Econ 623 Econometrics II

Topic 5: Non-stationary Time Series

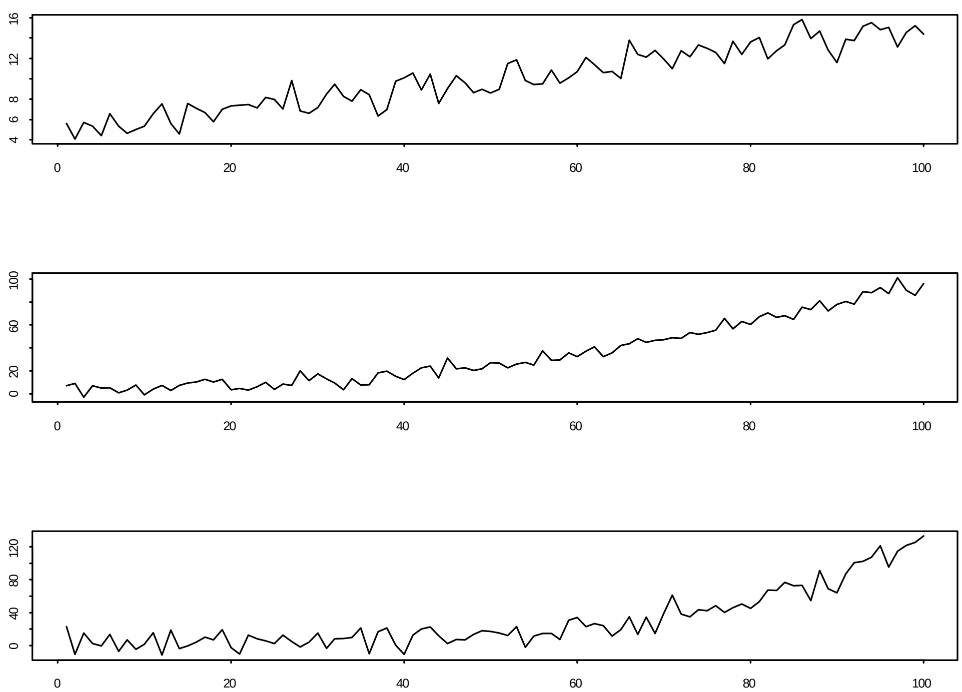
1 Types of non-stationary time series models often found in economics

- Deterministic trend (trend stationary):

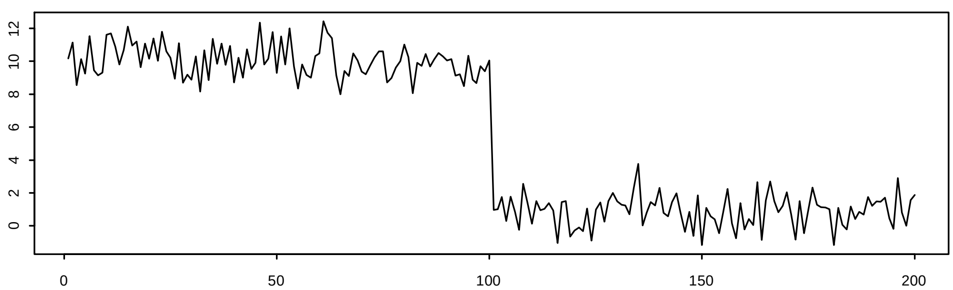
  $X_t = f(t) + \varepsilon_t,$

  where t is the time trend and ε_t represents a stationary error term (hence both the mean and the variance of ε_t are constant. Usually we assume $\varepsilon_t \sim ARMA(p, q)$ with mean 0 and variance σ^2)

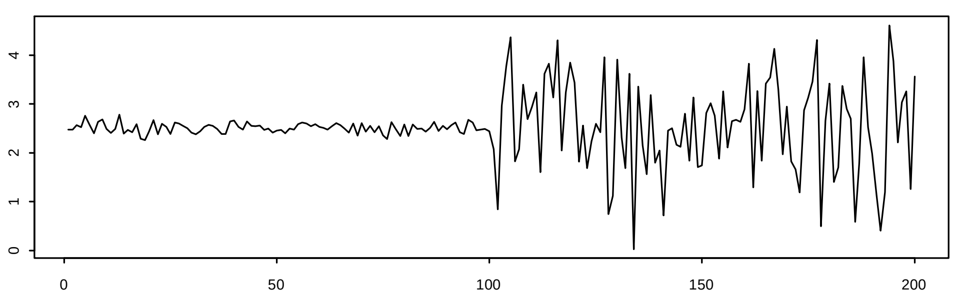
  - $f(t)$ is a deterministic function of time and hence the model is called the deterministic trend model.
  - Why non-stationary? $E(X_t) = f(t)$(depends on t, hence not a constant), $Var(X_t) = \sigma^2$(constant)
  - Since ε_t is stationary, X_t is fluctuating around $f(t)$, the deterministic trend. X_t has no obvious tendency for the amplitude of the fluctuations to increase or decrease since $Var(X_t)$ is a constant. (X_t could increase or decrease, however.)
  - $X_t - f(t)$ is stationary. For this reason a deterministic trend model is sometimes called a trend stationary model.
  - Assume $\varepsilon_t = \phi(L)e_t, e_t \sim iid(0, \sigma_e^2)$. If we label e_t the innovation or the shock to the system, the innovation must have a transient, diminishing effect on X. Why?



