{t=1}^n t^2 = n^{-1} \sum_{t=1}^n (t/n)^2 \to \int_0^1 r^2 dr = 1/3.$$

3.1.1 Case 1

Suppose that the data generating process (DGP) is a random walk, and we are estimating the parameter ρ by OLS in the regression

$$y_t = \rho y_{t-1} + u_t, \quad u_t \sim i.i.d(0, \sigma^2), \tag{20}$$

where $\rho = 1$ and we are interested in the asymptotic distributions of the OLS estimates $\hat{\rho}_n$:

$$\begin{aligned}\hat{\rho}_n &= \frac{\sum_{t=1}^n y_{t-1}y_t}{\sum_{t=1}^n y_{t-1}^2} \\ &= \frac{\sum_{t=1}^n y_{t-1}(y_{t-1}+u_t)}{\sum_{t=1}^n y_{t-1}^2} \\ &= 1 + \frac{\sum_{t=1}^n y_{t-1}u_t}{\sum_{t=1}^n y_{t-1}^2}\end{aligned}$$

Then

$$n(\hat{\rho}_n - 1) = \frac{n^{-1}\sum_{t=1}^n y_{t-1}u_t}{n^{-2}\sum_{t=1}^n y_{t-1}^2}.$$

By (19) (result b), (16) (result e) and CMT, we have

$$n(\hat{\rho} - 1) \rightarrow \frac{W(1)^2 - 1}{2\int_0^1 W(r)^2 dr}. \tag{21}$$

First we note that $(\hat{\rho}_n - 1)$ converges at the order of n, instead of $n^{1/2}$, as in the cases when $|\rho| < 1$. Therefore when the true coefficient is unity, $\hat{\rho}_n$ is superconsistent. Second, since $W(1) \sim N(0,1)$, $W(1)^2 \sim \chi^2(1)$. The probability that $\chi^2(1)$ is less than one is 0.68, therefore with probability 0.68 $n(\hat{\rho}_n - 1)$ will be negative, which implies that its limit distribution is skewed to the left. Recall that in the AR(1) regression with $|\rho| < 1$, the estimate $\hat{\rho}_n$ is downward biased. However, its limit distribution $\sqrt{n}(\hat{\rho}_n - \rho)$ is still symmetric around zero. While when the true value of ρ is unity, even the limit distribution of $n(\hat{\rho} - 1)$ is asymmetric with negative values twice as likely as positive values.

In practice, critical values for the random variable in (21) are found by computing the exact finite sample distribution of $n(\hat{\rho} - 1)$ assuming u_t is Gaussian. Then the critical value can be tabulated by Monte Carlo or by numerical approximation.

There are two commonly used approaches to test the hypothesis that $\rho_0 = 1$: Dickey-Fuller ρ-test and Dickey-Fuller t-test. The DF ρ-test is to compute the statistics $n(\hat{\rho}_n - 1)$ and compare the statistics with the critical values from the distribution in (21). The advantage of this approach is that we don't need to compute its standard deviation. Alternatively, we could use DF t-test which is based on the usual t statistics,

$$t_n = \frac{\hat{\rho}_n - 1}{\hat{\sigma}_{\hat{\rho}}}. \tag{22}$$

where $\hat{\sigma}_{\hat{\rho}}$ is the standard deviation of OLS estimated coefficient,

$$\hat{\sigma}_{\hat{\rho}}^2 = \frac{s_n^2}{\sum_{t=1}^n y_{t-1}^2}, \tag{23}$$

and

$$s_n^2 = \frac{1}{n}\sum_{t=1}^n (y_t - \hat{\rho}_n y_{t-1})^2.$$

Plug (23) into (22), we have

$$t_n = \frac{n^{-1}\sum_{t=1}^{n} y_{t-1}u_t}{\left(n^{-2}\sum_{t=1}^{n} y_{t-1}^2\right)^{1/2}(s_n^2)^{1/2}}.$$

If $\hat{\rho} \to \rho = 1$, which is true for OLS estimator in the present problem, then $s_n^2 \to \sigma^2$ by LLN. And by (19) and (16), we have the limit for t_n,

$$t_n \to \frac{(1/2)\sigma^2[W(1)^2 - 1]}{\left[\sigma^2\int_0^1 W(r)^2 dr\right]^{1/2}[\sigma^2]^{1/2}} = \frac{W(1)^2 - 1}{2\left(\int_0^1 W(r)^2 dr\right)^{1/2}}. \tag{24}$$

For the same reason as in (21), this t-statistics is asymmetric and skewed to the left.

