ECON 616: Lecture Two: Deterministic Trends, Nonstationary Processes

Ed Herbst

Last Updated: January 18, 2023

Background

- Overview: Chapters 15-16 from Hamilton (1994).
- Technical Details: Davidson and MacKinnon (2003)

Trends vs Cycles

A commond decomposition of macroeconomic time series is into trend and cycle.

If Y^T corresponds to real per capita GDP gdp_t of the United States. According to this components approach to time series, y_t is expressed as

$$y_t = \ln gdp_t = trend_t + fluctuations_t$$

we will examine regression techniques that decompose y_t in a trend and a cyclical component.

An identification problem

what features of the time series do we regard as trend and what do we regard as fluctuations around the trend?

Let's guess a linear deterministic time trend:

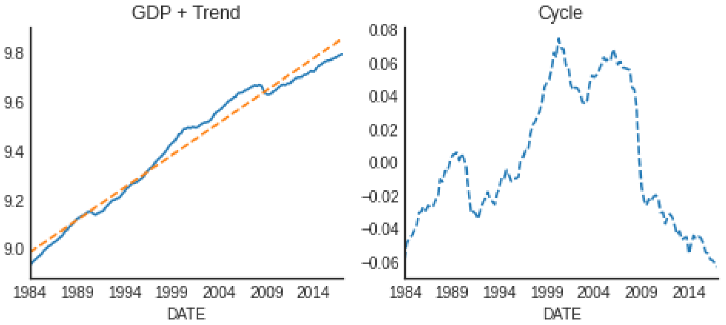
$$\mathrm{y_t} = \beta_1 + \beta_2 \mathrm{t} + \mathrm{u_t}$$

A decomposition of $\mathrm{y_t}$ into trend and fluctuations can be obtained by estimating β_1 and β_2:

$$\begin{aligned} \mathrm{y_t} &= \widehat{\mathrm{trend}}_\mathrm{t} + \widehat{\mathrm{fluctuations}}_\mathrm{t} \\ &= (\hat{\beta}_1 + \hat{\beta}_2 \mathrm{t}) + (\mathrm{y_t} - \hat{\beta}_1 - \hat{\beta}_2 \mathrm{t}). \end{aligned} \tag{1}$$

When $\mathrm{y_t}$ is logged, the coefficient β_2 has the interpretation of an average growth rate.

A Trend Cycle Decomposition of Log US Real GDP



Deterministic Trend Model

Consider the deterministic trend model

$$y_t = \beta_1 + \beta_2 t + u_t$$

with $\mathbb{E}[u_t] = 0$ and $\text{var}[u_t] = \sigma^2$. There are several difficulties associated with the large sample analysis of the OLS estimators $\hat{\beta}_{1,T}$ and $\hat{\beta}_{2,T}$. Taking $x_t = [1, t]'$,

1. The matrix $\frac{1}{T}\sum x_t x_t'$ does not converge to a non-singular matrix Q.
2. In a time series model, the disturbances u_t are in general dependent. This will change the limiting distribution of quantities such as $\sqrt{T}\frac{1}{T}\sum x_t u_t$.
3. If the u_t's are serially correlated, then the OLS estimator will in general be inefficient.

Rates of Convergence for OLS Estimator

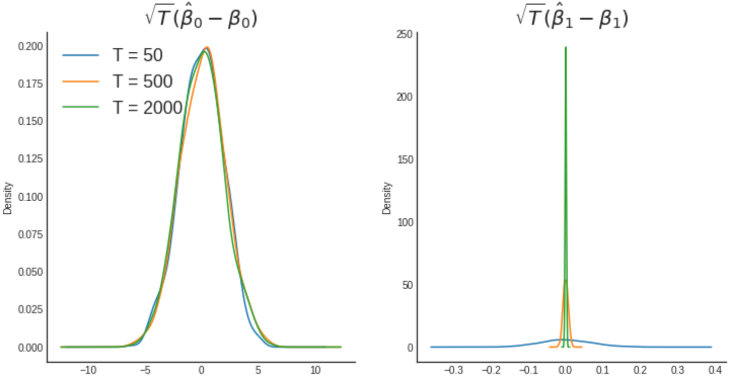
Roughly speaking, convergence rates tell us how fast we can learn the "true" value of a parameter in a sampling experiment.

If "standard" OLS then the variance of the $\hat{\beta}$ converges to zero at rate 1/T.