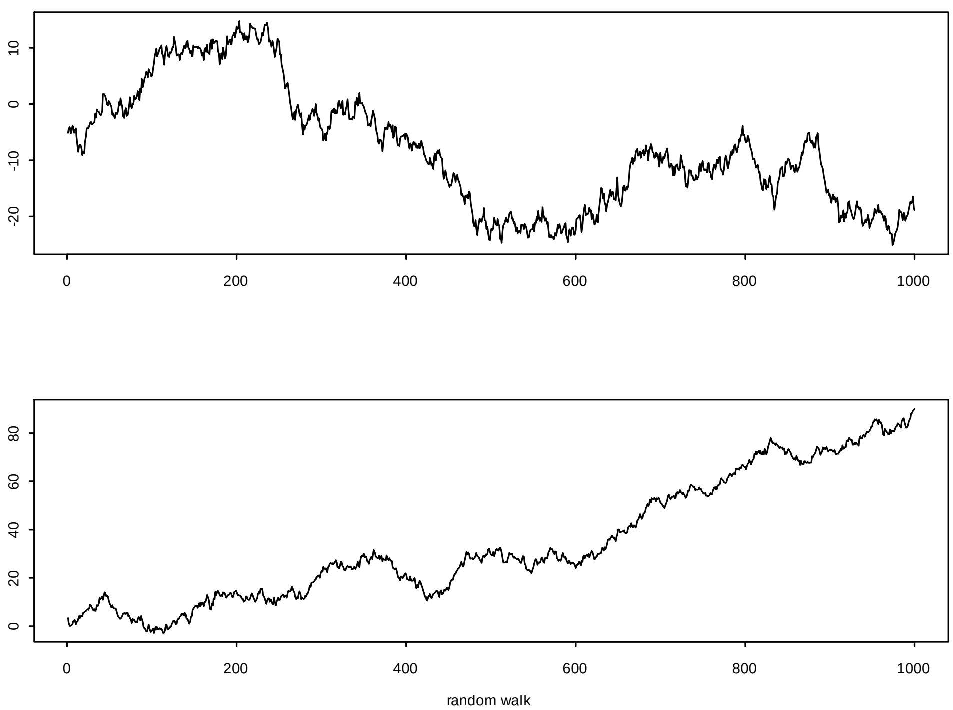
4
6
8
12
16
0
20
40
60
80
100
0
20
60
100
0
20
40
60
80
100
0
40
80
120
0
20
40
60
80
100



0
2
4
6
8
10
12
0
50
100
150
200



0
1
2
3
4
0
50
100
150
200


10
0
-10
-20
0
200
400
600
800
1000
80
60
40
20
0
0
200
400
600
800
1000
random walk

Example: $X_t = \delta_0 + \delta_1 t + \phi(L)e_t$, with $\phi(L) = 1 + \alpha L + \alpha^2 L^2 + \cdots$ and $|\alpha| < 1$

So $X_t = \delta_0 + \delta_1 t + \varepsilon_t, \varepsilon_t = \alpha\varepsilon_{t-1} + e_t$

$\varepsilon_t = \phi(L)e_t$ measures the deviation of the series from trend in period t. We wish to examine the effect of an innovation e_t on $\varepsilon_t, \varepsilon_{t+1}, ...$

$\varepsilon_t = e_t + \alpha e_{t-1} + \alpha^2 e_{t-2} + ...$

which gives $\frac{\partial \varepsilon_{t+s}}{\partial e_t} = \alpha^s$ (**impulse response**)

Therefore, when there is a shock to the economy, a deterministic trend model implies the shock has a transient effect. Sooner or later the system will be back to the deterministic trend.

- If $f(t) = c_1 + c_2 t$, we have a linear trend model. It is widely used.
- If $f(t) = Ae^{rt}$, we have an exponential growth curve
- If $f(t) = c_1 + c_2 t + c_3 t^2$, we have a quadratic trend model
- If $f(t) = \frac{1}{k + ab^t}$, we have a logistic curve

- Stochastic trend (unit root or difference stationarity):

  $X_t = \mu + X_{t-1} + \varepsilon_t$

  where ε_t is a stationary process (hence both the mean and the variance of ε_t are constant. Usually we assume $\varepsilon_t \sim ARMA(p, q)$ with mean 0 and variance σ^2)

  - Another representation: $(1 - L)X_t = \mu + \varepsilon_t$
  - Solving $1 - L = 0$ for L, we have $L = 1$, justifying the terminology "unit root".
  - Pure random walk: $X_t = X_{t-1} + e_t, e_t \overset{iid}{\sim} N(0, \sigma_e^2)$
    1. $X_t = \sum_{j=0}^{t-1} e_{t-j}$ if we assume $X_0 = 0$ with probability 1
    2. $E(X_t) = 0, Var(X_t) = t\sigma_e^2 \to \infty$
    3. Non-stationarity. X_t is wandering around and can be anywhere.

- Random walk with a drift: $X_t = \mu + X_{t-1} + e_t, e_t \overset{iid}{\sim} N(0, \sigma_e^2)$
  1. $X_t = t\mu + \sum_{j=0}^{t-1} e_{t-j}$ if we assume $X_0 = 0$ with probability 1
  2. $E(X_t) = t\mu \rightarrow \infty,\ Var(X_t) = t\sigma_e^2 \rightarrow \infty$
  3. Non-stationarity.
  4. As $E(X_t) = t\mu$, a stochastic trend (or unit root) model could behave similar to a model with a linear deterministic trend.
- In a random walk model, $X_t = t\mu + \sum_{j=0}^{t} e_{t-j}$. If we label e_{t-j} the innovation or the shock to the system, the innovation has a permanent effect on X_t because

$$\frac{\partial X_t}{\partial e_{t-j}} = 1,\ \forall\ j > 0.$$

  This is true for all models with a unit root. Therefore, when there is a shock to the economy, a unit root model implies that the shock has a permanent effect. The system will begin with a new level every time.
- In a trend-stationary model

$$X_t = \delta_0 + \delta_1 t + \phi X_{t-1} + e_t,\ |\phi| < 1,$$

  we have

$$\frac{\partial X_t}{\partial e_{t-j}} = \phi^j \rightarrow 0.$$

  So the innovation has a transient effect on X_t

- A deterministic trend model and a stochastic trend (or unit root) model could behave similar to each other. However, knowing whether non-stationarity in the data is due to a deterministic trend or a stochastic trend (or unit root) would seem to be a very important question in economics. For example, macroeconomists are very interested in knowing whether economic recessions have permanent consequences for the level of future GDP, or instead represent temporary downturns with the lost output eventually made up during the recovery.
- If $(1-L)X_t = \varepsilon_t \sim ARMA(p,q)$, we say $X_t \sim ARIMA(p,1,q)$ or X_t is an I(1) process.
- If $(1-L)^d X_t = \varepsilon_t \sim ARMA(p,q)$, we say $X_t \sim ARIMA(p,d,q)$ or X_t is an I(d) process.

- Why linear time trends and unit roots?

  - Why linear trends? Indeed most GDP series, such as many economic and financial series seem to involve an exponential trend rather than a linear trend (apart from a stochastic trend). Suppose this is true, then we have

    $Y_t = \exp(\delta t)$

    However, if we take the natural log of this exponential trend function, we will have,

    $\ln Y_t = \delta t$

    Thus, it is common to take logs of the data before attempting to describe them. After we take logs, most economic and financial time series exhibit a linear trend. This is why researchers often take the logarithmic transformation before doing analysis.