3.1.2 Case 2

The DGP is still a random walk as in case 1 (20),

$$y_t = y_{t-1} + u_t, \quad u_t \sim i.i.d(0, \sigma^2),$$

but we include a constant term in the regression

$$y_t = \hat{\alpha} + \hat{\rho}y_{t-1} + \hat{u}_t.$$

The OLS estimates for the coefficients

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

Under the null hypothesis $H_0 : \alpha = 0, \rho = 1$, the deviations of the estimate vector from the hypothesis

$$\begin{bmatrix} \hat{\alpha}_n \\ \hat{\rho}_n - 1 \end{bmatrix} = \begin{bmatrix} n & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum u_t \\ \sum u_{t-1}y_t \end{bmatrix}. \tag{25}$$

Recall in a regression with a constant and time trend, the estimates have different convergent rates. The situation is similar in this case. The order in probability for each terms are

$$\begin{bmatrix} \hat{\alpha}_n \\ \hat{\rho}_n \end{bmatrix} = \begin{bmatrix} O_p(n) & O_p(n^{3/2}) \\ O_p(n^{3/2}) & O_p(n^2) \end{bmatrix}^{-1} \begin{bmatrix} O_p(n^{1/2}) \\ O_p(n) \end{bmatrix}. \tag{26}$$

As we did before, now we need a rescaling matrix

$$H_n = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n \end{bmatrix}.$$

Premultiply (25) by H_n we have

$$\begin{bmatrix} n^{1/2}\hat{\alpha}_n \\ n(\hat{\rho}_n - 1) \end{bmatrix} = \begin{bmatrix} 1 & n^{-3/2}\sum y_{t-1} \\ n^{-3/2}\sum y_{t-1} & n^{-2}\sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum u_t \\ n\Sigma u_{t-1}y_t \end{bmatrix}. \tag{27}$$

By result (d) and (e) we have

$$\begin{bmatrix} 1 & n^{-3/2}\sum y_{t-1} \\ n^{-3/2}\sum y_{t-1} & n^{-2}\sum y_{t-1}^2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \sigma\int W(r)dr \\ \sigma\int W(r)dr & \sigma^2\int W(r)^2dr \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2dr \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}$$

and by result (a) and (b) we have

$$\begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-1}\sum u_{t-1}y_t \end{bmatrix} \rightarrow \begin{bmatrix} \sigma W(1) \\ (1/2)\sigma^2[W(1)^2-1] \end{bmatrix} = \sigma \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \begin{bmatrix} W(1) \\ (1/2)[W(1)^2-1] \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \begin{bmatrix} n^{1/2}\hat{\alpha}_n \\ n^1(\hat{\rho}_n-1) \end{bmatrix} &\rightarrow_d \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2dr \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ (1/2)[W(1)^2-1] \end{bmatrix} \\ &= \Delta^{-1} \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \int W(r)^2dr & -\int W(r)dr \\ -\int W(r)dr & 1 \end{bmatrix} \begin{bmatrix} W(1) \\ (1/2)[W(1)^2-1] \end{bmatrix} \\ &= \Delta^{-1} \begin{bmatrix} \sigma & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} W(1)\int W(r)^2dr - (1/2)[W(1)^2-1]\int W(r)dr \\ (1/2)[W(1)^2-1] - W(1)\int W(r)dr \end{bmatrix} \end{aligned}$$

where

$$\Delta \equiv \int W(r)^2dr - \left[\int W(r)dr\right]^2.$$

So the DF ρ statistics to test the null hypothesis that $\rho = 1$ has the following limit distribution

$$n(\hat{\rho}_n - 1) \rightarrow_d \frac{(1/2)[W(1)^2-1] - W(1)\int W(r)dr}{\int W(r)^2dr - \left[\int W(r)dr\right]^2}. \tag{28}$$

As in case 1, we can also use a t test,

$$t_n = \frac{\hat{\rho}_n - 1}{\hat{\sigma}_{\hat{\rho}_n}},$$

which converges to

$$\frac{(1/2)[W(1)^2-1] - W(1)\int W(r)dr}{\left\{\int W(r)^2dr - \left[\int W(r)dr\right]^2\right\}^{1/2}}.$$

The details can be found on page 493-494.

3.1.3 Case 3

Now, suppose that the true process is a random walk with drift:

$$y_t = \alpha + y_{t-1} + u_t, \quad u_t \sim i.i.d(0, \sigma^2).$$

Without loss of generality, we could set $y_0 = 0$. And we also estimate a linear regression with a constant,

$$y_t = \hat{\alpha} + \hat{\rho} y_{t-1} + \hat{u}_t.$$

Define

$$\xi_t \equiv u_1 + u_2 + \ldots + u_t,$$

then

$$y_t = \alpha t + \xi_t$$

and

$$\sum_{t=1}^{n} y_{t-1} = \alpha \sum_{t=1}^{n} t + \sum_{t=1}^{n} \xi_{t-1}.$$

Notice that these two terms have different divergent rates. We know that $\sum_{t=1}^{n} t = n(n+1)/2 = O(n^2)$, while $\sum_{t=1}^{n} \xi_{t-1} = O_p(n^{3/2})$ as $n^{-3/2} \sum_{t=1}^{n} \xi_{t-1}$ converges to a normal distribution with finite variance (result (d)). Therefore, pick the fastest divergent rate,

$$n^{-2} \sum_{t=1}^{n} y_{t-1} = \alpha n^{-2} \sum_{t=1}^{n} t + n^{-1/2} n^{-3/2} \sum_{t=1}^{n} \xi_{t-1} \to_p \alpha/2. \tag{29}$$