This isn't true for models with deterministic trends.

Let's look at the distributions of $\sqrt{\mathrm{T}}(\hat{\beta}_0 - \beta_0)$ and $\sqrt{\mathrm{T}}(\hat{\beta}_1 - \beta_1)$

A Monte Carlo



Some math

Facts:

$$\sum_{t=1}^{T} 1 = T, \quad \sum_{t=1}^{T} t = T(T+1)/2, \quad \sum_{t=1}^{T} t^2 = T(T+1)(2T+1)/6.$$

(Assume u_t's are independently distributed.)

$$\frac{1}{T}\sum x_t x_t' = \frac{1}{T}\begin{pmatrix} \sum 1 & \sum t \\ \sum t & \sum t^2 \end{pmatrix}$$

are not convergent!
On the other hand

$$\frac{1}{T^{3}}\sum x_t x_t' \longrightarrow \begin{pmatrix} 0 & 0 \\ 0 & 1/3 \end{pmatrix}$$

which is singular and not invertible!

Message: Trends change the rate of convergence of estimators!

More on Rates of Convergence

It turns out that $\hat{\beta}_{1,\mathrm{T}}$ and $\hat{\beta}_{2,\mathrm{T}}$ have different asymptotic rates of convergence. In particular, we will learn faster about the slope of the trend line than the intercept.

To analyze the asymptotic behavior of the estimators we define the matrix

$$\mathrm{G}_\mathrm{T} = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{T} \end{pmatrix}.$$

Note that the matrix is equivalent to its transpose, that is, $\mathrm{G}_\mathrm{T} = \mathrm{G}'_\mathrm{T}$.

Asymptotic Distributions

We will analyze the following quantity

$$G_T(\hat{\beta}_T - \beta) = \left(\frac{1}{T}\sum G_T^{-1}x_t x_t' G_T^{-1}\right)^{-1}\left(\frac{1}{T}\sum G_T^{-1}x_t u_t\right).$$

It can be easily verified that

$$\frac{1}{T}\sum G_T^{-1}x_t x_t' G_T^{-1} = \frac{1}{T}\begin{pmatrix} \sum 1 & \sum t/T \\ \sum t/T & \sum (t/T)^2 \end{pmatrix} \longrightarrow Q,$$

where

$$Q = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}.$$

Standardization

The term $\frac{1}{T}\sum G_T^{-1} x_t u_t$ has the components $\frac{1}{T}\sum u_t$ and $\frac{1}{T}\sum (t/T)u_t$ which converge in probability to zero based on the weak law of large numbers for non-identically distributed random variables.

Note: Without the proper standardization $\frac{1}{T}\sum tu_t$ will not converge to its expected value of zero. The variance of the random variable Tu_T is getting larger and larger with sample size which prohibits the convergence of the sample mean to its expectation. $\square$

Results

Result: Suppose

$$\mathrm{y_t} = \beta_1 + \beta_2 \mathrm{t} + \mathrm{u_t}, \quad \mathrm{u_t} \sim \mathrm{iid}(0, \sigma^2).$$

Let $\hat{\beta}_{\mathrm{i,T}}$, $\mathrm{i} = 1, 2$ be the OLS estimators of the intercept and slope coefficient, respectively. Then

$$\hat{\beta}_{1,\mathrm{T}} - \beta_1 \xrightarrow{\mathrm{p}} 0 \tag{2}$$

$$\mathrm{T}(\hat{\beta}_{2,\mathrm{T}} - \beta_2) \xrightarrow{\mathrm{p}} 0. \quad \square \tag{3}$$

CLT

I'm not going to show the details of proof for CLT, but

- We use a CLT for independently but not identically distributed random variables (Liapounov)
- Also, Cramer and Wold device that can be used to deduce the convergence of a random vector

based on the convergence of arbitrary linear combinations of its elements.