  - Why unit roots? After taking logs, Suppose a time series follows a unit root rather than a linear trend, we have $(1-L)\ln Y_t = \varepsilon_t$, where ε_t is a stationary process, such as ARMA(p,q). However, $(1-L)\ln Y_t = \ln(Y_t/Y_{t-1}) = \ln(1+\frac{Y_t-Y_{t-1}}{Y_{t-1}}) \approx \frac{Y_t-Y_{t-1}}{Y_{t-1}}$. So the rate of growth of the series Y_t is stationary. For example, if Y_t represents CPI and $\ln Y_t$ is a I(1) process, then $(1-L)\ln Y_t$ represent inflation rate and is a stationary process.

- Other forms of nonstationarity: structural break in mean, structural break in variance

2 Unit Root Tests and A Stationarity Test

- The theory of ARMA method relies on the assumption of stationarity.

- The assumption of stationarity is too strong for many macroeconomic time series. For example, many macroeconomic time series involve trend (both deterministic trends and stochastic trends – unit root).

- Tests for a unit root. Various test statistics are often used in practice: Augmented Dickey-Fuller (ADF) test, Phillips and Perron (PP) test.

- Test for stationarity: Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

- Three possible alternative hypotheses for the unit root tests

  1. no constant and no trend: $X_t = \phi X_{t-1} + e_t, |\phi| < 1$
  2. constant but no trend: $X_t = \alpha + \phi X_{t-1} + e_t, |\phi| < 1$
  3. constant and trend: $X_t = \alpha + \phi X_{t-1} + \delta t + e_t, |\phi| < 1$

- First consider the AR(1) process, $y_t = \phi y_{t-1} + e_t, \;\; u_t \sim iidN(0, \sigma^2)$

- Stationarity requires $|\phi| < 1$

- If $|\phi| = 1$, y_t is nonstationary and has a unit root

- Test $H_0 : \phi = 1$ against $H_1 : |\phi| < 1$

- How to estimate ϕ? OLS estimation, ie, $\widehat{\phi} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \phi + \frac{\sum y_{t-1} e_t}{\sum y_{t-1}^2}$. Stock (1994) shows that the OLS estimator of ϕ is superconsistent.

- What is the distribution or asymptotic distribution of $\widehat{\phi} - \phi$?

- Recall from the standard model, $y = X\beta + e$

  - If X is non-stochastic, $\widehat{\beta} - \beta \sim N(0, \sigma^2 (X'X)^{-1})$

  - If X is stochastic and stationary with some restrictions on autocovariance, $\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$ where $Q = p\lim \frac{X'X}{T}$

2.1 Classical Asymptotic Theory for Covariance Stationary Process

- Consider the AR(1) process, $y_t = \phi y_{t-1} + e_t,\ \ e_t \sim iidN(0, \sigma^2)$ with $|\phi| < 1$
- For this model $E(y_t) = 0$, $Var(y_t) = \frac{\sigma^2}{1-\phi^2} \equiv \sigma_y^2$, $\gamma_j = \phi^j Var(y_t)$.
- What is the asymptotic property of $\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$?
- We have $E(\overline{y}) = 0$, $Var(\overline{y}) = E(\overline{y}^2) = \frac{Var(y_t)}{T}\left(1 + 2\frac{T-1}{T}\phi + 2\frac{T-2}{T}\phi^2 + \cdots + 2\frac{1}{T}\phi^{T-1}\right) \rightarrow 0$. So $\overline{y} \xrightarrow{r^2} 0 \Rightarrow \overline{y} \xrightarrow{p} 0$
- Also, $TVar(\overline{y}) \rightarrow \sum_{j=-\infty}^{\infty} \gamma_j$, which is called the **long-run variance**.
- The consistency is generally true for any covariance stationary process with condition $\sum_{j=0}^{\infty} |\gamma_j| < \infty$.
- **Martingale Difference Sequence (MDS):** $E(y_t | y_{t-1}, ..., y_1) = 0$ for all t
- MDS is serially uncorrelated but not necessarily independent.
- MDS does not need to be stationary.

- **Theorem** (White, 1984): Let $\{y_t\}$ be a MDS with $\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t$ Suppose that (a) $E(y_t^2) = \sigma_t^2 > 0$ with $\frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 \to \sigma^2 > 0$, (b) $E|y_t|^r < \infty$ for some $r > 2$ and all t, and (c) $\frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{p} \sigma^2$. Then $\sqrt{T}\overline{y}_T \xrightarrow{d} N(0, \sigma^2)$.

- **Theorem** (Andersen, 1971): Let

  $y_t = \sum_{t=0}^{\infty} \psi_j \varepsilon_{t-j}$

  where $\{\varepsilon_t\}$ is a sequence of iid variables with $E(\varepsilon_t^2) < \infty$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Let γ_j be the j^{th} order autocovariance. Then

  $\sqrt{T}\overline{y}_T \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty} \gamma_j)$

- To know more about asymptotic theory for linear processes, read Phillips and Solo (1991).

2.2 Dickey-Fuller Test

- Under H_0, however, we cannot claim $\sqrt{T}(\widehat{\phi}-\phi) \xrightarrow{d} N(0,\sigma^2 Q^{-1})$ because y_{t-1} is not stationary.
- If $y_t = \phi y_{t-1} + e_t$, and $H_0 : \phi = 1$, we have $\widehat{\phi} - 1 = \frac{\sum y_{t-1}e_t}{\sum y_{t-1}^2}$
- Consider $y_t^2 = (y_{t-1}+e_t)^2 \Longrightarrow y_{t-1}e_t = \frac{1}{2}(y_t^2 - y_{t-1}^2 - e_t^2) \Longrightarrow \sum y_{t-1}e_t = \frac{1}{2}(y_T^2 - y_0^2 - \sum e_t^2)$
- Let $y_0 = 0, \sum y_{t-1}e_t = \frac{1}{2}(y_T^2 - \sum e_t^2)$
- $e_t \sim iid(0,\sigma^2) \Longrightarrow e_t^2 \sim iid$ with $E(e_t^2) = \sigma^2$. By LLN, $\frac{1}{T}\sum e_t^2 \xrightarrow{p} \sigma^2$
- $y_T = y_0 + \sum_{t=1}^T e_t = \sum e_t \overset{a}{\sim} N(0, T\sigma^2) \Longrightarrow \frac{y_T}{\sqrt{T}\sigma} \xrightarrow{d} N(0,1) \Longrightarrow \frac{y_T^2}{T\sigma^2} \xrightarrow{d} \chi^2_{(1)}$
- $\frac{1}{T\sigma^2}\sum y_{t-1}e_t = \frac{1}{2}(\frac{y_T^2}{T\sigma^2} - \frac{\sum e_t^2}{T\sigma^2}) \xrightarrow{d} \frac{1}{2}(\chi^2_{(1)} - 1)$
- What is the limit behavior of $\sum y_{t-1}^2$? We need to introduce the Brownian Motion.