Similarly, $\sum_{t=1}^{n} y_{t-1}^2$, $\sum_{t=1}^{n} y_{t-1} u_t$ also have terms with different divergent rates:

$$\begin{aligned} \sum_{t=1}^{n} y_{t-1}^2 &= \sum_{t=1}^{n} [\alpha(t-1) + \xi_{t-1}]^2 \\ &= \alpha^2 \sum_{t=1}^{n} (t-1)^2 + \sum_{t=1}^{n} \xi_{t-1}^2 + 2\alpha \sum_{t=1}^{n} (t-1)\xi_{t-1} \end{aligned}$$

where $\sum_{t=1}^{n}(t-1)^2 = O_p(n^3)$ (result (h)), $\sum_{t=1}^{n} \xi_{t-1}^2 = O_p(n^2)$ (result (e)), and $\sum_{t=1}^{n}(t-1)\xi_{t-1} = O_p(n^{5/2})$ (result (f)). Norm the sequence with the inverse of the fastest divergent rate n^3,

$$n^{-3} \sum_{t=1}^{n} y_{t-1}^2 \to \alpha^2/3. \tag{30}$$

Finally,

$$\sum_{t=1}^{n} y_{t-1} u_t = \sum_{t=1}^{n} [\alpha(t-1) + \xi_{t-1}] u_t = \alpha \sum_{t=1}^{n} (t-1) u_t + \sum_{t=1}^{n} \xi_{t-1} u_t,$$

where $\sum_{t=1}^{n}(t-1)u_t = O_p(n^{3/2})$ (result (c)) and $\sum_{t=1}^{n} \xi_{t-1} u_t = O_p(n)$ (result (b)). Again norm it with the fastest divergent rate

$$n^{-3/2} \sum_{t=1}^{n} y_{t-1} u_t \to_p n^{-3/2} \sum_{t=1}^{n} \alpha(t-1) u_t. \tag{31}$$

Corresponding to the different rates, to derive a nondegenerate limit distribution for the estimates, again we need a scaling matrix. In this case, we need

$$H_n = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{3/2} \end{bmatrix}.$$

Premultiply the OLS estimator vector (in deviations from their true value) with H_n we got

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha) \\ n^{3/2}(\hat{\rho}_n - 1) \end{bmatrix} = \begin{bmatrix} 1 & n^{-2}\sum y_{t-1} \\ n^{-2}\sum y_{t-1} & n^{-3}\sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-3/2}\sum y_{t-1}u_t \end{bmatrix}.$$

From (29) and (30), the first term

$$\begin{bmatrix} 1 & n^{-2}\sum y_{t-1} \\ n^{-2}\sum y_{t-1} & n^{-3}\sum y_{t-1}^2 \end{bmatrix} \to_p \begin{bmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{bmatrix} \equiv Q.$$

From (13) and (31), we have

$$\begin{aligned} \begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-3/2}\sum y_{t-1}u_t \end{bmatrix} &\to_p \begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-3/2}\sum_{t=1}^n \alpha(t-1)u_t \end{bmatrix} \\ &\to_p N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{bmatrix} \right) = N(0, \sigma^2 Q). \end{aligned}$$

Therefore we have the following limit distribution for the OLS estimates

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha) \\ n^{3/2}(\hat{\rho}_n - 1) \end{bmatrix} \to_d N(0, Q^{-1} \cdot \sigma^2 Q \cdot Q^{-1}) = N(0, \sigma^2 Q^{-1}). \tag{32}$$

So in case 3, both estimated coefficients are asymptotically Gaussian, and the asymptotic distribution is the same as $\hat{\alpha}$ and $\hat{\delta}$ in the regression with deterministic trends. This is because here y_t has two components: a deterministic time trend and random walk, and the time trend dominates the random walk.

3.1.4 Case 4

Finally we consider that the true process is a random walk with or without drift,

$$y_t = \alpha + y_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \sigma^2),$$

where α may or may not be zero, and we run the following regression

$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t. \tag{33}$$

Without loss of generality, we assume that $y_0 = 0$. Note that when $\alpha \neq 0$, it is also a time trend, hence there will be an asymptotic collinear problem between y_t and t. Hence rewrite the regression as

$$\begin{aligned} y_t &= (1-\rho)\alpha + \rho[y_{t-1} - \alpha(t-1)] + (\delta + \rho\alpha)t + u_t \\ &= \alpha^* + \rho^* \xi_{t-1} + \delta^* t + u_t \end{aligned}$$

where $\alpha^* = (1-\rho)\alpha$, $\rho^* = \rho$, $\delta^* = (\delta + \rho\alpha)$, and $\xi_t = y_t - \alpha t$. With this transformation, under the null hypothesis $\rho = 1$, $\delta = 0$, ξ_t is a random walk:

$$\xi_t = u_1 + u_2 + \ldots + u_t.$$

Therefore, with this transformation, we regress y_t on a constant, a driftless random walk, and a deterministic time trend.

