Econ 623 Econometrics II

Topic 3: 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)

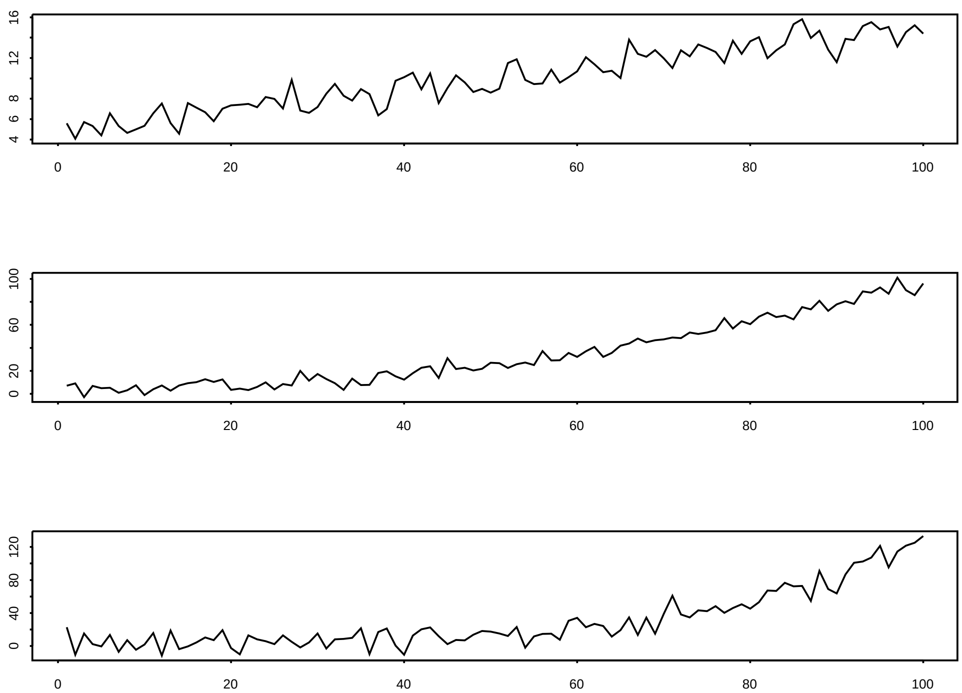
  – $f(t)$ is a deterministic function of time and hence the model is called the deterministic trend model.

  – 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?



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

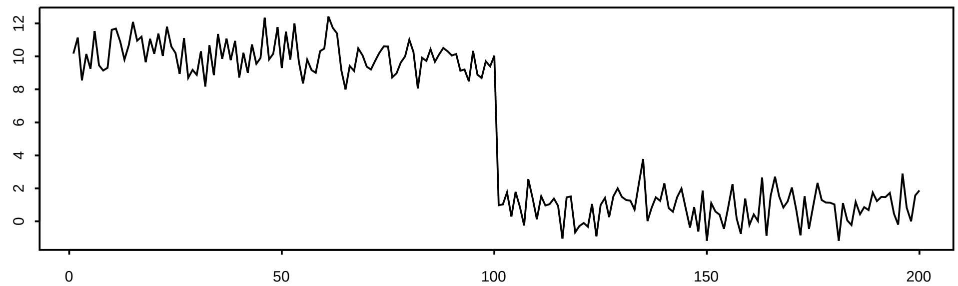
$\varepsilon_t = \phi(L)e_t$ measures the deviation of the series from trend in period t. We wish to examine the effect of an innovation e_t on $\varepsilon_t, \varepsilon_{t+1}, ...$

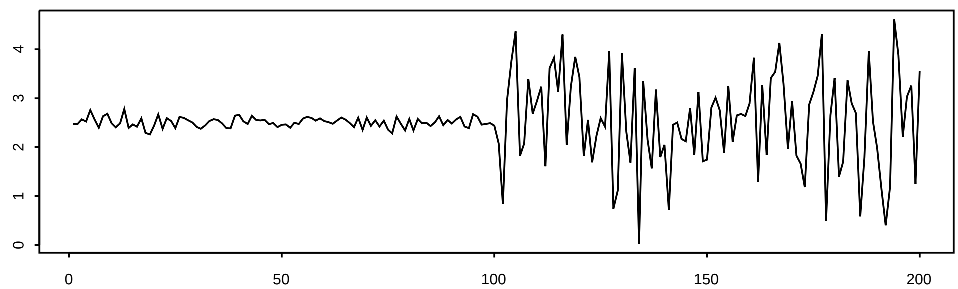
$\varepsilon_t = e_t + \alpha e_{t-1} + \alpha^2 e_{t-2} + ...$

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.