2.2.1 Brownian Motion

- Consider $y_t = y_{t-1} + e_t, y_0 = 0, e_t \sim iidN(0,1) \Longrightarrow y_t = \sum_{j=1}^{t} e_j \sim N(0,t)$
- $\forall t > s, y_t - y_s = \sum_{j=1}^{t} e_j - \sum_{j=1}^{s} e_j = e_{s+1} + ... + e_t \sim N(0,(t-s))$
- Also note that, $y_t - y_s$ is indept of $y_r - y_q$ if $t > s > r > q$
- What happen if we go from the discrete case to the continuous case? We have the standard Brownian Motion (BM). It is denoted by $B(t)$
- A **stochastic process** B_t over time is a *standard Brownian motion* if for small time interval Δ, the change in $B_t, \Delta B_t (\equiv B_{t+\Delta} - B_t)$, follows the following properties:

  1. $\Delta B_t = e\sqrt{\Delta}$ where $e \sim N(0,1)$
  2. ΔB_t and ΔB_s for changes over any non-overlapping short intervals are independent

- The above two properties imply the following:

  1. $\mathrm{E}(\Delta B_t) = 0$
  2. $\mathrm{Var}(\Delta B_t) = \Delta$
  3. $\sigma(\Delta B_t) = \sqrt{\Delta}$

- In general we have

$$B_T - B_0 = \sum_{i=1}^{N} e_i \sqrt{\Delta}$$

- Due to the above, we have

$$\mathrm{E}(B_T - B_0) = 0$$

$$\mathrm{Var}(B_T - B_0) = T$$

$$\sigma(B_T - B_0) = \sqrt{T}$$

- Therefore, for the Brownian motion, results for the mean and variance in *small* intervals also apply to *large* intervals.

- When $\Delta \to 0$, we write the Brownian motion as:

$$dB_t = e\sqrt{\Delta}$$

  where dB_t should be interpreted as the *change in the Brownian motion over an arbitrarily small time interval.*

- Sample paths of the Brownian motion are continuous everywhere but not smooth anywhere.

- Derivative of the Brownian motion does not exist anywhere and the process is infinitely "kinky".

- A Brownian motion is nonstationary as $\text{Var}(B_T)$ drifts up with t. The transition density is

$$B_{T+\Delta}|B_T \sim N(B_T, \Delta).$$

- A Brownian motion is also called a Wiener process. It has *no drift*. Now we extend it to the generalized Wiener process.

- We can define $e_t = e_{1t} + e_{2t}, e_{1t}, e_{2t} \sim N(0, 1/2), y_{t+\frac{1}{2}} = y_t + e_{1t}$

- Definition of the BM: $W(t) = \sigma B(t)$

- Note:

  1. $B(0) = 0$
  2. $B(t) \sim N(0,t)$ or $B(t)$ is a Gaussian process
  3. $E(B(t+h) - B(t)) = 0, E(B(t+h) - B(t))^2 = |h|$
  4. $E(B(t)B(s)) = \min\{t,s\}$
  5. If $t > s > r > q, B(t) - B(s)$ is indept of $B(r) - B(q)$, ie, $B(t)$ has indept increment
  6. Since $B(t)$ is a Gaussian process, it is completely specified by its covariance matrix
  7. If $B(t)$ is a BM, then so is $B(t+a) - B(a), \lambda^{-1}B(\lambda^2 t), -B(t)$
  8. The sample path of BM is a continuous function of time with prob 1
  9. No point differentiable

- Return to the unit root test. $y_t = y_0 + \sum_{j=1}^{t} e_j$. Define $p_t = \sum_{j=1}^{t} e_j = p_{t-1} + e_t$
- Change the time index from $T = \{1, ..., T\}$ to t with $0 \leq t \leq 1$
- Define $Y_T(t) = \frac{1}{\sigma\sqrt{T}} p_{[Tt]}$, if $\frac{j-1}{T} \leq t \leq \frac{j}{T}, j = 1, ..., T$. $[Tt]$ = integer part of Tt
- For instance, $Y_T(1) = \frac{1}{\sigma\sqrt{T}} p_T$
- $Y_T(t) = \begin{cases} 0 & if \;\; 0 \leq t \leq \frac{1}{T} \\ \frac{p_1}{\sigma\sqrt{T}} & if \;\; \frac{1}{T} \leq t \leq \frac{2}{T} \\ \frac{p_{j-1}}{\sigma\sqrt{T}} & if \;\; \frac{j-1}{T} \leq t \leq \frac{j}{T} \end{cases}$
- Two important theorems

  1. Functional CLT: $Y_T(t) \xrightarrow{d} B(t)$
  2. Continuous mapping theorem: if f is a continuous function, $f(Y_T(t)) \xrightarrow{d} f(B(t))$

- $\sum_{t=1}^{T} y_t = \sum_{j=1}^{T} p_{j-1} + \sum_{j=1}^{T} e_j$ if $y_0 = 0$.
- $\int_{(j-1)/T}^{j/T} Y_T(t)dt = \frac{p_{j-1}}{\sigma T\sqrt{T}} \Longrightarrow \sum p_{j-1} = \sigma T\sqrt{T} \sum_{j=1}^{T} \int_{(j-1)/T}^{j/T} Y_T(t)dt = \sigma T\sqrt{T} \int_0^1 Y_T(t)dt$
- $\sum_{t=1}^{T} y_t = \sigma T\sqrt{T} \int_0^1 Y_T(t)dt + \sqrt{T}\frac{\sum e_j}{\sqrt{T}} \Longrightarrow \frac{\sum_{t=1}^{T} y_t}{T\sqrt{T}} = \sigma \int_0^1 Y_T(t)dt + \frac{1}{T}\frac{\sum e_j}{\sqrt{T}}$
- By FCLT and CMT, $\frac{\sum_{t=1}^{T} y_t}{T\sqrt{T}} \xrightarrow{d} \sigma \int_0^1 B(t)dt$

