(b) Ordinary Least Squares (OLS) Method

$$S(\phi_1) = \sum_{t=2}^{T} (y_t - \phi_1 y_{t-1})^2$$

is minimized with respect to $\phi_1$.

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{T} y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2} = \phi_1 + \frac{\sum_{t=2}^{T} y_{t-1}\epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} = \phi_1 + \frac{(1/T)\sum_{t=2}^{T} y_{t-1}\epsilon_t}{(1/T)\sum_{t=2}^{T} y_{t-1}^2}$$

$$\longrightarrow \phi_1 + \frac{\mathrm{E}(y_{t-1}\epsilon_t)}{\mathrm{E}(y_{t-1}^2)} = \phi_1$$

OLSE of $\phi_1$ is a consistent estimator.

The following equations are utilized.

$$E(y_{t-1}\epsilon_t) = 0$$

$$E(y_{t-1}^2) = \mathrm{Var}(y_{t-1}) = \gamma(0)$$

8. Asymptotic distribution of OLSE $\hat{\phi}_1$:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) \longrightarrow N(0, 1 - \phi_1^2)$$

**Proof:**

$y_{t-1}\epsilon_t$, $t = 1, 2, \cdots, T$, are distributed with mean zero and variance $\dfrac{\sigma_\epsilon^4}{1 - \phi_1^2}$.

From the central limit theorem,

$$\frac{(1/T)\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sqrt{\sigma_\epsilon^4/(1-\phi_1^2)}/\sqrt{T}} \longrightarrow N(0,1)$$

Rewriting,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0, \frac{\sigma_\epsilon^4}{1-\phi_1^2}).$$

Next,

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \mathrm{E}(y_{t-1}^2) = \gamma(0) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$$

yields:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \frac{(1/\sqrt{T}) \sum_{t=1}^{T} y_{t-1}\epsilon_t}{(1/T) \sum_{t=1}^{T} y_{t-1}^2} \longrightarrow N(0, 1 - \phi_1^2)$$

9. Some formulas:

   (a) Central Limit Theorem

   Random variables $x_1$, $x_2$, $\cdots$, $x_T$ are mutually independently distributed with mean $\mu$ and variance $\sigma^2$.

   Define $\overline{x} = (1/T) \sum_{t=1}^{T} x_t$.

Then,

$$\frac{\overline{x} - \mathrm{E}(\overline{x})}{\sqrt{\mathrm{V}(\overline{x})}} = \frac{\overline{x} - \mu}{\sigma / \sqrt{T}} \longrightarrow N(0, 1)$$

(b) Central Limit Theorem II

Random variables $x_1, x_2, \cdots, x_T$ are distributed with mean $\mu$ and variance $\sigma^2$.

Define $\overline{x} = (1/T) \sum_{t=1}^{T} x_t$.

Then,

$$\frac{\overline{x} - \mathrm{E}(\overline{x})}{\sqrt{\mathrm{V}(\overline{x})}} \longrightarrow N(0, 1)$$

(c) Let $x$ and $y$ be random variables.

$y$ converges in distribution to a distribution, and $x$ converges in probability to a fixed value.

Then, $xy$ converges in distribution.

For example, consider:

$$y \longrightarrow N(\mu, \sigma^2), \qquad x \longrightarrow c.$$

Then, we obtain:

$$xy \longrightarrow N(c\mu, c^2\sigma^2)$$

10. **AR(1) +drift:** $y_t = \mu + \phi_1 y_{t-1} + \epsilon_t$

Mean:

Using the lag operator,

$$\phi(L) y_t = \mu + \epsilon_t$$

where $\phi(L) = 1 - \phi_1 L$.

Multiply $\phi(L)^{-1}$ on both sides. Then, when $|\phi_1| < 1$, we have:

$$y_t = \phi(L)^{-1}\mu + \phi(L)^{-1}\epsilon_t.$$

Taking the expectation on both sides,

$$\begin{aligned} \mathrm{E}(y_t) &= \phi(L)^{-1}\mu + \phi(L)^{-1}\mathrm{E}(\epsilon_t) \\ &= \phi(1)^{-1}\mu = \frac{\mu}{1-\phi_1} \end{aligned}$$

# 15 Unit Root (単位根) and Cointegration (共和分)

## 15.1 Unit Root (単位根) Test (Dickey-Fuller (DF) Test)

1. Why is a unit root problem important?

   (a) Economic variables increase over time in general.

   One of the assumptions of OLS is stationarity on $y_t$ and $x_t$.

   This assumption implies that $\frac{1}{T}X'X$ converges to a fixed matrix as $T$ is large.

That is, asymptotic normality of OLS estimator does not hold.

(b) In nonstationary time series, the unit root is the most important.

In the case of unit root, OLSE of the first-order autoregressive coefficient is consistent.