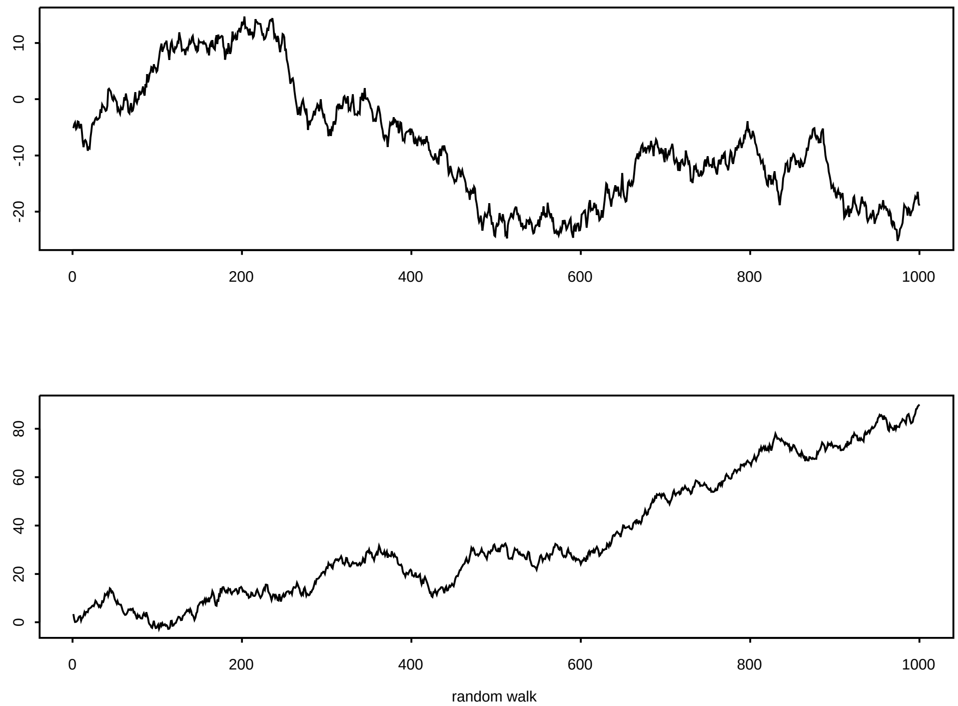
- If $f(t) = c_1 + c_2 t$, we have a linear trend model. It is widely used.
- If $f(t) = Ae^{rt}$, we have an exponential growth curve





– 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. $E(X_t) = t\mu \to \infty,\ Var(X_t) = t\sigma_e^2 \to \infty$
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

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

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.

- Explosive trend: $X_{t+1} = \phi X_t + \varepsilon_t, \phi > 1, X_0 = 0$.

- Why explosive trends?

  – Let P_t be the stock price (or house price) at time t before the dividend payout or rental, D_t be the dividend payoff from the asset at time t, and r be the discount rate ($r > 0$). The standard no arbitrage condition implies that

  $$P_t = \frac{1}{1+r} E_t(P_{t+1} + D_{t+1}), \tag{1.1}$$

  and recursive substitution yields

  $$P_t = F_t + B_t, \tag{1.2}$$

  where $F_t = \sum_{i=1}^{\infty} (1+r)^{-i} E_t(D_{t+i})$ and

  $$E_t(B_{t+1}) = (1+r)B_t. \tag{1.3}$$

  the asset price is decomposed into two components, a "fundamental" component, F_t, that is determined by expected future dividends, and a supplementary solution corresponding to the "bubble" component, B_t. When $B_t = 0$, $P_t = F_t$. Otherwise, $P_t = F_t + B_t$ and price embodies the explosive component B_t. Consequently, under bubble conditions, P_t will manifest the explosive behavior inherent in B_t.

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

2 Processes with Deterministic Trends

- Model with a simple linear trend

$$X_t = \delta t + e_t, t = 1, ..., T.$$

- Let the OLS estimator of δ be $\widehat{\delta}_T$, ie,

$$\widehat{\delta}_T = \frac{\sum X_t t}{\sum t^2} = \delta + \frac{\sum e_t t}{\sum t^2}.$$

- If $e_t \sim N(0, \sigma^2)$, then

$$\widehat{\delta}_T \sim N\left(\delta, \frac{\sigma^2}{\sum t^2}\right) = N\left(\delta, \frac{6\sigma^2}{T(T+1)(2T+1)}\right).$$

This is the exact small-sample normal distribution. The corresponding t statistic

$$\frac{\sqrt{T(T+1)(2T+1)}\widehat{\delta}_T}{\sqrt{6\widehat{\sigma}_T^2}}$$

has exact small-sample t distribution, where $\widehat{\sigma}_T^2 = \frac{1}{T-1}\sum\left(X_t - \widehat{\delta}_T t\right)$.

- If $e_t \overset{iid}{\sim} (0, \sigma^2)$ and has finite fourth moment, we have to find the asymptotic distributions for $\widehat{\delta}_T$.

