

LECTURE 12
UNIT ROOT, WEAK CONVERGENCE, FUNCTIONAL CLT

(Davidson (2000), Chapter 14; Phillips' Lectures on Unit Roots, Cointegration and Nonstationarity; White (1999), Chapter 7)

Unit root processes

Definition 1 (*Random walk*) The process $\{X_t\}$ is a random walk if it satisfies

- (a) $X_t = X_{t-1} + \varepsilon_t$ for all $t = 1, 2, \dots$,
- (b) $X_0 = 0$,
- (c) $\{\varepsilon_t\}$ is iid such that $E\varepsilon_t = 0$, $0 < E\varepsilon_t^2 < \infty$.

Random walk is not a stationary process:

$$\begin{aligned} X_t &= X_{t-1} + \varepsilon_t \\ &= (X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= X_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j} \\ &= \sum_{j=0}^{t-1} \varepsilon_{t-j}. \end{aligned}$$

The process has mean zero, however,

$$\text{Var}(X_t) = t\sigma^2.$$

A random walk with drift is defined as

$$\begin{aligned} X_t &= \mu + X_{t-1} + \varepsilon_t \\ &= \mu t + \sum_{j=0}^{t-1} \varepsilon_{t-j}. \end{aligned}$$

Random walk is a member of a more general class of nonstationary processes.

Definition 2 (*Unit root process*) The process $\{X_t\}$ is a unit root process if it satisfies $(1 - L)X_t = U_t$, where $\{U_t\}$ is a mean zero covariance stationary process with short memory.

For example, consider the process $\Phi(L)(1 - L)X_t = \Theta(L)\varepsilon_t$, where $\Phi(L)$ and $\Theta(L)$ are lag polynomials of finite order and have all the roots outside the unit circle. The first difference of X_t ,

$$\begin{aligned} \Delta X_t &= X_t - X_{t-1} \\ &= (1 - L)X_t \end{aligned}$$

is an $ARMA(p, q)$ process. The process $\{X_t\}$ is referred as containing a unit root since its autoregressive polynomial $\Phi(L)(1 - L)$ has a unit root. We say that $\{X_t\}$ is an autoregressive integrated moving average process $ARIMA(p, d, q)$,

$$\Phi(L)(1 - L)^d X_t = \Theta(L)\varepsilon_t,$$

where d indicates the number of time one has to difference $\{X_t\}$ in order to obtain a covariance stationary process with short memory (in the above example, $d = 1$).

Definition 3 (*Integrated or difference stationary processes*) The process $\{X_t\}$ is integrated of order d , denoted as $X_t = I(d)$, if it satisfies $(1 - L)^d X_t = U_t$, where $\{U_t\}$ is a mean zero covariance stationary process with short memory.

An $I(1)$ or unit root process with drift is defined as

$$(1 - L) X_t = \mu + U_t.$$

Integrated processes are referred as difference stationary, since, for example, in the case of $I(1)$, the first difference is a stationary process. This is compared to another class of nonstationary processes:

Definition 4 (*Trend stationary process*) The process $\{X_t\}$ is trend stationary if it satisfies $X_t = \mu + \beta t + U_t$, where $\{U_t\}$ is a mean zero covariance stationary process with short memory.

A trend stationary process evolves along the line of deterministic trend. Deviations from the trend are only of a short-run nature. The shocks have only temporary effect, and its mean, $\mu + \beta t$ describes the behavior of the process in the long-run. On the other hand, an integrated process accumulates its shocks and, therefore, is not ergodic. The shocks have permanent effect, and the process deviates from its expected value for very long periods of time.