OLSE is $\sqrt{T}$-consistent in the case of stationary AR(1) process, but OLSE is $T$-consistent in the case of nonstationay AR(1) process.

(c) A lot of economic variables increase over time.

It is important to check an economic variable is trend stationary (i.e., $y_t = a_0 + a_1 t + \epsilon_t$) or difference stationary (i.e., $y_t = b_0 + y_{t-1} + \epsilon_t$).

Consider $k$-step ahead prediction for both cases.

$$\text{(Trend Stationarity)} \qquad y_{t+k|t} = a_0 + a_1(t+k)$$

$$\text{(Difference Stationarity)} \qquad y_{t+k|t} = b_0 k + y_t$$

2. **The Case of $|\phi_1| < 1$:**

$$y_t = \phi_1 y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{i.i.d. } N(0, \sigma_\epsilon^2), \qquad y_0 = 0, \qquad t = 1, \cdots, T$$

Then, OLSE of $\phi_1$ is:

$$\hat{\phi}_1 = \frac{\displaystyle\sum_{t=1}^{T} y_{t-1} y_t}{\displaystyle\sum_{t=1}^{T} y_{t-1}^2}.$$

In the case of $|\phi_1| < 1$,

$$\hat{\phi}_1 = \phi_1 + \frac{\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\displaystyle\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow \phi_1 + \frac{\mathrm{E}(y_{t-1}\epsilon_t)}{\mathrm{E}(y_{t-1}^2)} = \phi_1.$$

Note as follows:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow \mathrm{E}(y_{t-1}\epsilon_t) = 0.$$

By the central limit theorem,

$$\frac{\overline{y\epsilon} - \mathrm{E}(\overline{y\epsilon})}{\sqrt{\mathrm{V}(\overline{y\epsilon})}} \longrightarrow N(0,1)$$

where

$$\overline{y\epsilon} = \frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t.$$

$$
\begin{aligned}
\mathrm{E}(\overline{y\epsilon}) &= 0, \\
\mathrm{V}(\overline{y\epsilon}) &= \mathrm{V}\big(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t\big) = \mathrm{E}\Big(\big(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\epsilon_t\big)^2\Big) \\
&= \frac{1}{T^2}\mathrm{E}\Big(\sum_{t=1}^{T}\sum_{s=1}^{T} y_{t-1}y_{s-1}\epsilon_t\epsilon_s\Big) = \frac{1}{T^2}\mathrm{E}\Big(\sum_{t=1}^{T} y_{t-1}^2\epsilon_t^2\Big) = \frac{1}{T}\sigma_\epsilon^2\gamma(0).
\end{aligned}
$$

Therefore,

$$
\frac{\overline{y\epsilon}}{\sqrt{\sigma_\epsilon^2\gamma(0)/T}} = \frac{1}{\sigma_\epsilon\sqrt{\gamma(0)}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0,1),
$$

which is rewritten as:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{t-1}\epsilon_t \longrightarrow N(0, \sigma_\epsilon^2\gamma(0)).$$

Using $\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \mathrm{E}(y_{t-1}^2) = \gamma(0)$, we have the following asymptotic distribution:

$$\sqrt{T}(\hat{\phi}_1 - \phi_1) = \frac{\dfrac{1}{\sqrt{T}}\displaystyle\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} y_{t-1}^2} \longrightarrow N\left(0, \frac{\sigma_\epsilon^2}{\gamma(0)}\right) = N\left(0, 1-\phi_1^2\right).$$

Note that $\gamma(0) = \dfrac{\sigma_\epsilon^2}{1 - \phi_1^2}$.

3. In the case of $\phi_1 = 1$, as expected, we have:

$$\sqrt{T}(\hat{\phi}_1 - 1) \longrightarrow 0.$$

That is, $\hat{\phi}_1$ has the distribution which converges in probability to $\phi_1 = 1$ (i.e., degenerated distribution).

Is this true?

4. **The Case of $\boldsymbol{\phi_1 = 1}$:** $\Longrightarrow$ Random Walk Process

$y_t = y_{t-1} + \epsilon_t$ with $y_0 = 0$ is written as:

$$y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \cdots + \epsilon_1.$$

Therefore, we can obtain:

$$y_t \sim N(0, \sigma_\epsilon^2 t).$$

The variance of $y_t$ depends on time $t$. $\implies$ $y_t$ is nonstationary.

5. Remember that $\hat{\phi}_1 = \phi_1 + \dfrac{\sum y_{t-1}\epsilon_t}{\sum y_{t-1}^2}$.

   (a) First, consider the numerator $\sum y_{t-1}\epsilon_t$.

We have $y_t^2 = (y_{t-1} + \epsilon_t)^2 = y_{t-1}^2 + 2y_{t-1}\epsilon_t + \epsilon_t^2$.