- Note that

$$T^{3/2}\left(\widehat{\delta}_T - \delta\right) = \frac{T^{-1/2}\sum e_t t/T}{T^{-3}\sum t^2}.$$

- Also note that $\{e_t t/T\}$ is a MDS with (1) $\sigma_t^2 = E(e_t^2 t^2/T^2) = \sigma^2 t^2/T^2$ and $\frac{1}{T}\sum \sigma^2 t^2/T^2 \to \sigma^2/3$; (2) $E|e_t t/T|^4 < \infty$; (3) $\frac{1}{T}\sum e_t^2 t^2/T^2 \overset{p}{\to} \sigma^2/3$. The reason for (3) to hold is because

$$\begin{aligned} E\left(\frac{1}{T}\sum e_t^2 t^2/T^2 - \frac{1}{T}\sum \sigma_t^2\right)^2 &= E\left(\frac{1}{T}\sum\left(e_t^2 - \sigma^2\right)t^2/T^2\right)^2 \\ &= \frac{1}{T^6}E\left(e_t^2 - \sigma^2\right)^2 \sum t^4 \to 0. \end{aligned}$$

- Applying CLT for MDS to $\{e_t t/T\}$, we get $T^{-1/2}\sum e_t t/T \overset{d}{\to} N(0, \sigma^2/3)$. Since $T^{-3}\sum t^2 \to 1/3$, $T^{3/2}\left(\widehat{\delta}_T - \delta\right) \overset{d}{\to} N(0, 3\sigma^2)$.

- Consider the t statistic:

$$\frac{\sqrt{T(T+1)(2T+1)}\left(\widehat{\delta}_T - \delta\right)}{\sqrt{6\widehat{\sigma}_T^2}} = \frac{\sqrt{T(T+1)(2T+1)}T^{3/2}\left(\widehat{\delta}_T - \delta\right)}{\sqrt{6\widehat{\sigma}_T^2}} \overset{d}{\to} N\left(0, 1\right).$$

.

- Model with a simple linear trend

$$X_t = \alpha + \delta t + e_t.$$

- If $e_t \overset{iid}{\sim} (0, \sigma^2)$ and has finite fourth moment, then

$$\begin{bmatrix} T^{-1/2} \sum e_t \\ T^{-1/2} \sum (t/T)\, e_t \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \right).$$

- Need to find the asymptotic distributions for $\widehat{\alpha}_T$ and $\widehat{\delta}_T$ and the rate of convergence is different.

$$\begin{bmatrix} T^{1/2} (\widehat{\alpha}_T - \alpha) \\ T^{3/2} \left(\widehat{\delta}_T - \delta \right) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \right).$$

- The corresponding t statistics converge to $N(0, 1)$.

3 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, $X_t = \phi X_{t-1} + e_t,\ \ u_t \overset{iid}{\sim} N(0, \sigma^2)$

- Stationarity requires $|\phi| < 1$

- If $|\phi| = 1$, X_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, $\widehat{\phi} = \frac{\sum X_t X_{t-1}}{\sum X_{t-1}^2} = \phi + \frac{\sum X_{t-1} e_t}{\sum X_{t-1}^2}$. 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, $\sqrt{T}(\widehat{\beta} - \beta) \overset{d}{\to} N(0, \sigma^2 Q^{-1})$ where $Q = p\lim \frac{X'X}{T}$

3.1 Classical Asymptotic Theory for Covariance Stationary Process

- Consider the AR(1) process, $X_t = \phi X_{t-1} + e_t,\;\; e_t \sim iidN(0, \sigma^2)$ with $|\phi| < 1$
- For this model $E(X_t) = 0$, $Var(X_t) = \frac{\sigma^2}{1-\phi^2} \equiv \sigma_X^2$, $\gamma_j = \phi^j Var(X_t)$.
- What is the asymptotic property of $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$?
- We have $E(\overline{X}) = 0$, $Var(\overline{X}) = E(\overline{X}^2) = \frac{Var(X_t)}{T}\left(1 + 2\frac{T-1}{T}\phi + 2\frac{T-2}{T}\phi^2 + \cdots + 2\frac{1}{T}\phi^{T-1}\right) \to 0$. So $\overline{X} \overset{r^2}{\to} 0 \Rightarrow \overline{X} \overset{p}{\to} 0$
- Also, $TVar(\overline{X}) \to \sum_{j=-\infty}^{\infty} \gamma_j$, 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):** $E(X_t | X_{t-1}, ..., X_1) = 0$ for all t
- MDS is serially uncorrelated but not necessarily independent.
- MDS does not need to be stationary.

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

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

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

  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

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

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

3.2 Dickey-Fuller Test

- Under H_0, however, we cannot claim $\sqrt{T}(\widehat{\phi}-\phi) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$ because X_{t-1} is not stationary.