Result

$$\mathrm{y_t} = \beta_1 + \beta_2 \mathrm{t} + \mathrm{u_t}, \quad \mathrm{u_t} \sim \mathrm{iid}(0, \sigma^2).$$

Let $\hat{\beta}_{\mathrm{i,T}}$, $\mathrm{i} = 1, 2$ be the OLS estimators of the intercept and slope coefficient, respectively. The sampling distribution of the OLS estimators has the following large sample behavior

$$\sqrt{\mathrm{T}}\mathrm{G_T}(\hat{\beta}_\mathrm{T} - \beta) \Longrightarrow \mathcal{N}(0, \sigma^2 \mathrm{Q}^{-1})$$

Note

This is equivalent to

$$\begin{bmatrix} \sqrt{\mathrm{T}}(\hat{\beta}_{1,\mathrm{T}} - \beta) \\ \mathrm{T}^{3/2}(\hat{\beta}_{2,\mathrm{T}} - \beta_2) \end{bmatrix} \Longrightarrow \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \right). \quad \square$$

Having Said All this

When we consider the case where the variance is unknown:

$$\hat{\sigma}^2 = \frac{1}{\mathrm{T}-2}\sum(\mathrm{y_t} - \hat{\beta}_1 - \hat{\beta}_2\mathrm{t})^2$$

Despite the fact that β_1 and β_2 have different asymptic rates of convergence, the t statistics still have $\mathrm{N}(0,1)$ limited distribution because the standard error estimates have offsetting behaviour.

OLS and Serial Dependence

$$y_t = \beta t + u_t$$

u_t are serially correlated, that is, $\mathbb{E}[u_t u_{t-h}] \neq 0$ for some h $\implies$ OLS not efficient.
Let's look at example with MA(1) errors.

$$u_t = \epsilon_t + \theta\epsilon_{t-1}, \quad \epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2).$$

can verify

$$\begin{aligned}
\mathbb{E}[u_t^2] &= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})^2] = (1+\theta^2)\sigma_\epsilon^2 && (4)\\
\mathbb{E}[u_t u_{t-1}] &= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] = \theta\sigma_\epsilon^2 && (5)\\
\mathbb{E}[u_t u_{t-h}] &= 0 \quad h > 1. && (6)
\end{aligned}$$

The OLS estimator

$$\hat{\beta}_T - \beta = \frac{\sum t u_t}{\sum t^2}.$$

To find the limiting distribution, note that

$$\frac{1}{T^3} \sum_{t=1}^{T} t^2 = \frac{T(T+1)(2T+1)}{6T} \longrightarrow \frac{1}{3}.$$

Denominator

The denominator can be manipulated as follows

$$
\begin{aligned}
\sum \mathrm{t}u_\mathrm{t} &= \sum \mathrm{t}(\epsilon_\mathrm{t} + \theta\epsilon_{\mathrm{t}-1}) \\
&= \begin{array}{ccccc} 0 & +\epsilon_1 & +2\epsilon_2 & +3\epsilon_3 & +\ldots \\ +\theta\epsilon_0 & +2\theta\epsilon_1 & +3\theta\epsilon_2 & +4\theta\epsilon_3 & +\ldots \end{array} \\
&= \sum_{\mathrm{t}=1}^{\mathrm{T}-1} (\mathrm{t} + \theta(\mathrm{t}+1))\epsilon_\mathrm{t} + \theta\epsilon_0 + \mathrm{T}\epsilon_\mathrm{T} \\
&= \sum_{\mathrm{t}=1}^{\mathrm{T}-1} (1+\theta)\mathrm{t}\epsilon_\mathrm{t} + \sum_{\mathrm{t}=1}^{\mathrm{T}-1} \theta\epsilon_\mathrm{t} + \theta\epsilon_0 + \mathrm{T}\epsilon_\mathrm{T} \\
&= \sum_{\mathrm{t}=1}^{\mathrm{T}} (1+\theta)\mathrm{t}\epsilon_\mathrm{t} \underbrace{- \theta\mathrm{T}\epsilon_\mathrm{T} + \theta\sum_{\mathrm{t}=1}^{\mathrm{T}} \epsilon_{\mathrm{t}-1}}_{\text{asymp. negligible}} . \qquad (7)
\end{aligned}
$$

OLS, Continued

After standardization by $T^{-3/2}$ we obtain

$$T^{-3/2}\sum tu_t = \frac{1}{\sqrt{T}}(1+\theta)\sum_{t=1}^{T}(t/T)\epsilon_t - \frac{1}{\sqrt{T}}\theta\epsilon_T + \frac{\theta}{T}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_{t-1}.$$

1. First term obeys CLT
2. Second Term goes to zero
3. Third Term goes to zero

Thus ,

$$T^{3/2}(\hat{\beta}_T - \beta) \Longrightarrow \left(0, 3\sigma_\epsilon^2(1+\theta)^2\right).$$

Remark

Consider the following model with iid disturbances

$$\mathrm{y_t} = \beta \mathrm{t} + \mathrm{u_t}, \quad \mathrm{u_t} \sim \mathrm{iid}(0, \sigma_\epsilon^2(1+\theta^2)).$$

The unconditional variance of the disturbances is the same as in the model with moving average disturbances. It can be verified that

$$\mathrm{T}^{3/2}(\hat{\beta}_\mathrm{T} - \beta) \Longrightarrow \big(0, 3\sigma_\epsilon^2(1+\theta^2)\big).$$

If θ is positive then the limit variance of the OLS estimator in the model with iid disturbances is smaller than in the trend model with moving average disturbances.