The OLS estimates in this regression are

$$\begin{bmatrix} \hat{\alpha}_n^* \\ \hat{\rho}_n^* \\ \hat{\delta}_n^* \end{bmatrix} = \begin{bmatrix} n & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum \xi_{t-1} t \\ \sum t & \sum \xi_{t-1} t & \sum t^2 \end{bmatrix} \begin{bmatrix} \sum y_t \\ \sum \xi_{t-1} y_t \\ \sum t y_t \end{bmatrix}.$$

The hypothesis is that $\alpha = c$, any constant, $\rho = 1$ and $\delta = 0$. Correspondingly, in the transformed system $\alpha^* = 0$, $\rho^* = 1$, and $\delta^* = c$. The deviations of the estimates from these true values are given by

$$\begin{bmatrix} \hat{\alpha}_n^* \\ \hat{\rho}_n^* - 1 \\ \hat{\delta}_n^* - c \end{bmatrix} = \begin{bmatrix} n & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum \xi_{t-1} t \\ \sum t & \sum \xi_{t-1} t & \sum t^2 \end{bmatrix} \begin{bmatrix} \sum u_t \\ \sum \xi_{t-1} u_t \\ \sum t u_t \end{bmatrix}. \tag{34}$$

Note that these three estimates have different convergent rates (we are already familiar with them!) α_n^* is $n^{1/2}$ convergent, $\hat{\rho}_n^*$ is n convergent, and $\hat{\delta}_n^*$ is $n^{3/2}$ convergent. Therefore we need a rescaling matrix

$$H_n = \begin{bmatrix} n^{1/2} & 0 & 0 \\ 0 & n & 0 \\ 0 & 0 & n^{3/2} \end{bmatrix}.$$

Premultiply (34) with H_n we have that

$$\begin{bmatrix} n^{1/2}\hat{\alpha}_n^* \\ n(\hat{\rho}_n^* - 1) \\ n^{3/2}(\hat{\delta}_n^* - c) \end{bmatrix} = \begin{bmatrix} 1 & n^{-3/2}\sum \xi_{t-1} & n^{-2}\sum t \\ n^{-3/2}\sum \xi_{t-1} & n^{-2}\sum \xi_{t-1}^2 & n^{-5/2}\sum \xi_{t-1} t \\ n^{-2}\sum t & n^{-5/2}\sum \xi_{t-1} t & n^{-3}\sum t^2 \end{bmatrix} \begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-1}\sum \xi_{t-1} u_t \\ n^{-3/2}\sum t u_t \end{bmatrix}.$$

The limit distribution of each term in the above equation can be found in the proposition. Plug them in and we get

$$\begin{bmatrix} n^{1/2}\hat{\alpha}_n^* \\ n(\hat{\rho}_n^* - 1) \\ n^{3/2}(\hat{\delta}_n^* - c) \end{bmatrix} \to_d \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr & \frac{1}{2} \\ \int W(r)dr & \int W(r)^2 dr & \int rW(r)dr \\ \frac{1}{2} & \int rW(r)dr & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ (1/2)[W(r)^2 - 1] \\ W(1) - \int W(r)dr \end{bmatrix}. \tag{35}$$

The DF unit root ρ test in this case is given by the middle row of (35). Note that it does not depend on either σ or α. The DF t test can be derived in a similar way (see page 500 in Hamilton).

3.2 Unit Root Tests with Serially Correlated Errors

3.2.1 BN Decomposition and Phillips-Solo Device

Beveridge and Nelson (1981) proposed that any time series that displays some degree of nonstationarity can be decomposed into two additive parts: a stationary (also called cyclitical or transitory) part and a nonstationary (also called long-run or permanent) part. Let

$$u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \tag{36}$$

where (a) $\epsilon_t \sim WN(0, \sigma^2)$ and (b) $\sum_{j=0}^{\infty} j \cdot |c_j| < \infty$. The BN-decomposition tells that we could rewrite the lag operator as

$$C(L) = C(1) + (L-1)\tilde{C}(L)$$

where $C(1) = \sum_{j=0}^{\infty} c_j$, $\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$, and $\tilde{c}_j = \sum_{j+1}^{\infty} c_k$. Since we assume that $\sum_{j=0}^{\infty} j \cdot |c_j| < \infty$, we have $\sum_{j=0}^{\infty} |\tilde{c}_j| < \infty$. Phillips and Solo (1992) verified that with conditions (a), (b), and that $C(1) \neq 0$, u_t can be represented in the form of

$$\begin{aligned} u_t &= (C(1) + (L-1)\tilde{C}(L))\epsilon_t \\ &= C(1)\epsilon_t - \tilde{C}(L)(\epsilon_t - \epsilon_{t-1}) \end{aligned}$$