- If $X_t = \phi X_{t-1} + e_t$, and $H_0 : \phi = 1$, we have $\widehat{\phi} - 1 = \frac{\sum X_{t-1}e_t}{\sum X_{t-1}^2}$

- Consider $X_t^2 = (X_{t-1} + e_t)^2 \Longrightarrow X_{t-1}e_t = \frac{1}{2}(X_t^2 - X_{t-1}^2 - e_t^2) \Longrightarrow \sum X_{t-1}e_t = \frac{1}{2}(X_T^2 - X_0^2 - \sum e_t^2)$

- Let $X_0 = 0, \sum X_{t-1}e_t = \frac{1}{2}(X_T^2 - \sum e_t^2)$

- $e_t \overset{iid}{\sim} (0, \sigma^2) \Longrightarrow e_t^2 \sim iid$ with $E(e_t^2) = \sigma^2$. By LLN, $\frac{1}{T}\sum e_t^2 \xrightarrow{p} \sigma^2$

- $X_T = X_0 + \sum_{t=1}^T e_t = \sum e_t \overset{a}{\sim} N(0, T\sigma^2) \Longrightarrow \frac{X_T}{\sqrt{T}\sigma} \xrightarrow{d} N(0,1) \Longrightarrow \frac{X_T^2}{T\sigma^2} \xrightarrow{d} \chi^2_{(1)}$

- $\frac{1}{T\sigma^2}\sum X_{t-1}e_t = \frac{1}{2}(\frac{X_T^2}{T\sigma^2} - \frac{\sum e_t^2}{T\sigma^2}) \xrightarrow{d} \frac{1}{2}(\chi^2_{(1)} - 1)$

- What is the limit behavior of $\sum X_{t-1}^2$? We need to introduce the Brownian Motion.

3.2.1 Brownian Motion

- Consider $X_t = X_{t-1} + e_t, X_0 = 0, e_t \sim iidN(0,1) \Longrightarrow X_t = \sum_{j=1}^{t} e_j \sim N(0,t)$

- $\forall t > s, X_t - X_s = \sum_{j=1}^{t} e_j - \sum_{j=1}^{s} e_j = e_{s+1} + ... + e_t \sim N(0,(t-s))$

- Also note that, $X_t - X_s$ is indept of $X_r - X_q$ 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 continuous in time with probability 1
  9. No point differentiable

- Return to the unit root test. $X_t = X_0 + \sum_{j=1}^{t} e_j$. 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: $Y_T(\cdot) \xrightarrow{d} B(\cdot)$
  2. Continuous mapping theorem: if f is a continuous function, $f(Y_T(t)) \xrightarrow{d} f(B(t))$

- Note that $Y_T(1) \xrightarrow{d} N(0,1)$, $Y_T(1/2) \xrightarrow{d} N(0,1/2)$, $Y_T(r) \xrightarrow{d} N(0,r)$, $\forall r \in [0,1]$.
- $\forall r_1, r_2 \in [0,1]$, $r_2 > r_1$, $Y_T(Tr_2) - Y_T(Tr_2) \xrightarrow{d} N(0, r_2 - r_1)$.
- $\sum_{t=1}^{T} X_t = \sum_{t=1}^{T} p_{t-1} + \sum_{t=1}^{T} e_t$ if $X_0 = 0$.
- $\int_{(j-1)/T}^{j/T} Y_T(t)dt = \frac{p_{j-1}}{\sigma T\sqrt{T}} \Longrightarrow \sum p_{j-1} = \sigma T\sqrt{T} \sum_{j=1}^{T} \int_{(j-1)/T}^{j/T} Y_T(t)dt = \sigma T\sqrt{T} \int_0^1 Y_T(t)dt$
- $\sum_{t=1}^{T} X_t = \sigma T\sqrt{T} \int_0^1 Y_T(t)dt + \sqrt{T}\frac{\sum e_j}{\sqrt{T}} \Longrightarrow \frac{\sum_{t=1}^{T} X_t}{T\sqrt{T}} = \sigma \int_0^1 Y_T(t)dt + \frac{1}{T}\frac{\sum e_j}{\sqrt{T}}$
- By FCLT and CMT, $\frac{\sum_{t=1}^{T} X_t}{T\sqrt{T}} \xrightarrow{d} \sigma \int_0^1 B(t)dt$

- $\sum_{t=1}^{T} X_{t-1}^2 = X_0^2 + \cdots + X_{T-1}^2 = X_1^2 + \cdots + X_{T-1}^2 = \sum_{t=1}^{T} X_t^2 - X_T^2 = \sum_{t=1}^{T} p_t^2 - p_T^2$
- 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, $\frac{\sum_{t=1}^{T} X_{t-1}^2}{T^2} \xrightarrow{d} \sigma^2 \int_0^1 B^2(t)dt$
- Under H_0, $T(\widehat{\phi} - \phi) \xrightarrow{d} \frac{1/2\left(\chi_{(1)}^2 - 1\right)}{\int_0^1 B^2(t)dt}$
- 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.