Suppose that $\{X_t\}$ is a random walk with the increments $\{U_t\}$ such that $EU_t^2 = 1$. We have

$$\begin{aligned} n^{-1/2} X_n &= n^{-1/2} \sum_{t=1}^n U_t \\ &\rightarrow_d N(0, 1). \end{aligned}$$

Let $[a]$ be the integer part of a . For $0 \leq r \leq 1$ define the following function (partial sum)

$$\begin{aligned} W_n(r) &= n^{-1/2} X_{[nr]} \\ &= n^{-1/2} \sum_{t=1}^{[nr]} U_t. \end{aligned} \tag{1}$$

For fixed r ,

$$\begin{aligned} W_n(r) &= n^{-1/2} \sum_{t=1}^{[nr]} U_t \\ &= n^{-1/2} [nr]^{1/2} \left([nr]^{-1/2} \sum_{t=1}^{[nr]} U_t \right) \\ &\rightarrow_d r^{1/2} N(0, 1) \\ &= N(0, r). \end{aligned}$$

Further, for fixed $0 \leq r_1 < r_2 \leq 1$,

$$W_n(r_2) - W_n(r_1) \rightarrow_d N(0, r_2 - r_1),$$

and, for fixed $0 \leq r_1 < r_2 < \dots < r_k \leq 1$,

$$\begin{pmatrix} W_n(r_1) \\ W_n(r_2) \\ \vdots \\ W_n(r_k) \end{pmatrix} \rightarrow_d N \left(0, \begin{pmatrix} r_1 & r_1 & \dots & r_1 \\ r_1 & r_2 & \dots & r_2 \\ \dots & \dots & \dots & \dots \\ r_1 & r_2 & \dots & r_k \end{pmatrix} \right). \tag{2}$$

Definition 5 (*Brownian motion*) The continuous process $\{W(t) : t \geq 0\}$ is a standard Brownian motion if

- (a) $W(0) = 0$ with probability one,
- (b) For any $0 \leq r_1 < r_2 < \dots < r_k$ and k , $W(r_1), W(r_2) - W(r_1), \dots, W(r_k) - W(r_{k-1})$ are independent,
- (c) For $0 \leq s < t$, $W(t) - W(s) \sim N(0, t - s)$.

A Brownian motion is viewed as a random function, $W(t, \omega)$, where $\omega \in \Omega$ is an outcome of the random experiment (a single outcome determines the whole sample path $W(\cdot, \omega)$). For a fixed ω , $W(\cdot, \omega) : [0, \infty) \rightarrow R$ is a continuous function.

The result in (2) can be alternatively stated as

$$\begin{pmatrix} W_n(r_1) \\ W_n(r_2) \\ \vdots \\ W_n(r_k) \end{pmatrix} \rightarrow_d \begin{pmatrix} W(r_1) \\ W(r_2) \\ \vdots \\ W(r_k) \end{pmatrix} \quad (3)$$

for fixed $0 \leq r_1 < r_2 < \dots < r_k \leq 1$. However, a stronger result is available. We can treat $W_n(r)$ as a random function on the zero-one interval, and discuss convergence in distributions of such a function to $W(r)$ as $n \rightarrow \infty$.

Weak convergence

The concept of weak convergence is an extension of convergence in distribution of random variables or vectors (elements of R^k) to metric spaces.

Definition 6 (*Metric, metric space*) Let \mathcal{S} be a set. A metric is a mapping $d : \mathcal{S} \times \mathcal{S} \rightarrow R$ such that

- (a) $d(x, y) \geq 0$ for all $x, y \in \mathcal{S}$,
- (b) $d(x, y) = 0$ if and only if $x = y$,
- (c) $d(x, y) = d(y, x)$ for all $x, y \in \mathcal{S}$,
- (d) $d(x, y) \leq d(x, z) + d(y, z)$ for all $x, y, z \in \mathcal{S}$.

The pair (\mathcal{S}, d) is called a metric space.