- $\sum_{t=1}^{T} y_{t-1}^2 = y_0^2 + y_1^2 + ... + y_{T-1}^2 = y_1^2 + ... + y_{T-1}^2 = \sum_{t=1}^{T} y_t^2 - y_T^2 = \sum_{t=1}^{T} p_t^2 - p_T^2$

- Recall $p_{j-1}^2 = \sigma^2 T^2 \int_{j-1/T}^{j/T} Y_T^2(t)dt \implies \sum p_{j-1}^2 = \sigma^2 T^2 \int_0^1 Y_T^2(t)dt$ and $p_T^2 = \sigma^2 T Y_T^2(1)$

- Therefore, $\frac{\sum_{t=1}^{T} y_{t-1}^2}{T^2} \xrightarrow{d} \sigma^2 \int_0^1 B^2(t)dt$

- Under H_0, $T(\widehat{\phi} - \phi) \xrightarrow{d} \frac{\chi_{(1)}^2 - 1}{2\int_0^1 B^2(t)dt}$

- Phillips (1987) proved the above result under very general conditions.

- Note:

  1. This limiting distribution is non-standard
  2. The numerator and denominator are not independent.
  3. It is called the Dickey-Fuller distribution since Dickey and Fuller (1976) uses Monte-Carlo simulation to find the critical values of $\frac{\chi_{(1)}^2 - 1}{2\int_0^1 B^2(t)dt}$ and tabulate them.

2.2.2 Case 1 (no drift or trend in the regression):

- In case 1, we test

$$\begin{cases} H_0 : X_t = X_{t-1} + e_t \\ H_1 : X_t = \rho X_{t-1} + e_t \text{ with } |\rho| < 1 \end{cases}$$

- This is equivalent to

$$\begin{cases} H_0 : \Delta X_t = e_t \\ H_1 : \Delta X_t = \beta X_{t-1} + e_t \text{ with } \beta < 0 \end{cases}$$

- We estimate the following model using the OLS method:

$$\Delta X_t = \beta X_{t-1} + e_t, e_t \sim iid(0, \sigma_e^2)$$

- The OLS estimator of β is defined by $\widehat{\beta}$.The OLS estimator of ϕ is defined by $\widehat{\phi}$. They are biased but consistent.

- The test statistic (known as the **DF** Z test or **coefficient** statistic) is defined as $T\widehat{\beta}$

- Under H_0, the sampling distribution is not a normal distribution, but the DF distribution. See Table B.5 for the critical values of this distribution for different sample sizes.

- Alternatively, we can use t-statistic $\frac{\widehat{\beta}}{\widehat{se}(\widehat{\beta})}$. Although it is called the Dickey-Fuller t-statistic, the sampling distribution is no longer a t distribution, but $\frac{\chi^2_{(1)}-1}{2\left(\int_0^1 B^2(t)dt\right)^{1/2}}$. See Table B.6 for the critical values of this distribution for different sample sizes.

2.2.3 Case 2 (constant but no trend in the regression):

- We test

  $$\begin{cases} H_0 : \Delta X_t = e_t \\ H_1 : \Delta X_t = \alpha + \beta X_{t-1} + e_t \text{ with } \beta < 0 \end{cases}$$

- In case 2, we estimate the following model using the OLS method:

  $\Delta X_t = \alpha + \beta X_{t-1} + e_t, e_t \sim iidN(0, \sigma_e^2)$

- The DF Z test statistic is $T\widehat{\beta}$. Under H_0, the sampling distribution is not a normal distribution, but $\frac{1/2(\chi^2_{(1)}-1)-B(1)\int_0^1 B(t)dt}{\int_0^1 B^2(t)dt-\left(\int_0^1 B(t)dt\right)^2}$. See Table B.5 for the critical values of this distribution for different sample sizes.

- Alternatively, we can use the DF t-statistic $\frac{\widehat{\beta}}{\widehat{se}(\widehat{\beta})}$. The sampling distribution is no longer a t distribution, but $\frac{1/2(\chi^2_{(1)}-1)-B(1)\int_0^1 B(t)dt}{\left(\int_0^1 B^2(t)dt-\left(\int_0^1 B(t)dt\right)^2\right)^{1/2}}$. See Table B.6 for the critical values of this distribution for different sample sizes.

2.2.4 Case 3 (constant and trend in the regression):

- We test

  $$\begin{cases} H_0 : \Delta X_t = e_t \\ H_1 : \Delta X_t = \alpha + \beta X_{t-1} + \delta t + e_t \text{ with } \beta < 0 \end{cases}$$

  or

  $$\begin{cases} H_0 : \Delta X_t = \alpha + e_t \\ H_1 : \Delta X_t = \alpha + \beta X_{t-1} + \delta t + e_t \text{ with } \beta < 0 \end{cases}$$

- In this case, we estimate the following model using the OLS method:

  $$\Delta X_t = \alpha + \beta X_{t-1} + \delta t + e_t, e_t \sim iidN(0, \sigma_e^2)$$

- The Dickey-Fuller Z test statistic is $T\widehat{\beta}$. Under H_0, the sampling distribution is not a normal distribution, but another new distribution (it is also different from the ones in Case 1 and 2). See Table B.5 for the critical values of this distribution for different sample sizes.

- Alternatively, we can use the Dickey-Fuller t-statistic $\frac{\widehat{\beta}}{\widehat{se}(\widehat{\beta})}$. The sampling distribution is no longer a t distribution, but another new distribution (it is also different from ones in Case 1 and 2). See Table B.6 for the critical values of this distribution for different sample sizes.