1. $T^{-1/2}\sum_{t=1}^{T} e_t \xrightarrow{d} \sigma B(1); \sum_{t=1}^{T} e_t \sim O_p(T^{1/2})$

2. $T^{-1}\sum_{t=1}^{T} p_{t-1}e_t \xrightarrow{d} \frac{1}{2}\sigma^2 \left\{B^2(1) - 1\right\}; \sum_{t=1}^{T} p_{t-1}e_t \sim O_p(T)$

3. $T^{-3/2}\sum_{t=1}^{T} te_t \xrightarrow{d} \sigma B(1) - \sigma \int_0^1 W(r)dr; \sum_{t=1}^{T} te_t \sim O_p(T^{3/2})$

4. $T^{-3/2}\sum_{t=1}^{T} p_{t-1} \xrightarrow{d} \sigma \int_0^1 B(r)dr; \sum_{t=1}^{T} p_{t-1} \sim O_p(T^{3/2})$

5. $T^{-2}\sum_{t=1}^{T} p_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 B^2(r)dr; \sum_{t=1}^{T} p_{t-1}^2 \sim O_p(T^2)$

6. $T^{-5/2}\sum_{t=1}^{T} tp_{t-1} \xrightarrow{d} \sigma \int_0^1 rB(r)dr; \sum_{t=1}^{T} tp_{t-1} \sim O_p(T^{5/2})$

7. $T^{-3}\sum_{t=1}^{T} tp_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 rB^2(r)dr; \sum_{t=1}^{T} tp_{t-1}^2 \sim O_p(T^3)$

8. $\sum_{t=1}^{T} t = T(T+1)/2 \sim O(T^2)$

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

- In case 1, we test

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

- 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:

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

- The OLS estimator of β is defined by $\widehat{\beta}_T$, which is biased but consistent.

- The test statistic (known as the **DF** Z test or the **coefficient** statistic) is defined as $T\widehat{\beta}_T$. Under H_0, $T\widehat{\beta}_T \xrightarrow{d} \frac{1/2\left(\chi^2_{(1)}-1\right)}{\int_0^1 B^2(r)dr}$.

- $\frac{1/2\left(\chi^2_{(1)}-1\right)}{\int_0^1 B^2(r)dr}$ is not a normal distribution. See Table B.5 for the critical values of this distribution. The finite sample distribution of $\widehat{\beta}_T$ is not normal. See Table B.5 for the critical values for different sample sizes.

- Alternatively, we can use t-statistic $\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)}$. Under H_0, it can be shown that

$$\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)} = \frac{T\widehat{\beta}_T\left(T^{-2}\sum X^2_{t-1}\right)^{1/2}}{\widehat{\sigma}_T} \xrightarrow{d} \frac{1/2\left(\chi^2_{(1)}-1\right)}{\left[\int_0^1 B^2(r)dr\right]^{1/2}}.$$

  Although it is called the Dickey-Fuller t-statistic, it is no longer a t distribution. See Table B.6 for the critical values of this distribution and the critical values of the finite sample distribution for different sample sizes.

3.2.3 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:

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

- The DF Z test statistic is $T\widehat{\beta}_T$. Under H_0, $T\widehat{\beta}_T \xrightarrow{d} \frac{1/2(\chi^2_{(1)}-1)-B(1)\int_0^1 B(r)dr}{\int_0^1 B^2(r)dr-\left(\int_0^1 B(r)dr\right)^2}$. See Table B.5 for the critical values of this distribution and the critical values of the finite sample distribution for different sample sizes.

- Proof of the Asymptotic Distribution:

The OLS estimates are

$$\begin{bmatrix} \widehat{\alpha}_T \\ \widehat{\beta}_T \end{bmatrix} = \begin{bmatrix} T & \sum X_{t-1} \\ \sum X_{t-1} & \sum X_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum e_t \\ \sum X_{t-1} e_t \end{bmatrix} = \begin{bmatrix} O_p(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{bmatrix}^{-1} \begin{bmatrix} O_p(T^{1/2}) \\ O_p(T) \end{bmatrix}.$$

Let $g(T) = diag(T^{1/2}, T)$, then

$$\begin{bmatrix} T^{1/2}\widehat{\alpha}_T \\ T\widehat{\beta}_T \end{bmatrix} = g(T) \begin{bmatrix} T & \sum X_{t-1} \\ \sum X_{t-1} & \sum X_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum e_t \\ \sum X_{t-1} e_t \end{bmatrix}$$

$$= \left\{ g^{-1}(T) \begin{bmatrix} T & \sum X_{t-1} \\ \sum X_{t-1} & \sum X_{t-1}^2 \end{bmatrix} g^{-1}(T) \right\}^{-1} g^{-1}(T) \begin{bmatrix} \sum e_t \\ \sum X y_{t-1} e_t \end{bmatrix}$$

$$= \begin{bmatrix} 1 & T^{-3/2} \sum X_{t-1} \\ T^{-3/2} \sum X_{t-1} & T^{-2} \sum X_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2} \sum e_t \\ T^{-1} \sum X_{t-1} e_t \end{bmatrix}$$

$$\xrightarrow{d} \begin{bmatrix} 1 & \sigma \int_0^1 B(r)dr \\ \sigma \int_0^1 B(r)dr & \sigma^2 \int_0^1 B^2(r)dr \end{bmatrix}^{-1} \begin{bmatrix} \sigma B(1) \\ \frac{1}{2}\sigma^2 \left\{ B^2(1) - 1 \right\} \end{bmatrix}.$$

- Alternatively, we can use t-statistic $\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)} \xrightarrow{d} \frac{1/2(\chi^2_{(1)}-1) - B(1)\int_0^1 B(r)dr}{\left(\int_0^1 B^2(r)dr - \left(\int_0^1 B(r)dr\right)^2\right)^{1/2}}$ under H_0. Although it is called the Dickey-Fuller t-statistic, it is no longer a t distribution. See Table B.6 for the critical values of this distribution and the critical values of the finite sample distribution for different sample sizes.