Positive serial correlated data are less informative than iid data.

Stochastic Trends

We looked at stationary model and deterministic trend models so far.

Now we will examine univariate models with a stochastic trend of the form

$$\mathrm{y_t} = \phi_0 + \mathrm{y_{t-1}} + \epsilon_\mathrm{t} \quad \epsilon_\mathrm{t} \sim \mathrm{iid}(0, \sigma^2)$$

This particular model is called a random walk with drift.

The variable $\mathrm{y_t}$ is said to be integrated of order one.

Cointegration

Moreover, we will consider bivariate models with a common stochastic trend

$$y_{1,t} = \gamma y_{2,t} + u_{1,t} \tag{8}$$
$$y_{2,t} = y_{2,t-1} + u_{2,t} \tag{9}$$

where $[u_{1,t}, u_{2,t}]' \sim \mathrm{iid}(0, \Omega)$. Both $y_{1,t}$ and $y_{2,t}$ have a stochastic trend. However, there exists a linear combination of $y_{1,t}$ and $y_{2,t}$, namely,

$$y_{1,t} - \gamma y_{2,t} = u_t$$

that is stationary. Therefore, $y_{1,t}$ and $y_{2,t}$ are called cointegrated.

Background

In the late 80s and early 90s, this was a super hot research area.

- Dickey and Fuller (1979) examined the sampling distribution of estimators for autoregressive time series with a unit root and provided tables with critical values for unit root tests.
- In Phillips (1986) and (1987) published two papers on spurious regression and time series regressions with a unit root that employ the mathematical theory of convergence of probability measures for metric spaces. This marks a "technological breakthrough" and the field started to grow at an exponential rate thereafter.

Three Choices

Consider the first order autoregressive model with mean zero:

$$y_t = \phi y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$$

Three cases

- $|\phi| < 1$: stationarity! we talked about this last week
- $|\phi| > 1$: explosive! We will not analyze explosive processes in this course.
- $|\phi| = 1$. This is the unit root and will be the focus of this

part of the lecture. If $\phi = 1$ then the AR(1) model simplifies to

$$y_t = y_{t-1} + \epsilon_t$$

With $\Delta = 1 - L$, we have $\Delta y_t = \epsilon_t$ form a stationary process, the random walk is called integrated of order one, denoted by I(1).

Difference b/w Stationary AR and Unit Root

Suppose that the AR process is initialized by $y_0 \sim \mathcal{N}(0,1)$.
Then y_t can be expressed as

$$y_t = \phi^t y_0 + \sum_{\tau=1}^{t} \phi^{\tau-1} \epsilon_{t+1-\tau}$$

- The unconditional mean of y_t is given by

$$\mathbb{E}[y_t] = \phi^{t-1}\mathbb{E}[y_0] + \sum_{\tau=1}^{t} \phi^{\tau-1}\mathbb{E}[\epsilon_\tau] = 0$$

Differences, continued

The unconditional variance is y_t is given by

$$\begin{aligned}
\mathrm{var}[y_t] &= \phi^{2(t-1)}\mathrm{var}[y_0] + \sum_{\tau=1}^{t} \phi^{2(\tau-1)}\mathrm{var}[\epsilon_\tau] \qquad (10) \\
&= \phi^{2(t-1)}\mathrm{var}[y_0] + \sigma^2 \sum_{\tau=1}^{t} \phi^{2(\tau-1)} \\
&= \begin{cases} \phi^{2(t-1)}\mathrm{var}[y_0] + \sigma^2 \frac{1-\phi^{2t}}{1-\phi^2} & \longrightarrow \quad \frac{\sigma^2}{1-\phi^2} \quad \text{if} \quad |\phi| < 1 \\ \mathrm{var}[y_0] + \sigma^2 t & \longrightarrow \quad \infty \quad \text{if} \quad |\phi| = 1 \end{cases}
\end{aligned}$$

as $t \to \infty$.