Test Statistic	1%	2.5%	5%	10%
Case 1	-2.56	-2.34	-1.94	-1.62
Case 2	-3.43	-3.12	-2.86	-2.57
Case 3	-3.96	-3.66	-3.41	-3.13

  When the sample size is 100, the critical values for the Dickey-Fuller t-statistic are given in the table below

Test Statistic	1%	2.5%	5%	10%
Case 1	-2.60	-2.24	-1.95	-1.61
Case 2	-3.51	-3.17	-2.89	-2.58
Case 3	-4.04	-3.73	-3.45	-3.15

2.2.5 Which case to use in practice?

- Use Case 3 test for a series with an obvious trend, such as GDP
- Use Case 2 test for a series without an obvious trend, such as interest rates. For variables such as interest rate we should use Case 2 rather than Case 1 since if the data were to be described by a stationary process, surely the process would have a positive mean. However, if you strongly believe a series has a zero mean when it is stationary, use Case 1.

2.3 Augmented DF (ADF) Test

- The model in the null hypothesis considered in the DF test is a class of highly restrictive unit root models.

- Suppose the model in the null hypothesis is $X_t = X_{t-1} + u_t, u_t = \kappa u_{t-1} + e_t$. Unless $\kappa = 0$, the DF test is not applicable.

- The above model implies the following AR(2) model for X_t,

$$X_t = (1+\kappa)X_{t-1} - \kappa X_{t-2} + e_t = X_{t-1} + \kappa \Delta X_{t-1} + e_t$$

- In general, if

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + ... + \rho_p X_{t-p} + e_t, e_t \sim iid(0, \sigma_e^2), \tag{2.1}$$

  then the DF test is not applicable.

- Model (2.1) can be represented by

$$X_t = \rho X_{t-1} + \phi_1 \Delta X_{t-1} + ... + \phi_{p-1} \Delta X_{t-p+1} + e_t, e_t \sim iid(0, \sigma_e^2),$$

  where $\rho = \sum_{i=1}^{p} \rho_i$

- So if $\rho = 1$, $\{X_t\}$ is a unit root process; if $|\rho| < 1$, $\{X_t\}$ is a stationary process.

- In Case 2, the ADF test is based on the OLS estimator of β from the following regression model,

  $\Delta X_t = \alpha + \beta X_{t-1} + \phi_1 \Delta X_{t-1} + ... + \phi_{p-1} \Delta X_{t-p+1} + e_t, e_t \sim iid(0, \sigma_e^2)$

- In Case 3, the ADF test is based on the OLS estimator of β from the following regression model,

  $\Delta X_t = \alpha + \delta t + \beta X_{t-1} + \phi_1 \Delta X_{t-1} + ... + \phi_{p-1} \Delta X_{t-p+1} + e_t, e_t \sim iid(0, \sigma_e^2)$

- Based on the OLS estimator of β, the ADF test statistic, $\frac{\widehat{\beta}}{\sqrt{\widehat{Var}(\widehat{\beta})}}$, can be used. It has the same sampling distribution as the DF t statistic.

2.4 Phillips-Perron (PP) Test

- $X_t = X_{t-1} + u_t, \Phi(L)u_t = \Theta(L)e_t, e_t \sim iidN(0, \sigma_e^2)$, so u_t is an ARMA process. Phillips and Perron (1988) proposed a nonparametric method of controlling for possible serial correlation in u_t.

- There are two PP test statistics. One of them (so-called PP Z test) is the analogue of $T\widehat{\beta}$. The other (so-called PP t test) is the analogue of $\frac{\widehat{\beta}}{\sqrt{\widehat{Var}(\widehat{\beta})}}$. PP Z test has the same sampling distribution as $T\widehat{\beta}$ under all three cases in the previous section. PP t test has the same sampling distribution as $\frac{\widehat{\beta}}{\sqrt{\widehat{Var}(\widehat{\beta})}}$ under all three cases in the previous section.

- In Case 2, both PP tests are based on the OLS estimator of β from the following regression model,

  $\Delta X_t = \alpha + \beta X_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim$ ARMA

- In this case, the PP Z test statistic is defined by $T\widehat{\beta} - \frac{1}{2}(T^2\widehat{\sigma}^2_{\widehat{\beta}} \div s_T^2)(\lambda^2 - \gamma_0)$, where $s_T^2 = (T-2)^{-1}\sum(X_t - \widehat{\alpha} - \widehat{\phi}X_{t-1})^2, \lambda = \sigma\Theta(1), \gamma_0 = E(u_t^2)$

- In Case 3, both PP tests are based on the OLS estimator of β from the following regression model,

  $\Delta X_t = \alpha + \delta t + \beta X_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim$ ARMA

2.5 Kwiatkowski, Phillips, Schmidt, and Shin (KPSS, 1992) Test

- Unlike the tests we have thus far introduced, KPSS test is a test for stationarity or trend-stationarity, that is, it tests for the null of stationarity or trend stationarity against the alternative of a unit root.

- It is assumed that one can decompose X_t into the sum of a deterministic trend, a random walk and a stationary error

$$X_t = \xi t + r_t + u_t \tag{2.2}$$

$$r_t = r_{t-1} + e_t \tag{2.3}$$

  where $e_t \sim iidN(0, \sigma_e^2)$. The initial value r_0 is assumed to be fixed.

- If $\sigma_e^2 = 0$, X_t is trend-stationary. If $\sigma_e^2 > 0$, X_t is a non-stationary.

- KPSS test is a one-sided Lagrange Multiplier (LM) test
- The KPSS statistic is based on the residuals from the OLS regression of X_t on the exogenous variables $Z_t = 1$ or $(1, t)$:

$$X_t = \delta' Z_t + \epsilon_t$$

- The LM statistic is be defined as:

$$\frac{\sum_{t=1}^{T} S(t)^2}{T^2 f_0},$$

where f_0 is an estimator of the long-run variance and $S(t)$ is a cumulative residual function:

$$S(t) = \sum_{s=1}^{t} \hat{\epsilon}_s$$

- Critical values for the KPSS test statistic are:

Level of Significance	10%	5%	2.5%	1%
Crit value (case 2)	0.347	0.463	0.574	0.739
Crit value (case 3)	0.119	0.146	0.176	0.216

- If the LM test statistic is larger than the critical value, we have to reject the null of stationarity (or trend-stationarity).