3.2.4 Case 2b (H_0: Random Walk with Drift):

- What if the true model is a random walk with drift, ie, $X_t = \alpha + X_{t-1} + e_t = X_0 + t\alpha + \sum_{j=1}^{t} e_j$? ($\alpha \neq 0$)

- $\sum X_{t-1} = TX_0 + T(T-1)\alpha/2 + \sum p_{t-1} \sim O_p(T^2)$

- $T^{-2} \sum X_{t-1} \xrightarrow{p} \alpha/2$, $T^{-3} \sum X_{t-1}^2 \xrightarrow{p} \alpha^2/3$, $\sum X_{t-1}e_t \sim O_p(T^{3/2})$

- Let $g(T) = diag(T^{1/2}, T^{3/2})$, then

$$\left[\begin{array}{c} T^{1/2}\widehat{\alpha}_T \\ T^{3/2}\widehat{\beta}_T \end{array}\right] = \left\{ g^{-1}(T) \left[\begin{array}{cc} T & \sum X_{t-1} \\ \sum X_{t-1} & \sum X_{t-1}^2 \end{array}\right] g^{-1}(T) \right\}^{-1} \left[\begin{array}{c} T^{-1/2}\sum e_t \\ T^{-3/2}\sum X_{t-1}e_t \end{array}\right]$$

- The first term converges to $\left[\begin{array}{cc} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{array}\right]^{-1}$. Since $T^{-3/2}\sum X_{t-1}e_t \xrightarrow{p} T^{-3/2}\sum \alpha(t-1)e_t$, the second term converges to

$$\left[\begin{array}{c} T^{-1/2}\sum e_t \\ T^{-3/2}\sum X_{t-1}e_t \end{array}\right] \xrightarrow{d} N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \sigma^2 \left[\begin{array}{cc} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{array}\right] \right).$$

  And hence,

$$\left[\begin{array}{c} T^{1/2}\widehat{\alpha}_T \\ T^{3/2}\widehat{\beta}_T \end{array}\right] \xrightarrow{d} N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \sigma^2 \left[\begin{array}{cc} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{array}\right]^{-1} \right)$$

3.2.5 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:

  $$X_t = \alpha + \phi X_{t-1} + \delta t + e_t, e_t \overset{iid}{\sim} N(0, \sigma_e^2)$$

- $X_t = \alpha(1-\phi) + \phi(X_{t-1} - \alpha(t-1)) + (\delta + \phi\alpha)t + e_t = \alpha^* + \phi p_{t-1} + \delta^* t + e_t$, where $p_{t-1} = X_{t-1} - \alpha(t-1)$ is a pure random walk if $\phi = 1$ (ie $\beta = 0$) and $\delta = 0$.

- Let $g(T) = diag(T^{1/2}, T, T^{3/2})$, then

$$\begin{bmatrix} T^{1/2}\widehat{\alpha}_T^* \\ T\left(\widehat{\phi}_T - 1\right) \\ T^{3/2}\left(\widehat{\delta}_T^* - \alpha\right) \end{bmatrix} = \left\{ g^{-1}(T) \begin{bmatrix} T & \sum p_{t-1} & \sum t \\ \sum p_{t-1} & \sum p_{t-1}^2 & \sum tp_{t-1} \\ \sum t & \sum tp_{t-1} & \sum t^2 \end{bmatrix} g^{-1}(T) \right\}^{-1} \begin{bmatrix} T^{-1/2}\sum e_t \\ T^{-1}\sum p_{t-1}e_t \\ T^{-3/2}\sum te_t \end{bmatrix}$$

$$\xrightarrow{d} \begin{bmatrix} 1 & \sigma\int Bdr & 1/2 \\ \sigma\int B(r)dr & \sigma^2\int B^2dr & \sigma\int rBdr \\ 1/2 & \sigma\int rBdr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} \sigma B(1) \\ \frac{1}{2}\sigma^2\left[B^2(1) - 1\right] \\ \sigma\left[B(1) - \int Bdr\right] \end{bmatrix}$$

$$= \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & \int Bdr & 1/2 \\ \int Bdr & \int B^2dr & \int rBdr \\ 1/2 & \int rBdr & 1/3 \end{bmatrix}^{-1} \begin{bmatrix} B(1) \\ \frac{1}{2}\left[B^2(1) - 1\right] \\ B(1) - \int Bdr \end{bmatrix}.$$

- So the aymptotic distribution of $\widehat{\phi}_T$ (and hence $\widehat{\beta}_T$) is independent of σ and α.