Differences, continued

The conditional expectation of y_t given y_0 is

$$\mathbb{E}[y_t|y_0] = \phi^{\tau-1} y_0 \longrightarrow \left\{ \begin{array}{lll} 0 & \text{if} & |\phi| < 1 \\ y_0 & \text{if} & \phi = 1 \end{array} \right\}$$

as $t \to \infty$.
In the unit root case, the best prediction of future y_t is the initial y_0 at all horizons, that is, "no change".

In the stationary case, the conditional expectation converges to the unconditional mean. For this reason, stationary processes are also called "mean reverting".

Result

Stationary and unit root processes differ in their behavior over long time horizons.

Suppose that $\sigma^2 = 1$, and $\mathrm{y}_0 = 1$. Then the conditional mean and variance of a process y_t with $\phi = 0.995$ is given by

Horizon t	1	2	5	10	20	50	100
$\mathbb{E}[\mathrm{y}_\mathrm{t} \mid \mathrm{y}_0]$	0.995	0.990	0.975	0.951	0.905	0.778	0.606
$\mathbb{V}[\mathrm{y}_\mathrm{t} \mid \mathrm{y}_0]$	1.000	1.990	4.901	9.563	18.21	39.52	63.46

If interestered in long run predictions, very important to distinguish these two cases.

But note: long run predictions face serious extrapolation problem.

Frequentist Approach

To get a unit root test of the null hypothesis $\mathrm{H_0} : \phi = 1$, we have to find the sampling distribution of a suitable test statistic such as the t ratio

$$\frac{\hat{\phi}_{\mathrm{T}} - 1}{\sqrt{\sigma^2 / \sum \mathrm{y}_{\mathrm{t}-1}^2}}$$

Under the generating mechanism

$$\mathrm{y_t} = \phi_0 + \mathrm{y_{t-1}} + \epsilon_{\mathrm{t}}, \quad \mathrm{iid}(0, \sigma^2)$$

For stationary processes used a variety of WLLN and CLTs, unfortunately, these don't apply.

Heuristic Overview of Asymptotics

Assume that $\phi_0 = 0$, $\sigma = 1$, and $y_0 = 0$. Thus, the process y_t can be represented as

$$y_T = \sum_{t=1}^{T} \epsilon_t$$

Summations will range from $t = 1$ to T unless stated otherwise. The central limit theorem for iid random variables implies

$$\frac{y_T}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum \epsilon_t \Longrightarrow \mathcal{N}(0, 1)$$

This suggests that

$$\frac{1}{T} \sum y_t = \frac{1}{\sqrt{T}} \sum \left[\sqrt{\frac{t}{T}} \frac{1}{\sqrt{t}} \sum_{\tau=1}^{t} \epsilon_\tau \right]$$

will not converge to a constant in probability but instead to a random variable.

Need a more elegant approach!

A Twist on our framework

We used $T = \{0, \pm 1, \pm 2, \ldots\}$.

Consider $S = [0, 1]$. Consider random elements $W(s)$ that correspond to functions this interval.

We will place some probability Q on these functions and show that Q can be helpful in the approximation of the distribution of $\sum y_t$

Defining probability distributions on function spaces is a pain.

Wiener Measure

Let $\mathcal{C}$ be the space of continuous functions on the interval $[0, 1]$.

We will define a probability distribution for the function space $\mathcal{C}$.

This probability distribution is called "Wiener measure".

Whenever we draw an element from the probability space we obtain a function $\mathrm{W(s)}$, $\mathrm{s} \in [0, 1]$. Let $\mathrm{Q}[\cdot]$ denote the expectation operator under the Wiener measure.

Properties of W(s)

- If we repeatedly draw functions under the Wiener measure and evaluate these functions at a particular value $\mathrm{s} = \mathrm{s}'$, then

$$\mathrm{Q}[\{\mathrm{W}(\mathrm{s}') \leq \mathrm{w}\}] = \frac{1}{\sqrt{2\pi \mathrm{s}'}} \int_{-\infty}^{\mathrm{w}} \mathrm{e}^{-\mathrm{u}^2/2\mathrm{s}'} \mathrm{du}$$

  that is,

$$\mathrm{W}(\mathrm{s}') \sim \mathcal{N}(0, \mathrm{s}')$$

  If $\mathrm{s}' = 0$ then the equations is interpreted to mean $\mathrm{Q}[\{\mathrm{W}(0) = 0\}] = 1$. Thus $\mathrm{W}(0) = 0$ with probability one.