Therefore, we obtain:

$$y_{t-1}\epsilon_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - \epsilon_t^2).$$

Taking into account $y_0 = 0$, we have:

$$\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}y_T^2 - \frac{1}{2}\sum_{t=1}^{T}\epsilon_t^2.$$

Divided by $\sigma_\epsilon^2 T$ on both sides, we have the following:

$$\frac{1}{\sigma_\epsilon^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma_\epsilon\sqrt{T}}\right)^2 - \frac{1}{2\sigma_\epsilon^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2.$$

From $y_t \sim N(0, \sigma_\epsilon^2 t)$, we obtain the following result:

$$\left(\frac{y_T}{\sigma_\epsilon \sqrt{T}}\right)^2 \sim \chi^2(1).$$

Moreover, the second term is derived from:

$$\frac{1}{T}\sum_{t=1}^{T} \epsilon_t^2 \longrightarrow \mathrm{E}(\epsilon_t^2) = \sigma_\epsilon^2.$$

Therefore,

$$\frac{1}{\sigma_\epsilon^2 T}\sum_{t=1}^{T} y_{t-1}\epsilon_t = \frac{1}{2}\left(\frac{y_T}{\sigma_\epsilon \sqrt{T}}\right)^2 - \frac{1}{2\sigma_\epsilon^2}\frac{1}{T}\sum_{t=1}^{T}\epsilon_t^2 \longrightarrow \frac{1}{2}(\chi^2(1) - 1).$$

(b) Next, consider $\sum y_{t-1}^2$.

$$\mathrm{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) = \sum_{t=1}^{T} \mathrm{E}(y_{t-1}^2) = \sum_{t=1}^{T} \sigma_\epsilon^2 (t-1) = \sigma_\epsilon^2 \frac{T(T-1)}{2}.$$

Thus, we obtain the following result:

$$\frac{1}{T^2}\mathrm{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) \longrightarrow \text{ a fixed value.}$$

Therefore,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 \longrightarrow \text{ a distribution.}$$

6. Summarizing the results up to now, $T(\hat{\phi}_1 - \phi_1)$, not $\sqrt{T}(\hat{\phi}_1 - \phi_1)$, has limiting distribution in the case of $\phi_1 = 1$.

$$T(\hat{\phi}_1 - \phi_1) = \frac{(1/T)\sum y_{t-1}\epsilon_t}{(1/T^2)\sum y_{t-1}^2} \longrightarrow \text{ a distribution.}$$

7. Basic Concepts of Random Walk Process:

   (a) Model: $y_t = y_{t-1} + \epsilon_t, \qquad y_0 = 0, \qquad \epsilon_t \sim N(0,1).$

   Then,

$$y_t = \epsilon_t + \epsilon_{t-1} + \cdots + \epsilon_1.$$

Therefore,

$$y_t \sim N(0, t).$$

$\Longrightarrow$ Nonstationary Process (i.e., variance depends on time $t$.)

Difference between $y_s$ and $y_t$ ($s > t$) is:

$$y_s - y_t = \epsilon_s + \epsilon_{s-1} + \cdots + \epsilon_{t+2} + \epsilon_{t+1}.$$

The distribution of $y_s - y_t$ is:

$$y_s - y_t \sim N(0, s - t).$$

(b) Rewrite as follows:

$$
\begin{aligned}
y_t &= y_{t-1} + \epsilon_t \\
&= y_{t-1} + e_{1,t} + e_{2,t} + \cdots + e_{N,t},
\end{aligned}
$$

where $\epsilon_t = e_{1,t} + e_{2,t} + \cdots + e_{N,t}$.

$e_{1,t}, e_{2,t}, \cdots, e_{N,t}$ are iid with $e_{i,t} \sim N(0, 1/N)$.

That is, suppose that there are $N$ subperiods between time $t$ and time $t+1$.

The limit when $N \to \infty$ is a **continuous time (連続時間)** process known as **standard Brownian motion** or **Wiener process**.

The value of this process at time $r$ is denoted by $W(r)$ for $0 \le r \le 1$.

**Definition:**

Standard Brownian motion $W(r)$ denotes a continuous-time variable at time $r$ and a stochastic function.

$W(r)$ for $r \in [0, 1]$ satisfies the following:

i. $W(0) = 0$

ii. For any time periods $0 \leq r_1 < r_2 < \cdots < r_k \leq 1$, $W(r_2) - W(r_1)$, $W(r_3) - W(r_2)$, $\cdots$, $W(r_k) - W(r_{k-1})$ are independently multivariate normal with $W(s) - W(t) \sim N(0, s - t)$ for $s > t$.

iii. $W(r)$ is continuous in $r$ with probability 1.

An example:

$$\sigma W(r) \sim N(0, \sigma^2 r),$$

which denotes the Brownian motion with variance $\sigma^2$.

Another example;

$$W(r)^2 \sim r \times \chi^2(1).$$