Then for a random walk process with innovations u_t, we could represent it as

$$\begin{aligned} y_t &= y_{t-1} + u_t \\ &= y_0 + \sum_{j=1}^{t} u_j \\ &= y_0 + \sum_{j=1}^{t} C(1)\epsilon_j - \tilde{C}(L)\sum_{j=0}^{t}(\epsilon_j - \epsilon_{j-1}) \\ &= y_0 + C(1)\sum_{j=1}^{t}\epsilon_j - \tilde{C}(L)\epsilon_t + \tilde{C}(L)\epsilon_0 \\ &= y_0 + \eta_0 - \eta_t + C(1)\sum_{j=1}^{t}\epsilon_j \end{aligned}$$

where $\eta_0 = \tilde{C}(L)\epsilon_0$ is the initial condition, $\eta_t = \tilde{C}(L)\epsilon_t = \sum_{j=1}^{\infty} \tilde{c}_j \epsilon_{t-j}$ is a stationary process (note that $\tilde{c}_j$ is absolutely summable), and $C(1)\sum_{j=1}^{t} \epsilon_j$ is a nonstationary random walk process.

Rewrite y_t as

$$y_t = \sum_{s=1}^{t} u_s = C(1)\sum_{s=1}^{t} \epsilon_s + \theta\epsilon_0 - \theta\epsilon_t.$$

Note that $\xi_t = \sum_{s=1}^{t} \epsilon_s$ is a random walk with serially uncorrelated error and we have that $n^{-1/2}\xi_{[nr]} \to \sigma W(r)$, while $\epsilon_0 - \epsilon_t$ are bounded in probability, hence we would expect that $n^{-1/2} y_t = C(1)n^{-1/2}\xi_t + o_p(1) \to \lambda W(r)$. The following proposition summarizes some important limit theories for unit root process with serially correlated error.

Proposition 2 *Let $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, where $\sum_{j=0}^{\infty} j \cdot |c_j| < \infty$ and $\epsilon \sim i.i.d.(0, \sigma^2, \mu_4)$. Define that*

$$\gamma_h = E(u_t u_{t-h}) = \sigma^2 \sum_{j=0}^{\infty} c_j c_{j+h},$$

$$\lambda = \sigma \sum_{j=0}^{\infty} c_j = \sigma C(1),$$

$$\xi_t = u_1 + u_2 + \ldots u_t, \quad \xi_0 = 0.$$

In the above notation, λ^2 is known as the long run variance of u_t, which is in gerenal different from the variance of u_t, which is γ_0.

(a) $n^{-1/2}\sum_{t=1}^{n} u_t \to_d \lambda \cdot W(1)$;

(b) $n^{-1/2}\sum_{t=1}^{n} u_{t-j}\epsilon_t \to_d N(0, \sigma^2\gamma_0)$ for $j = 1, 2, \ldots$;

(c) $n^{-1}\sum_{t=1}^{n} u_t u_{t-j} \to \gamma_j$ for $j = 1, 2, \ldots$;

(d) $n^{-1}\sum_{t=1}^{n} \xi_{t-1}\epsilon_t \to_d (1/2)\sigma\lambda[W(1)^2 - 1]$;

(e)
$$n^{-1}\sum_{t=1}^{n} \xi_{t-1}u_{t-h} \to_d \begin{cases} (1/2)[\lambda^2[W(1)^2 - \gamma_0] & \text{for} \quad h = 0 \\ (1/2)[\lambda^2[W(1)^2 - \gamma_0] + \sum_{j=0}^{h-1}\gamma_h & \text{for} \quad h = 1, 2, \ldots \end{cases}$$

(f) $n^{-3/2}\sum_{t=1}^{n} \xi_{t-1} \to_d \lambda \int_0^1 W(r)dr$;

(g) $n^{3/2}\sum_{t=1}^{n} tu_{t-j} \to_d \lambda \left[W(1) - \int_0^1 W(r)dr\right]$;

(h) $n^{-2}\sum_{t=1}^{n} \xi_{t-1}^2 \to_d \lambda^2 \int_0^1 W(r)^2 dr$;

(i) $n^{-5/2}\sum_{t=1}^{n} t\xi_{t-1} \to_d \lambda \int_0^1 rW(r)dr$;

(j) $n^{-3}\sum_{t=1}^{n} t\xi_{t-1}^2 \to_d \lambda^2 \int_0^1 rW(r)^2 dr$;

(k) $n^{-(v+1)}\sum_{t=1}^{n} t^v \to 1/(v+1)$ for $v = 0, 1, 2, \ldots$.