- In general ϵ_t is not a white noise. As a result, the long run variance is different from the short run variance. One way to estimate the long run variance is to nonparametrically estimate the spectral density at frequency zero, that is, compute a weighted sum of the autocovariances, with the weights being defined by a kernel function. For example, a nonparametric estimate based on Bartlett kernel is

$$\hat{f}_0 = \sum_{h=-(T-1)}^{T-1} \hat{\gamma}(h) K(h/l),$$

where l is a bandwidth parameter (which acts as a truncation lag in the covariance weighting), and $\hat{\gamma}(h)$ is the h-th sample autocovariance of the residuals $\hat{\epsilon}$, defined as,

$$\hat{\gamma}(h) = \sum_{t=h+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-h}/T,$$

and K is the Barlett kernel function defined by

$$K(x) = \begin{cases} 1 - |x| & \text{if } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

3 Revisit to Box-Jenkins

- If the null hypothesis cannot be rejected, difference the data (ie, $X_t - X_{t-1} = u_t$) to achieve stationarity.
- In the case where both a stochastic trend and a linear deterministic trend are involved, the first difference transformation leads to a stationary process.
- Test for a unit root in the differenced data. If the null hypothesis cannot be rejected, difference the differenced data until stationarity is achieved.
- ARMA(p, q)+one unit root =ARIMA$(p, 1, q)$
- ARMA(p, q)+d unit roots =ARIMA(p, d, q)

4 Limitations of Box-Jenkins

- The Data Generating Process has to be time-invariant. This assumption can be too restrictive for an economy undergoing a period of transition and for data covering a long sample period.
- What if the data is explosive, ie $\rho > 1$.
- What if a nonlinear deterministic trend is involved.
- What if the data follow a trend stationary process.
- Unit root tests and model selection are done in two separate steps.

5 Model Selection

5.1 Why use model selection criteria

- Select p, q in the ARMA model.
- Select lag order, the trend order and other regressors.

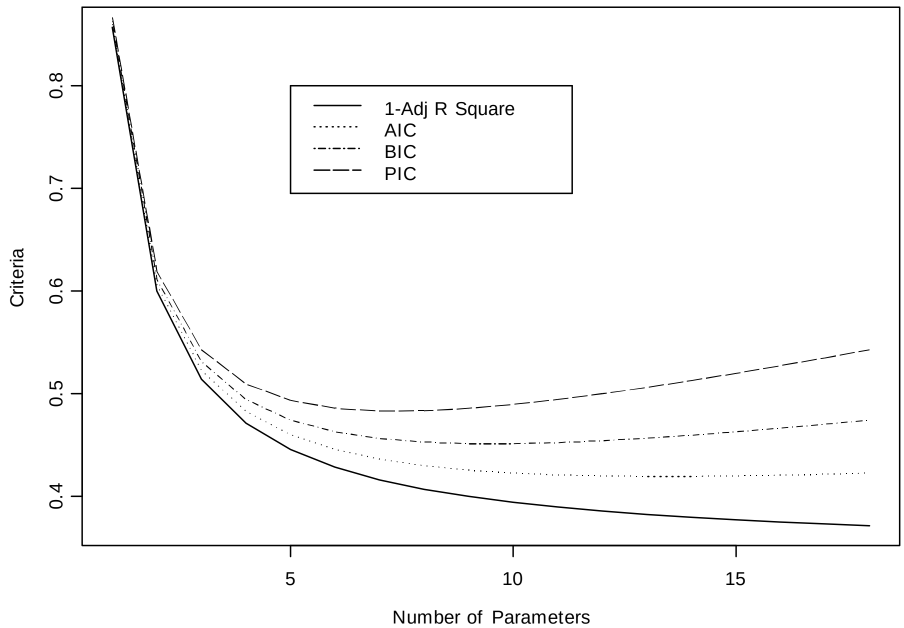
5.2 Conventional model selection criteria

- $1 - R^2$
- Minimize $1 - \bar{R}^2$
- Minimize AIC (Akaike information criterion, Akaike, 1969). AIC$= -\frac{2l}{T} + \frac{2k}{T}$, where k is the number of parameters, l is the log-likelihood value, and T is the sample size
- Minimize BIC (Bayesian information criterion, Schwarz, 1978): BIC$= -\frac{2l}{T} + \frac{k}{T} \ln T$

5.3 Model selection criteria recently developed

- In AIC/BIC, the penalty depends only on the number of regressors while in PIC it depends not only on the number of regressors but also on the type of regressors

- In the regression context with the normality assumption (ie, $Y = X\beta + \varepsilon$), PIC$= -\frac{2l}{T} + \frac{\ln(|X'X/\hat{\sigma}_K^2|)}{T}$, where $\hat{\sigma}_K^2$ is a consistent estimate of the residual variance of a reference model. Usually the reference model is chosen to be the biggest plausible model.



5.4 Which criterion to use in practice

- Comparison of these criteria

- BIC and PIC are consistent. This means in large samples they will select the true model

- AIC in not consistent. In large samples it tends to choose a model with too many lags

- PIC is applicable to ARIMA+Trend processes

- AIC and BIC are provided by most packages when a regression is run.

- A Gauss library COINT developed by Ouliaris and Phillips calculates PIC

- In the context of stationary ARMA models, Phillips and Ploberger (1994) found that PIC is superior to BIC using simulated data.

- PIC is found in Schiff and Phillips (2000) to be superior to BIC in a class of AR+Trend models for forecasting New Zealand GDP

6 Cointegration

- Definition: two variables (Y_t, X_t) is said to be cointegrated if each of them taken individually is I(1) (ie non-stationary with a unit root), while some linear combinations of them (say $Y_t - \beta X_t$) is stationary (ie I(0)).

- If (Y_t, X_t) is cointegrated, the regression model $Y_t = \alpha + \beta X_t + e_t$ is called a cointegration relation, where e_t is I(0).

- $(1, -\beta)$ is called the cointegrating vector. It is easy to see any multiple of it is also a cointegrating vector. So normalisation is important.

- Cointegration is a long run equilibrium relationship, that is the manner in which the two variables drift upward together.

- Since e_t is I(0), any deviation from the equilibrium will come back to the equilibrium in the long run.