An example of a metric space is R^k with $d(x, y) = \|x - y\|$, where $\|\cdot\|$ denotes the Euclidean norm. Another example is the space of all continuous functions on the zero-one interval $f : [0, 1] \rightarrow R$. This space is denoted as $C[0, 1]$. For any $f, g \in C[0, 1]$ define the uniform metric

$$d_u(f, g) = \sup_{0 \leq r \leq 1} |f(r) - g(r)|.$$

$(C[0, 1], d_u)$ is a metric space. Sample paths of a Brownian motion are in $C[0, 1]$.

A metric gives the distance between the elements of \mathcal{S} , which allows us to introduce the notion of an open set. The set $\{x \in \mathcal{S} : d(x, y) < r\}$ is called the open sphere with the center at y and radius r . For the metric space (\mathcal{S}, d) the Borel σ -field, denoted $\mathcal{B}_{\mathcal{S}}$, is the smallest σ -field containing all open subsets of \mathcal{S} . The following requirements are imposed on (\mathcal{S}, d) : completeness and separability. Completeness is required since we want the limit of a convergent sequence of the elements of \mathcal{S} to be an element of \mathcal{S} as well. The second requirement is separability. The metric space (\mathcal{S}, d) is separable if it contain a countable dense subset; a subset is dense if every point of \mathcal{S} is arbitrary close to one of its elements. Separability ensures that the probabilities can be assigned to the elements of $\mathcal{B}_{\mathcal{S}}$. For example, the real line with the Euclidean metric is a separable space, since the set of rational numbers is countable and dense. The metric space $(C[0, 1], d_u)$ is complete and separable.

For $A \in \mathcal{B}_{\mathcal{S}}$, its boundary, denoted ∂A , is the set of all points not interior to A . Let μ be a probability measure on $(\mathcal{S}, d), \mathcal{B}_{\mathcal{S}}$. The set A is a continuity set of μ if $\mu(\partial A) = 0$.

Definition 7 (*Weak convergence*) Let μ_n, μ be probability measures on the complete and separable measurable metric space $((\mathcal{S}, d), \mathcal{B}_{\mathcal{S}})$. The sequence of probability measures μ_n is said to converge weakly to μ , denoted $\mu_n \Rightarrow \mu$, if for all continuity sets $A \in \mathcal{B}_{\mathcal{S}}$ of μ , $\mu_n(A) \rightarrow \mu(A)$.

Let X_n, X be the random elements on the complete and separable measurable metric space $((\mathcal{S}, d), \mathcal{B}_{\mathcal{S}})$, i.e.

$$\begin{aligned} X &: (\Omega, \mathcal{F}) \rightarrow ((\mathcal{S}, d), \mathcal{B}_{\mathcal{S}}), \\ X_n &: (\Omega, \mathcal{F}) \rightarrow ((\mathcal{S}, d), \mathcal{B}_{\mathcal{S}}), \end{aligned}$$

and measurable. For $A \in \mathcal{B}_{\mathcal{S}}$ define

$$\begin{aligned} \mu_n(A) &= P\{\omega \in \Omega : X_n(\omega) \in A\}, \\ \mu(A) &= P\{\omega \in \Omega : X(\omega) \in A\}, \end{aligned}$$

where P is a probability measure on (Ω, \mathcal{F}) . We say that X_n converges weakly to X , denoted as $X_n \Rightarrow X$, if $\mu_n \Rightarrow \mu$.

The notion of a continuity set is similar to that of a continuity point of a CDF. Let F_n, F be CDFs. We say that $F_n \rightarrow F$ if $F_n(x) \rightarrow F(x)$ for all continuity points x of F . We require convergence of measures only for the continuity points in order to avoid disappearance of the probability mass. For example, suppose $X_n = 1/n$ with probability one. This random variable has the CDF

$$F_n(x) = \begin{cases} 0, & x < 1/n, \\ 1 & x \geq 1/n. \end{cases}$$

The limit of X_n is $X = 0$ with probability one, and its distribution function is