The proof of all these results can be found in the appendix of Chapter 17 in Hamilton. In the class, we will discuss (a), (e) and (f) as examples. First to prove (a),

$$n^{-1/2}\sum_{t=1}^{[nr]} u_t = n^{-1/2}C(1)\sum_{t=1}^{[nr]} \epsilon_t + n^{-1/2}(\eta_{[nr]} - \eta_0).$$

By (6) and CMT we have that

$$n^{-1/2}C(1)\sum_{t=1}^{[nr]} \epsilon_t \to \sigma C(1)W(r).$$

η_0 is the initial condition and $\eta_{[nr]}$ is a zero mean stationary process, both are bounded in probability,

$$n^{-1/2}(\eta_{[nr]} - \eta_0) \to 0.$$

Therefore, we obtain the limit

$$n^{-1/2}\sum_{t=1}^{[nr]} u_t \to \sigma C(1)W(r)$$

and when $r = 1$,

$$n^{-1/2}\sum_{t=1}^{n} u_t \to \sigma C(1)W(1) \tag{37}$$

Second, to prove

$$n^{-1}\sum_{t=1}^{n}\xi_{t-1}u_{t-j}\rightarrow\left\{\begin{array}{ll}(1/2)[\lambda^2(W(1)^2-\gamma_0)] & \text{for}\quad j=0\\(1/2)[\lambda^2(W(1)^2-\gamma_0)]+\sum_{i=0}^{j-1}\gamma_i & \text{for}\quad j>0\end{array}\right. \tag{38}$$

First, let $h=0$ and we have

$$n^{-1}\sum_{t=1}^{n}\xi_{t-1}u_t=n^{-1}(1/2)\xi_n^2-n^{-1}(1/2)\sum_{t=1}^{n}u_t^2.$$

We know that $n^{-1/2}\xi_n\rightarrow\lambda W(1)$. By CMT, we have that

$$(1/2)n^{-1}\xi_n^2\rightarrow(1/2)\lambda^2W(1)^2.$$

In result (c), we have $n^{-1}\sum_{t=1}^{n}u_t^2\rightarrow\gamma_0$. Therefore,

$$n^{-1}\sum_{t=1}^{n}\xi_{t-1}u_t\rightarrow(1/2)[\lambda^2(W(1)^2-\gamma_0)].$$

Next, let $h=1$. Note that

$$\xi_{t-1}u_{t-1}=(\xi_{t-2}+u_{t-1})u_{t-1}=\xi_{t-2}u_{t-1}+u_{t-1}u_{t-1}.$$

We already got the limit for $n^{-1}\sum_{t-1}^{n}\xi_{t-1}u_t$, therefore,

$$n^{-1}\sum_{t=1}^{n}\xi_{t-1}u_{t-1}\rightarrow(1/2)[\lambda^2(W(1)^2-\gamma_0)]+\gamma_0.$$

Similar for $h=2,3,\ldots$.

Thirdly, consider result (f),

$$n^{-3/2}\sum_{t=1}^{n}\xi_{t-1}\rightarrow\lambda\int_0^1W(r)dr. \tag{39}$$

Define

$$S_n(r)=\left\{\begin{array}{ll}0 \quad \text{for} \quad r\in[0,1/n)\\ n^{-1/2}\xi_t \quad \text{for} \quad r\in[t/n,(t+1)/n)\\ n^{-1/2}\xi_n \quad \text{for} \quad r=1\end{array}\right. \tag{40}$$

then we have

$$S_n(r)\rightarrow\lambda W(r).$$

By CMT,

$$\int_0^1S_n(r)dr\rightarrow\lambda\int_0^1W(r)dr,$$

and we have

$$\int_0^1S_n(r)dr=n^{-3/2}\sum_{t=1}^{n}\xi_t.$$

3.2.2 Phillips-Perron Tests for Unit Roots

We will discuss case 2 only and other cases can be derived similarly. Let the true DGP be a random walk with serially correlated errors,

$$y_t = \alpha + \rho y_{t-1} + u_t, \quad u_t = C(L)\epsilon_t$$

where $C(L)$ and ϵ_t satisfy the conditions in proposition 2. When $|\rho| < 1$, OLS estimates of ρ is not consistent when the errors are serially correlated. However, when $\rho = 1$, OLS estimates $\hat{\rho}_n \to 1$. Therefore, Phillips and Perron (1988) proposed estimating the regression with OLS and then correct the estimates with serial correlation.

Under the null hypothesis $H_0 : \alpha = 0, \rho = 1$, the deviations of the OLS estimates vector from the hypothesis

$$\begin{bmatrix} n^{1/2}\hat{\alpha}_n \\ n(\hat{\rho}_n - 1) \end{bmatrix} = \begin{bmatrix} 1 & \sum n^{-3/2} y_{t-1} \\ n^{-3/2}\sum y_{t-1} & n^{-2}\sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-1}\sum y_{t-1}u_t \end{bmatrix}. \quad (41)$$

Use result (f) and (h) in proposition 2,

$$\begin{aligned} \begin{bmatrix} 1 & \sum n^{-3/2} y_{t-1} \\ n^{-3/2}\sum y_{t-1} & n^{-2}\sum y_{t-1}^2 \end{bmatrix}^{-1} &\to \begin{bmatrix} 1 & \lambda \int W(r)dr \\ \lambda \int W(r)dr & \lambda^2 \int W(r)^2 dr \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}^{-1}, \end{aligned}$$

and use result (a) and (e) in proposition 2,

$$\begin{aligned} \begin{bmatrix} n^{-1/2}\sum u_t \\ n^{-1}\sum y_{t-1}u_t \end{bmatrix} &\to_d \begin{bmatrix} \lambda W(1) \\ (1/2)[\lambda^2 W(1)^2 - \gamma_0] \end{bmatrix} \\ &= \begin{bmatrix} \lambda W(1) \\ (1/2)[\lambda^2 W(1)^2 - 1] \end{bmatrix} + \begin{bmatrix} 0 \\ (1/2)(\lambda^2 - \gamma_0) \end{bmatrix} \\ &= \lambda \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} \lambda W(1) \\ (1/2)[\lambda^2 W(1)^2 - 1] \end{bmatrix} + \begin{bmatrix} 0 \\ (1/2)(\lambda^2 - \gamma_0) \end{bmatrix}. \end{aligned}$$