- Cointegration and common trend:

  - cointegrated variables share common trends
  - Suppose both Y_t and X_t have a stochastic trend
  - $Y_t = \mu_{Yt} + \varepsilon_{Yt}$, where μ_{Yt} is I(1) ie $\mu_{Yt} = \mu_{Yt-1} + \omega_{Yt}$ and ε_{Yt} is stationary

    $X_t = \mu_{Xt} + \varepsilon_{Xt}$, where μ_{Xt} is I(1) ie $\mu_{Xt} = \mu_{Xt-1} + \omega_{Xt}$ and ε_{Yt} is stationary
  - Now if Y_t and X_t are cointegrated, then there must exist a parameter β such that $Y_t - \beta X_t$ is stationary, ie, $\mu_{Yt} + \varepsilon_{Yt} - \beta(\mu_{Xt} + \varepsilon_{Xt}) = \mu_{Yt} - \beta\mu_{Xt} + \varepsilon_{Yt} - \beta\varepsilon_{Xt}$ is stationary. Since $\varepsilon_{Yt} - \beta\varepsilon_{Xt}$ is stationary, for $Y_t - \beta X_t$ to be stationary, $\mu_{Yt} - \beta\mu_{Xt}$ must be 0. Therefore $\mu_{Yt} = \beta\mu_{Xt}$
  - The equality thus necessarily requires that Y_t and X_t must have a common stochastic trend

- The concept of cointegration is originally due to Granger (1981) but popularised by Engle and Granger (1987, Econometrica)

- Estimation of cointegration systems:

  1. OLS estimator is not only consistent but also "super-consistent". This means the estimator converges to its limiting distribution at a faster rate than the estimator in the stationary case.

  2. Super-consistency is also true in the cointegration model involved the auto-correlation.

- Test for cointegration (between Y_t and X_t)

  Basic Idea:

  1. Test Y_t and X_t for I(1)
  2. Obtain the fitted residual $\widehat{e}_t$ from a regression of Y_t on X_t (different cointegrating regression models lead to different cases)
  3. Use the ADF or PP test to test for a unit root in $\widehat{e}_t$. Since the test is based on $\widehat{e}_t$, not e_t, we cannot use the Dickey-Fuller tables as before. We must use modified critical values.
  4. If H_0 is rejected Y_t and X_t are cointegrated. If H_0 cannot be rejected, the regression of Y_t on X_t is spurious.

- Case 2 (constant and no trend in the regression):

  - We estimate the following model (with the constant term but without trend) using the OLS method

    $Y_t = \alpha + X_t'\beta + e_t$

  - The ADF t-statistic or the PP t-statistic can be used. The critical values for the ADF t-statistic or the PP t-statistic are given in the table below.

- Case 3 (constant and trend in the regression):

  - We estimate the following model (with the constant term and the trend) using the OLS method:

    $Y_t = \alpha + X_t'\beta + \delta t + e_t$

  - The ADF t-statistic or the PP t-statistic can be used. The critical values for the ADF t-statistic or the PP t-statistic are given in the table below.

- Which case to use in practice?

  - Use Case 3 test if either Y_t or X_t or both involve a possible trend, such as GDP

  - Use Case 2 test if neither Y_t nor X_t involve a possible trend, such as interest rates and exchange rates.

k	2		3		4		5		6	
case	2	3	2	3	2	3	2	3	2	3
1%	-3.90	-4.32	-4.29	-4.66	-4.64	-4.97	-4.96	-5.25	-5.25	-5.52
5%	-3.34	-3.78	-3.74	-4.12	-4.10	-4.43	-4.42	-4.72	-4.71	-4.98
10%	-3.04	-3.50	-3.45	-3.84	-3.81	-4.15	-4.13	-4.43	-4.42	-4.70

- Error Correction Model (ECM):

  1. Since the trends of cointegrated variables are linked, the dynamic paths of such variables must bear some relation to the current deviation from the equilibrium relationship. This connection between the change in a variable and the deviation from equilibrium is examined via the error correction representation.
  2. The set of cointegrated variables is said to be in equilibrium if $Y_t - \alpha - \beta X_t = 0$
  3. Thus the deviation from the long run equilibrium is $e_t = Y_t - \alpha - \beta X_t$ and this error must be stationary.
  4. If the system is to return to equilibrium the movements of at least some of the variables must respond to the magnitude of the disequilibrium.
  5. ECM: $\Delta Y_t = \alpha_0 + \delta(Y_{t-1} - \alpha - \beta X_{t-1}) + \gamma \Delta X_t + \varepsilon_t$, where (α, β) captures the long-run equilibrium relationship. δ captures the short-run dynamics and can be interpreted as the speed of adjustment of the system towards the long run equilibrium.
  6. Since all the variables in the ECM are I(0), the classical inferences follow.
  7. Estimation of ECM.
     (a) NLS. The ECM is a non-linear model and can be estimated by the NLS method, ie, $\min_{\{\alpha_0,\delta,\alpha,\beta,\gamma\}} \varepsilon_t^2$
     (b) OLS. Rewrite the ECM as $\Delta Y_t = \alpha_0 + \delta Y_{t-1} - \alpha^* - \beta^* X_{t-1} + \gamma \Delta X_t + \varepsilon_t$, where $\alpha^* = \delta\alpha, \beta^* = \delta\beta$. The new model is a linear model. However, α and β cannot be directly estimated.

7 Spurious Regression

- $Y_t = a + bZ_t + u_t$
- Classical approach requires that both Y_t and Z_t and hence u_t are stationary
- Now consider $Y_t = Y_{t-1} + \varepsilon_{1t}$, $Z_t = Z_{t-1} + \varepsilon_{2t}$, where $\varepsilon_{jt} \sim iid(0, \sigma_j^2)$
- Let $Y_0 = Z_0 = 0$, we have $Y_t = \sum_{j=1}^{t} \varepsilon_{1j}$, $Z_t = \sum_{j=1}^{t} \varepsilon_{2t}$
- If ε_{1t} and ε_{2t} are indept, the regression is a spurious regression.
- Why spurious?
- Using Monte Carlo simulations Granger and Newbold (1974) find that t statistic rejects the null hypothesis $b = 0$ far more often than it should and tends to reject it more and more frequently the larger is the sample size.
- Using asymptotic theory Phillips (1986) finds that $t \xrightarrow{p} +\infty$, hence it will reject the null hypothesis $b = 0$ all the time asymptotically. Also he shows $R^2 \xrightarrow{p} 1$
- Another type of spurious regression: $Y_t = a + bt + u_t$ and $Y_t = Y_{t-1} + \varepsilon_t$
- Note:

  1. If Y_t and Z_t are stationary, the classical theory applies
  2. If Y_t and Z_t are integrated with different orders, the regression is meaningless
  3. If Y_t and Z_t are integrated with the same order, but u_t is integrated with the same order, we have a spurious regression
  4. If Y_t and Z_t are integrated with the same order, but u_t is integrated with a lower order, we have a cointegration