- The Dickey-Fuller Z test statistic is $T\widehat{\beta}_T$. Under H_0, the asymptotic distribution is not a normal distribution. See Table B.5 for the critical values of the asymptotic distribution distribution and the critical values of the finite sample distribution for different sample sizes.

- Alternatively, we can use the Dickey-Fuller t-statistic $\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)}$ whose asymptotic distribution distribution is no longer a t distribution. 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

3.2.6 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.

3.3 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

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + e_t, e_t \sim iid(0, \sigma_e^2), \tag{3.4}$$

  then the DF test is not applicable.

- Model (3.4) can be represented by

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

  where $\phi = \sum_{i=1}^{p} \phi_i$

- So if $\phi = 1$, $\{X_t\}$ is a unit root process; if $|\phi| < 1$, $\{X_t\}$ 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, $\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)}$, can be used. It has the same asymptotic distribution as the DF t statistic.

3.4 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 $T\widehat{\beta}_T$. The other (so-called PP t test) is the analogue of $\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)}$. PP Z test has the same sampling distribution as $T\widehat{\beta}_T$ under all three cases in the previous section. PP t test has the same sampling distribution as $\frac{\widehat{\beta}_T}{\widehat{se}(\widehat{\beta}_T)}$ 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 $T\widehat{\beta}_T - \frac{1}{2}(T^2\widehat{\sigma}^2_{\widehat{\beta}_T} \div s_T^2)(\lambda^2 - \gamma_0)$, where $s_T^2 = (T-2)^{-1}\sum(X_t - \widehat{\alpha}_T - \widehat{\phi}_T X_{t-1})^2, \lambda = \sigma\Theta(1), \gamma_0 = E(u_t^2)$

- 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

3.5 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

$$X_t = \xi t + r_t + u_t \tag{3.5}$$

$$r_t = r_{t-1} + e_t \tag{3.6}$$

  where $e_t \overset{iid}{\sim} N(0, \sigma_e^2)$. 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

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

4 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)

5 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 $\phi > 1$.
- 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.

6 Explosive Models

- Model: $X_{t+1} = \phi X_t + \varepsilon_t, \phi > 1, X_0 = 0$. Let $\widehat{\phi}$ be the LS estimator of ϕ. We will show that when $\varepsilon_t \sim N(0, \sigma^2)$,

$$\frac{\phi^T}{\phi^2-1}\left(\widehat{\phi} - \phi\right) \Rightarrow Cauchy. \tag{6.7}$$

When $\varepsilon_t \overset{iid}{\sim} (0, \sigma^2)$ but is not necessarily normally distributed, then

$$\frac{\phi^T}{\phi^2-1}\left(\widehat{\phi} - \phi\right) \Rightarrow P/Q,$$

where P and Q are the limits of P_T and Q_T defined by

$$P_T = \sum_{t=1}^{T} \phi^{-t}\varepsilon_t \text{ and } Q_T = \sum_{t=1}^{T} \phi^{-(T-t)}\varepsilon_t. \tag{6.8}$$

- To prove (6.7), note that

$$X_t = \varepsilon_t + \phi\varepsilon_{t-1} + \ldots + \phi^{t-1}\varepsilon_1 + \phi^t X_0 = \varepsilon_t + \phi\varepsilon_{t-1} + \ldots + \phi^{t-1}\varepsilon_1.$$

- Let $P_t = X_t/\phi^t = \phi^{-t}\varepsilon_t + \phi^{-t+1}\varepsilon_{t-1} + \ldots + \phi^{-1}\varepsilon_1$ which is normally distributed. P_t is a martingale since $E(P_t|I_{t-1}) = \phi^{-t+1}\varepsilon_{t-1} + \ldots + \phi^{-1}\varepsilon_1 = P_{t-1}$ and $E(P_t^2) = \sum_{s=1}^{t} \phi^{-2s}\sigma^2 = \sum_{s=1}^{t} \sigma^2(1-\phi^{2(t-1)})/(1-\phi^t)$, and hence $\sup_t E(P_t^2) = \sum_{s=1}^{\infty} \sigma^2(1-\phi^{2(t-1)})/(1-\phi^2) = \sigma^2/(\phi^2-1)$.

- By the Martingale Convergence Theorem, $P_t \stackrel{a.s.}{\to} P = \phi^{-1}\varepsilon_1 + \ldots + \phi^{-t}\varepsilon_t + \ldots \sim N(0, \sigma^2/(\phi^2 - 1))$.