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 & x \geq 0. \end{cases}$$

However, $F_n(x)$ does not converge to $F(x)$ for all $x \in \mathbb{R}$. For $x > 0$, $\lim_{n \rightarrow \infty} F_n(x) = 1$, and for $x < 0$, $\lim_{n \rightarrow \infty} F_n(x) = 0$. However, since for all n , $0 < 1/n$, $\lim_{n \rightarrow \infty} F_n(0) = 0$. Thus, the limit of $F_n(x)$ is

$$\tilde{F}(x) = \begin{cases} 0, & x \leq 0, \\ 1 & x > 0, \end{cases}$$

which is not a CDF. As a result, the probability mass disappears. The reason is that $x = 0$ is not a continuity point of F . Hence, we have to redefine the limit of F_n as F .

The continuous mapping theorem can be extended to the case of weak convergence. In this case, we have to deal with the mappings that are functions of functions. Such a mapping is called a functional.

Theorem 1 (*CMT*) Suppose that $h : ((\mathcal{S}, d), \mathcal{B}_{\mathcal{S}}) \rightarrow (\mathcal{R}, \mathcal{B})$ is a mapping continuous with probability one. Let X_n, X be the random elements on $((\mathcal{S}, d), \mathcal{B}_{\mathcal{S}})$. If $X_n \Rightarrow X$, then $h(X_n) \Rightarrow h(X)$.

For example, suppose that $f \in C[0, 1]$. Let $h(f) = \int_0^1 f(r) dr$. The integral is a continuous functional. Suppose that $f_n \rightarrow f$ in the uniform metric, i.e. $d_u(f_n, f) \rightarrow 0$. Then,

$$\begin{aligned} |h(f_n) - h(f)| &= \left| \int_0^1 f_n(r) dr - \int_0^1 f(r) dr \right| \\ &\leq \int_0^1 |f_n(r) - f(r)| dr \\ &\leq \sup_{0 \leq r \leq 1} |f_n(r) - f(r)| \\ &= d_u(f_n, f) \\ &\rightarrow 0. \end{aligned}$$

Functional CLT (FCLT)

Consider partial sums $W_n(\cdot)$. It is a step function on zero-one interval. The function is $W_n(r)$ constant for $t/n \leq r < (t+1)/n$, since for such r 's, $[rn] = t$. When r hits $(t+1)/n$, the function jumps from $n^{-1/2}X_t$ to $n^{-1/2}(X_t + U_{t+1})$. Thus, this function is continuous on the right:

$$\lim_{r \downarrow r_0} W_n(r) = W_n(r_0),$$

and has left limit, i.e. $\lim_{r \uparrow r_0} W_n(r)$ exists. Such functions are referred as cadlag.

Let $D[0, 1]$ be the space of cadlag functions on the zero-one interval. The space of continuous functions, $C[0, 1]$ is a subspace of $D[0, 1]$, and therefore, $W(\cdot) \in D[0, 1]$. It turns out that $(D[0, 1], d_u)$ is not separable. A space is not separable if it contains noncountable discrete set; a set is discrete if its points are separated. Consider, for example, the following set of functions on $D[0, 1]$. Let

$$f_\theta(x) = \begin{cases} 0, & x < \theta, \\ 1, & x \geq \theta. \end{cases}$$

Define $\{f_\theta : \theta \in [0, 1]\}$. This set is noncountable. However, $d_u(f_{\theta_1}, f_{\theta_2}) = \sup_x |f_{\theta_1}(x) - f_{\theta_2}(x)| = 1$ for $\theta_1 \neq \theta_2$. The distance between any two elements is always 1 regardless of how close θ_1 and θ_2 are (as long as they differ). Therefore, the set is discrete.

The appropriate metric to study weak convergence on $D[0, 1]$ is the Billingsley's metric, d_B , which is a modification of d_u (see, for example Davidson (1994, Chapter 28)). The space $((D[0, 1], d_B), \mathcal{B}_D)$ is complete and separable. The following result establishes convergence of re-scaled partial sums of iid random variables to a Brownian motion.