Substitute these two results into (41),

$$\begin{aligned} \begin{bmatrix} n^{1/2}\hat{\alpha}_n \\ n(\hat{\rho}_n - 1) \end{bmatrix} \to & \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \lambda W(1) \\ (1/2)[\lambda^2 W(1)^2 - 1] \end{bmatrix} \\ & + \begin{bmatrix} 1 & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ (1/2)(\lambda^2 - \gamma_0)/\lambda \end{bmatrix} \end{aligned}$$

To test $\rho = 1$,

$$\begin{aligned} n(\hat{\rho}_n - 1) \to & \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} \lambda W(1) \\ (1/2)[\lambda^2 W(1)^2 - 1] \end{bmatrix} \\ & + \frac{\lambda^2 - \gamma_0}{2\lambda^2} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2 dr \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ = & \frac{(1/2)[W(1)^2 - 1] - W(1)\int W(r)dr}{\int W(r)^2 dr - \left[\int W(r)dr\right]} + \frac{(1/2)(\lambda^2 - \gamma_0)}{\lambda^2 \left\{\int W(r)^2 dr - \left[\int W(r)dr\right]\right\}}. \end{aligned}$$

The first term describes the asymptotic distribution of $n(\hat{\rho}-1)$ as if u_t is $i.i.d$ as in the previous subsection (28). The second term is a correction for serial correlation. When u_t is serially uncorrelated, $C(1)=1$, then $\lambda^2=\gamma_0=\sigma^2$. Then this term disappears. The asymptotics for the t-statistics can be derived in a similar way.

3.2.3 Augmented Dickey-Fuller Tests for Unit Roots

An alternative unit root test with serially correlated errors is augmented Dickey-Fuller test. Recall that I used an example of AR(2) process early in this lecture

$$(1-\phi_1 L-\phi_2 L^2)y_t=\epsilon_t,$$

with one unit root and another root $|\lambda_2|<1$. Then we could rewrite it

$$y_t=y_{t-1}+u_t, \quad u_t=(1-\lambda_2)^{-1}\epsilon_t=\theta(L)\epsilon_t.$$

So this is a unit root process with serially correlated errors. To correct for the serial correlation, define

$$\rho=\phi_1+\phi_2, \quad \kappa=-\phi_2.$$

Then we have the following equivalent polynomial,

$$\begin{aligned} & (1-\rho L)-\kappa L(1-L) \\ = \; & 1-\rho L-\kappa L+\kappa L^2 \\ = \; & 1-(\phi_1+\phi_2-\phi_2)L-\phi_2 L^2. \end{aligned}$$

Therefore, the original AR(2) process can be written as

$$[(1-\rho L)-\kappa L(1-L)]y_t=\epsilon_t,$$

or

$$y_t=\rho y_{t-1}+\kappa\Delta y_{t-1}+\epsilon_t \tag{42}$$

This approach can be generalized to AR(p) process, where we define

$$\rho=\phi_1+\phi_2+\ldots+\phi_p,$$

and

$$\kappa_j=-[\phi_{j+1}+\phi_{j+2}+\ldots+\phi_p] \quad \text{for} \quad j=1,2,\ldots,p-1.$$

Note that when the process contain a unit root, which means one root of

$$1-\phi_1 z-\phi_2 z^2-\ldots-\phi_p z^p=0$$

is unity,

$$1-\phi_1-\phi_2-\ldots-\phi_p=0,$$

which implies that $\rho=1$. Therefore to test if a process contain a unit root is equivalent to test if $\rho=1$ in (42). Furthermore, (42) is a regression with serially uncorrelated errors. For simplicity, in our following discussion, we work with an AR(2) process. Again, we only consider case 2. Our regression

$$y_t = \kappa \Delta y_{t-1} + \alpha + \rho y_{t-1} + \epsilon_t \equiv x_t' \beta + \epsilon_t$$

where $x_t = (\Delta y_{t-1}, 1, y_{t-1})$, $\beta = (\kappa, \alpha, \rho)$. The deviation of the OLS estimates from the true β,

$$\hat{\beta}_n - \beta = \left[\sum_{t=1}^{n} x_t x_t' \right]^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t \right].$$