Properties of W(s)

- The random function W(s) has independent increments. If

$$0 \leq s_1 \leq s_2 \leq \ldots \leq s_k \leq 1$$

  Then the random variables

$$W(s_2) - W(s_1),\ W(s_3) - W(s_2)\ \ldots,\ W(s_k) - W(s_{k-1})$$

  are independent.
- The random function W(s) is continuous on $s \in [0, 1]$. Otherwise, contradiction.

More of W(s)

It can be shown that there indeed exists a probability distribution on $\mathcal{C}$ with these properties.

Rougly speaking, the Wiener measure is to the theory of stochastic processes, what the normal distribution is to the theory related to real valued random variables.

Note: $\mathrm{W}(1) \sim \mathcal{N}(0, 1)$.

Relating this back to our discrete processes

Define the partial sum process

$$Y_T(s) = \frac{1}{\sqrt{T}} \sum \{t \leq \lfloor Ts \rfloor\} \epsilon_t$$

where $\lfloor x \rfloor$ denotes the integer part of x. Since we assumed that $\epsilon_t \sim \text{iid}(0, 1)$, the partial sum process is a random step function. Interpolation:

$$\bar{Y}_T(s) = \frac{1}{\sqrt{T}} \sum \{t \leq \lfloor Ts \rfloor\} \epsilon_t + (Ts - \lfloor Ts \rfloor) \epsilon_{\lfloor Ts \rfloor + 1} / \sqrt{T}$$

Two ways to randomly generate continuous functions

- Draw a function W(s) from the Wiener distribution. We did not examine how to do the sampling in practice, but since the Wiener distribution is well-defined, it is theoretically possible.
- Generate a sequence $\epsilon_1, \ldots, \epsilon_T$, where $\epsilon_t \sim \text{iid}(0, 1)$ and compute $\bar{Y}_T(s)$.

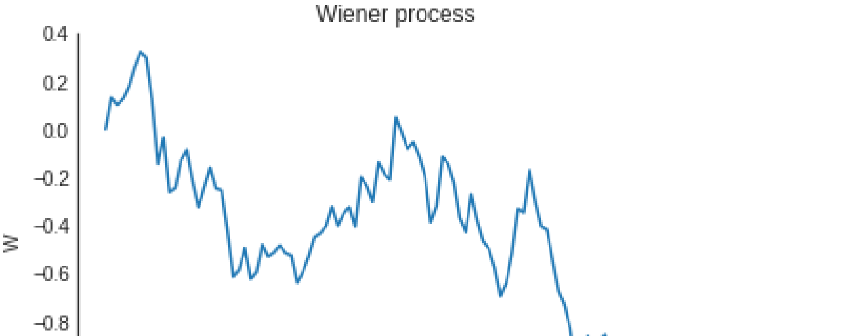
As $T \longrightarrow \infty$, these are basically the same.

Functional CLT: Let $\epsilon_t \sim \text{iid}(0, \sigma^2)$. Then

$$Y_T(s) = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{T} \{t \leq \lfloor Ts \rfloor\} \epsilon_t \Longrightarrow W(s) \quad \square$$

Simulation of Wiener Process

```
<ipython-input-3-6ff2a83aa180>:10: DeprecationWarning: scip
  W = scipy.zeros(N+1)
<ipython-input-3-6ff2a83aa180>:12: DeprecationWarning: scip
  t = scipy.linspace(0, T, N+1);
<ipython-input-3-6ff2a83aa180>:13: DeprecationWarning: scip
  W[1:N+1] = scipy.cumsum(scipy.sqrt(Delta)*scipy.random.st
<ipython-input-3-6ff2a83aa180>:13: DeprecationWarning: scip
  W[1:N+1] = scipy.cumsum(scipy.sqrt(Delta)*scipy.random.st
```



The upshot

The sum

$$\frac{1}{T}\sum y_{t-1}\epsilon_t$$

convergences to a stochastic integral; i.e.,

Suppose that $y_t = y_{t-1} + \epsilon_t$, where $\epsilon_t \sim \mathrm{iid}(0, \sigma^2)$ and $y_0 = 0$. Then

$$\frac{1}{\sigma^2 T}\sum y_{t-1}\epsilon_t \Longrightarrow \int W(s)dW(s)$$

where W(s) denotes a standard Wiener process.

we can use this to develop tests!