Theorem 2 (Donsker's FCLT, White (1999), Theorem 7.13) *Let $\{U_t\}$ be a sequence of iid random variables such that $EU_t = 0$ and $EU_t^2 = 1$. Let W_n be defined as in (1), and W be a standard Brownian motion. Then, $W_n \Rightarrow W$.*

The FCLT says that if, for $B \in \mathcal{B}_D$,

$$\begin{aligned} \mu_n &= P\{\omega \in \Omega : W_n(\cdot, \omega) \in B\}, \text{ and} \\ \mu &= P\{\omega \in \Omega : W(\cdot, \omega) \in B\}, \end{aligned} \tag{4}$$

then,

$$\mu_n \Rightarrow \mu.$$

Thus, for n large enough, we can approximate the distribution of the sample paths of a random walk by the distribution of the sample paths of a Brownian motion.

The weak convergence of the FCLT is a stronger result than in (3). The result in (3) gives joint convergence in distribution of a finite number of random variables, or convergence of finite dimensional distributions. On the other hand, the FCLT provides convergence of an infinite dimensional object. In fact, $W_n \Rightarrow W$ implies that $(W_n(r_1), W_n(r_2), \dots, W_n(r_k)) \rightarrow_d (W(r_1), W(r_2), \dots, W(r_k))$.

Convergence of finite dimensional distributions does not imply weak convergence unless the sequence of measures $\{\mu_n\}$ is uniformly tight; a measure P on $((\mathcal{S}, d), \mathcal{B}_S)$ is tight if for all $\varepsilon > 0$ there exists a compact set $K \in \mathcal{B}_S$, such that $P(K) > 1 - \varepsilon$. One can show that the sequence of probability measures in (4) is uniformly tight.

The result can be extended to the time-series case by the means of the Phillips-Solo device (Lecture 11). Suppose that

- $U_t = C(L)\varepsilon_t$,
- $\{\varepsilon_t\}$ is iid such that $E\varepsilon_t = 0$ and $E\varepsilon_t^2 = \sigma^2 < \infty$,
- $\sum_{j=0}^{\infty} j^{1/2} |c_j| < \infty$,

- $C(1) \neq 0$.

Let W_n be defined as in (1). Then,

$$\begin{aligned} W_n &\Rightarrow B, \text{ where} \\ B(r) &= \sigma C(1)W(r), \end{aligned}$$

and W is a standard Brownian motion. The Brownian motion B satisfies (a) and (b) of Definition 5, and (c) is replaced with

$$B(t) - B(s) \sim N(0, \sigma^2 C(1)^2 (t-s)) \text{ for } t > s.$$

The result can be extended to the vector case as well.

Definition 8 (*Vector Brownian motion*) The k -vector process $\{W(t) = (W_1(t), \dots, W_k(t))' : t \geq 0\}$ is a vector standard Brownian motion if W_1, \dots, W_k are independent scalar standard Brownian motion processes.

Suppose that the k -vector process $\{U_t : t = 1, 2, \dots\}$ satisfies

- $U_t = C(L)\varepsilon_t$,
- $\{\varepsilon_t\}$ is iid such that $E\varepsilon_t = 0$ and $E\varepsilon_t\varepsilon_t' = \Sigma$ finite,
- $\sum_{j=0}^{\infty} j^{1/2} \|C_j\| < \infty$,
- $C(1) \neq 0$.

Again, let $W_n(r) = n^{-1/2} \sum_{t=1}^{[nr]} U_t$. Then,

$$\begin{aligned} W_n &\Rightarrow B, \text{ where} \\ B(r) &= C(1)\Sigma^{1/2}W(r), \end{aligned}$$

and $W(r)$ is a k -vector Brownian motion. Notice that in this case,

$$B(t) - B(s) \sim N(0, (t-s)C(1)\Sigma C(1)') \text{ for } t > s.$$