Let $u_t = y_t - y_{t-1}$,

$$\sum_{t=1}^{n} x_t x_t' = \left[\begin{array}{ccc} \sum u_{t-1}^2 & \sum u_{t-1} & \sum u_{t-1} y_{t-1} \\ \sum u_{t-1} & n & \sum y_{t-1} \\ \sum y_{t-1} u_{t-1} & y_{t-1} & \sum y_{t-1}^2 \end{array} \right],$$

$$\sum_{t=1}^{n} x_t \epsilon_t = \left[\begin{array}{c} \sum u_{t-1} \epsilon_t \\ \sum \epsilon_t \\ \sum y_{t-1} \epsilon_t \end{array} \right].$$

u_t is stationary, so its coefficient is $n^{1/2}$ convergent. So the scaling matrix

$$H_n = \left[\begin{array}{ccc} \sqrt{n} & 0 & 0 \\ 0 & \sqrt{n} & 0 \\ 0 & 0 & n \end{array} \right].$$

Premultiply the coefficient vector with H_n,

$$H_n(\hat{\beta}_n - \beta) = \left\{ H_n^{-1} \left[\sum_{t=1}^{n} x_t x_t' \right] H_n^{-1} \right\}^{-1} H_n^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t \right]. \tag{43}$$

Define $\gamma_j = E(u_t u_{t-j})$, $\lambda = \sigma C(1) = \sigma/(1-\kappa)$, where $\sigma = E(\epsilon_t^2)$.

$$H_n^{-1} \left[\sum_{t=1}^{n} x_t x_t' \right] H_n^{-1} \to_d \left[\begin{array}{ccc} \gamma_0 & 0 & 0 \\ 0 & 1 & \lambda \int W(r)dr \\ 0 & \lambda \int W(r)dr & \lambda^2 \int W(r)^2 dr \end{array} \right] \equiv \left[\begin{array}{cc} V & \mathbf{0} \\ \mathbf{0} & Q \end{array} \right].$$

Here $V = \gamma_0$, while it would be a matrix with elements γ_j for a general AR(p) model, and

$$Q = \left[\begin{array}{cc} 1 & \lambda \int W(r)dr \\ \lambda \int W(r)dr & \lambda^2 \int W(r)^2 dr \end{array} \right].$$

Next, consider the second term in (43),

$$H_n^{-1} \left[\sum_{t=1}^{n} x_t \epsilon_t \right] = \left[\begin{array}{c} n^{-1/2} \sum u_{t-1} \epsilon_t \\ n^{-1/2} \sum \epsilon_t \\ n^{-1} \sum y_{t-1} \epsilon_t \end{array} \right]$$

Apply the usual CLT to the first element,

$$n^{-1/2} \sum u_{t-1} \epsilon_t \to_p h_1 \sim N(0, \sigma^2 V).$$

Apply result (a) and (d) of proposition 2 for the other two terms,

$$\begin{bmatrix} n^{-1/2}\sum \epsilon_t \\ n^{-1}\sum y_{t-1}\epsilon_t \end{bmatrix} \to_d h_2 \sim \begin{bmatrix} \sigma W(1) \\ (1/2)\sigma\lambda[W(1)^2 - 1] \end{bmatrix}.$$

Substituting the above results into (43) and we get

$$H_n(\hat{\beta}_n - \beta) \to_d \begin{bmatrix} V & \mathbf{0} \\ \mathbf{0} & Q \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} V^{-1}h_1 \\ Q^{-1}h_2 \end{bmatrix}. \tag{44}$$

Since the limit distribution is block diagonal, we can discuss the coefficients on the stationary components and the nonstationary components seperately. For the stationary components,

$$\sqrt{n}(\hat{\kappa}_n - \kappa) \to_d V^{-1}h_1 \sim N(0, \sigma^2 V^{-1}).$$

In this AR(2) problem, the variance is simply σ^2/γ_0. The limit distribution on the constant and the I(1) components are

$$\begin{bmatrix} n^{1/2}\hat{\alpha}_n \\ n(\hat{\rho}_n - 1) \end{bmatrix} \to_d Q^{-1}h_2 = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma/\lambda \end{bmatrix} \begin{bmatrix} 1 & \int W(r)dr \\ \int W(r)dr & \int W(r)^2 dr \end{bmatrix} \begin{bmatrix} \sigma W(1) \\ (1/2)\sigma\lambda[W(1)^2 - 1] \end{bmatrix}.$$

This implies that $n \cdot (\lambda/\sigma) \cdot (\hat{\rho}_n - 1)$ has the same distribution as in (28). Since $\lambda = \sigma C(1)$, $\lambda/\sigma = C(1) = 1/(1-\kappa)$. Therefore, the ADF ρ-test is

$$\frac{n(\hat{\rho}_n - 1)}{1 - \hat{\kappa}_n} \to_d \frac{(1/2)[W(1)^2 - 1] - W(1)\int W(r)dr}{\int W(r)^2 dr - \left[\int W(r)dr\right]^2}. \tag{45}$$

For the general AR(p) process, simply replace $(1 - \hat{\kappa}_n)$ with $(1 - \hat{\kappa}_{1,n} - \ldots - \hat{\kappa}_{p-1,n})$. The ADF t-test can be found in Hamilton's book.