Theorem

Suppose that $y_t = \phi y_{t-1} + \epsilon_t$, where $\epsilon_t \sim \mathrm{iid}(0, \sigma^2)$, $\phi = 1$, and $y_0 = 0$. The sampling distribution of the OLS estimator $\hat{\phi}_T$ of the autoregressive parameter $\phi = 1$ and the sampling distribution of the corresponding t-statistic have the following asymptotic approximations

$$z(\hat{\phi}_T) \implies \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(s)^2 ds} \tag{11}$$

$$t(\hat{\phi}_T) \implies \frac{\frac{1}{2}(W(1)^2 - 1)}{\left[\int_0^1 W(s)^2 ds\right]^{1/2}} \tag{12}$$

where W(s) denotes a standard Wiener process. $\square$

Back to Our Model

We will now analyze a simple bivariate system of cointegrated processes. Consider the model

$$y_{1,t} = \gamma y_{2,t} + u_{1,t} \tag{13}$$

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

where $[u_{1,t}, u_{2,t}]' \sim \mathrm{iid}(0, \Omega)$.
Clearly, $y_{2,t}$ is a random walk. Moreover, it can be easily verified that $y_{1,t}$ follows a unit root process.

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

Therefore,

$$y_{1,t} = y_{1,t-1} + \gamma u_{2,t} + u_{1,t} - u_{1,t-1} \tag{16}$$

Thus, both $y_{1,t}$ and $y_{2,t}$ are integrated processes.

Model Continued

However, the linear combination

$$[1,\ -\gamma]\begin{bmatrix} \mathrm{y}_{1,\mathrm{t}} \\ \mathrm{y}_{2,\mathrm{t}} \end{bmatrix} = \mathrm{y}_{1,\mathrm{t}} - \gamma \mathrm{y}_{2,\mathrm{t}} = \mathrm{u}_{1,\mathrm{t}} \tag{17}$$

is stationary. Therefore, $\mathrm{y}_{1,\mathrm{t}}$ and $\mathrm{y}_{2,\mathrm{t}}$ are cointegrated.

The vector $[1, -\gamma]'$ is called the cointegrating vector.

Note that the cointegrating vector is only unique up to normalization.

Rewriting the Model

The model can be rewritten as a VAR(1)

$$y_t = \Phi_1 y_{t-1} + \epsilon_t \tag{18}$$

The elements of the matrix Φ_1 and the definition of ϵ_t is given by

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} + \gamma u_{2,t} \\ u_{2,t} \end{bmatrix} \tag{19}$$

The matrix Φ_1 is of reduced rank in this example of cointegration. More generally cointegrated system can be casted in the form of a vector autoregression in levels of y_t.

Although both $y_{1,t}$ and $y_{2,t}$ follow univariate random walks, the cointegrated system cannot be expressed as a vector autoregression in differences $[\Delta y_{1,t}, \Delta y_{2,t}]'$. Consider

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} 1-L & \gamma L \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} = \Theta(L) u_t \tag{20}$$

Since $|\Theta(1)| = 0$ the moving average polynomial is not

VECM

The cointegrated model can be written in the so-called vector error correction model (VECM) form:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \left(\begin{bmatrix} 1 & -\gamma \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \right) + \begin{bmatrix} u_{1,t} + \gamma u_{2,t} \\ u_{2,t} \end{bmatrix} \tag{21}$$

The term

$$\left(\begin{bmatrix} 1 & -\gamma \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \right) = y_{1,t-1} - \gamma y_{2,t-1} \tag{22}$$

is called error correction term. In economic models it often reflects a long-run equilibrium relationship such as a constant ratio of consumption and output. If the economy is out of equilibrium in period $t-1$, that is, $y_{1,t-1} - \gamma y_{2,t-1} \neq 0$, then the economy adjusts toward its long-run equilibrium and ${}_{t-1}[\Delta y_t] \neq 0$. If the "true" cointegrating vector is known, then both the left-hand-side variables and the error correction term are stationary.

References

Davidson, R. and MacKinnon, J. G. (2003). Econometric Theory and Methods. Oxford University Press.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a):427–431.

Hamilton, J. (1994). Time Series Analysis. Princeton University Press, Princeton, New Jersey.

Phillips, P. (1986). Understanding spurious regressions in econometrics. Journal of Econometrics, 33(3):311–340.