- Let L be an integer and satisfies $\frac{1}{L} + \frac{L}{T} \to 0$ as $T \to \infty$. Consider

$$\begin{aligned}
\frac{1}{\phi^{2T}} \sum_{t=1}^{T} X_{t-1}^2 &= \frac{1}{\phi^{2T}} \sum_{t=1}^{T} \left(\frac{X_{t-1}}{\phi^{t-1}}\right)^2 \phi^{2(t-1)} \\
&= \frac{1}{\phi^{2T}} \sum_{t=L+1}^{T} \left(\frac{X_{t-1}}{\phi^{t-1}}\right)^2 \phi^{2(t-1)} + \frac{1}{\phi^{2T}} \sum_{t=1}^{L} \left(\frac{X_{t-1}}{\phi^{t-1}}\right)^2 \phi^{2(t-1)} \\
&= \frac{1}{\phi^{2T}} \sum_{t=L+1}^{T} \left(P + o_{a.s.}(1)\right)^2 \phi^{2(t-1)} + O_{a.s.}\left(\frac{L\phi^{2L}}{\phi^{2T}}\right) \\
&= \frac{1}{\phi^{2T}} \sum_{t=1}^{T} P^2 \phi^{2(t-1)} + o_{a.s.}(1) = \frac{1}{\phi^{2T}} P^2 \frac{\phi^{2T} - 1}{\phi^2 - 1} \stackrel{a.s.}{\to} \frac{P^2}{\phi^2 - 1}.
\end{aligned}$$

- Consider

$$
\begin{aligned}
\frac{1}{\phi^T}\sum_{t=1}^{T} X_{t-1}\varepsilon_t &= \frac{1}{\phi^T}\sum_{t=1}^{T}\left(\frac{X_{t-1}}{\phi^{t-1}}\right)\varepsilon_t\phi^{t-1} \\
&= \frac{1}{\phi^T}\sum_{t=L+1}^{T}\left(\frac{X_{t-1}}{\phi^{t-1}}\right)\varepsilon_t\phi^{t-1} + \frac{1}{\phi^T}\sum_{t=1}^{L}\left(\frac{X_{t-1}}{\phi^{t-1}}\right)\left(\varepsilon_t\phi^{t-1}\right) \\
&= \frac{1}{\phi^T}\sum_{t=L+1}^{T}\left(P + o_{a.s.}(1)\right)\varepsilon_t\phi^{t-1} + O_{a.s.}\left(\frac{L\phi^L}{\phi^T}\right) \\
&= \frac{1}{\phi^T}\sum_{t=1}^{T} P\varepsilon_t\phi^{t-1} + o_{a.s.}(1) = P\frac{1}{\phi^T}\sum_{t=1}^{T}\varepsilon_t\phi^{t-1} + o_{a.s.}(1) \\
&= P\left[\phi^{-1}\varepsilon_t + \phi^{-2}\varepsilon_{t-1} + \ldots + \phi^{-T}\varepsilon_1\right] + o_{a.s.}(1).
\end{aligned}
$$

- $Q_t = \phi^{-1}\varepsilon_t + \phi^{-2}\varepsilon_{t-1} + \ldots + \phi^{-t}\varepsilon_1$ is normally distributed and $Q_t \overset{a.s.}{\to} Q \sim N(0, \sigma^2/(\phi^2-1))$.

- Note that $Cov(P_T, Q_T) = \sum_{t=1}^{T}\sigma^2/\phi^{T+1} = T\sigma^2/\phi^{T+1} \to 0$. So P_T and Q_T are asymptotically independent, and P and Q are independent.

- So

$$
\frac{\phi^T}{\phi^2-1}\left(\widehat{\phi} - \phi\right) = \frac{1}{\phi^2-1}\frac{\frac{1}{\phi^T}\sum_{t=1}^{T} X_{t-1}\varepsilon_t}{\frac{1}{\phi^{2T}}\sum_{t=1}^{T} X_{t-1}^2} \overset{a.s.}{\to} \frac{PQ}{P^2} = \frac{Q}{P} = Cauchy.
$$

- Asymptotic distribution of OLS estimator of AR coefficient

Model	Asymptotic distribution
$X_t = \phi X_t + \varepsilon_t, \lvert\phi\rvert < 1, X_0 \sim O_p(1)$	$\sqrt{T}\left(\widehat{\phi} - \phi\right) \Rightarrow N(0, 1-\phi^2)$
$X_t = d + \phi X_t + \varepsilon_t, \lvert\phi\rvert < 1, X_0 \sim O_p(1)$	$\sqrt{T}\left(\widehat{\phi} - \phi\right) \Rightarrow N(0, 1-\phi^2)$
$X_t = \phi X_t + \varepsilon_t, \phi = 1, X_0 \sim O_p(1)$	$T\left(\widehat{\phi} - 1\right) \Rightarrow \int BdB / \int B^2$
$X_t = d + \phi X_t + \varepsilon_t, \phi = 1, d = 0, X_0 \sim O_p(1)$	$T\left(\widehat{\phi} - 1\right) \Rightarrow \frac{\int BdB - B(1)\int B}{\int B^2 - \left[\int B\right]^2}$
$X_t = d + \phi X_t + \varepsilon_t, \phi = 1, d \neq 0, X_0 \sim O_p(1)$	$T^{3/2}\left(\widehat{\phi} - 1\right) \Rightarrow N\left(0, \frac{12\sigma^2}{d^2}\right)$
$X_t = \phi X_t + \varepsilon_t,\ X_0 = 0, \lvert\phi\rvert > 1, \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$	$\frac{\phi^T}{\phi^2 - 1}\left(\widehat{\phi} - \phi\right) \Rightarrow Cauchy$