02/08/2023

Course Title: Econometrics

Reference:
1. Time Series Analysis (Ch 8 & 9) by J D Hamilton
2. Introduction to Econometrics by Christopher Dougherty
3. Introduction to Statistical Learning with Applications in R.
4. Econometrics – Green.
5. Econometrics Methods for Panel Data by Badi Baltage

Topics / Course Content / Syllabus:

1. Simple Linear Regression – Only some basics } Dougherty
2. Multiple Linear Regression } Hamilton & Dougherty
   - (A) Gauss-Markov Assumption
     - (a) Heteroskedasticity
     - (b) Auto Correlation
     - (c) Endogenity
     - (d) Normality
   - (B) Hypothesis Testing
   - (C) Prediction
3. Regression Models for Binary Data } Dougherty
   - (A) Logit & Profit Models (Logistic Regression)
4. Ridge & Lasso } ISLR
5. Basics of Univariate Time Series Analysis } Hamilton & Shumway & Stoffer
6. Multivariate Time Series Analysis } Hamilton

Marks Distribution:

R or Python

1. Mid Sem → 30%
2. End Sem → 50%
3. Project → 20%
   - (A) Step 1: Define your problem.
   - (B) Step 2: According find or collect a data.
   - (C) Step 3: Do some literature survey.
   - (D) Step 4: Analyse the data
   - (E) Step 5: Summarize your data analysis results.

- At most 2 persons in group
- After the End Sem
- 10 mins per group
- Present the idea before to Sir.

Contact: (1) Gmail: Rajumaiti@gmail.com

Simple Linear Regression:

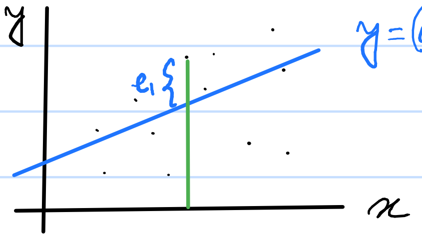
Sl	Study Var y	Explanatory Variable / Regressor / Covariate / Feature / Biomarker (x)
1	y_1	x_1
2	y_2	x_2
⋮	⋮	⋮
n	y_n	x_n

Objectives:

① How does a regressor (x) impact the study variable (y)

② Prediction (Given a value of x, you have to find the value of y)

Scatter Plot:



$y = \alpha + \beta x$

$x \uparrow \quad y \uparrow$

$y_i = \alpha + \beta x_i + e_i \leftarrow$ Statistics error $\quad i = 1(1)n$

$$e_1 = y_1 - \alpha - \beta x_1$$

$$\begin{pmatrix} \hat{\alpha}_{OLS} \\ \hat{\beta}_{OLS} \end{pmatrix} = \underset{(\alpha,\beta)}{\arg\min} \sum_{i=1}^{n} e_i^2 \quad \text{(Convex)}$$

$$= \underset{(\alpha,\beta)}{\arg\min} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\widehat{Cov}(x_i,y_i)}{\widehat{Var}(x_i)} = \frac{\widehat{Cov}(x_i,y_i)}{\widehat{Var}(x_i)\,\widehat{Var}(y_i)} \times \widehat{Var}(y_i)$$

$$= r \times \frac{s_y}{s_x} \quad \text{where, } r = \frac{\widehat{Cov}(x_i,y_i)}{\sqrt{\widehat{Var}(x_i)}\sqrt{\widehat{Var}(y_i)}}$$

And, $\hat{\alpha}_{OLS} = \bar{y} - \hat{\beta}_{OLS}\bar{x}$

$y_i = \alpha + \beta x_i + e_i$ and $y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{e}_i = \hat{y}_i + \hat{e}_i$

$\Rightarrow \hat{e}_i = y_i - \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

$\Rightarrow \boxed{\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i}$

How good is this regression line?

depends on $\sum_{i=1}^{n} \hat{e}_i^2$

What can we say about $\hat{\alpha}_{OL}$ & $\hat{\beta}_{OL}$ or what are the statistical properties $\hat{\alpha}_{OLS}$ & $\hat{\beta}_{OLS}$ have?

① Unbiasedness

Multiple Linear Regression:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + e_i, \quad i = 1(1)n$$

Here, you have a study variable y & k many regressors $(x_1, x_2, \dots, x_n)$

$\therefore \; y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i \;, \quad i = 1(1)n$

where,

$$\underset{\sim}{x}_i = \begin{bmatrix} 1 \\ x_{1i} \\ \vdots \\ x_{ki} \end{bmatrix}_{(k+1)\times 1}, \quad \underset{\sim}{\beta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1}$$

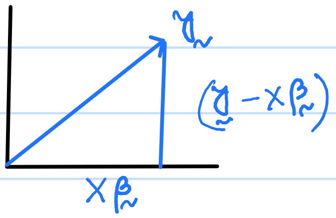
$$\therefore \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \underset{\sim}{x}_1^T \\ \underset{\sim}{x}_2^T \\ \vdots \\ \underset{\sim}{x}_n^T \end{bmatrix} \underset{\sim}{\beta} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \Rightarrow \underset{\sim}{y}^{n\times 1} = X^{n\times(k+1)} \underset{\sim}{\beta}^{(k+1)\times 1} + e^{n\times 1}$$

where,

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}_{n\times(k+1)}, \quad \underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}, \quad \underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\hat{\beta}_{OLS} = \arg\min_{\underset{\sim}{\beta}} \sum_{i=1}^{n} e_i^2$$

$$= \arg\min_{\underset{\sim}{\beta}} \underset{\sim}{e}^T \underset{\sim}{e} = \arg\min_{\beta} (\underset{\sim}{y} - X\underset{\sim}{\beta})^T (\underset{\sim}{y} - X\underset{\sim}{\beta})$$



$$\therefore (\underset{\sim}{y} - X\underset{\sim}{\beta})^T (X\underset{\sim}{\beta}) = 0$$

$$\Rightarrow \underset{\sim}{y}^T X \underset{\sim}{\beta} - \underset{\sim}{\beta}^T X^T X \underset{\sim}{\beta} = 0$$

$$\Rightarrow (\underset{\sim}{y}^T X - \underset{\sim}{\beta}^T X^T X)\underset{\sim}{\beta} = 0 \Rightarrow \underset{\sim}{y}^T X - \underset{\sim}{\beta}^T X^T X = 0 \quad [\text{Taking } \underset{\sim}{\beta} \neq 0]$$

$$\Rightarrow \underset{\sim}{\beta}^T X^T X = \underset{\sim}{y}^T X \Rightarrow (X^T X)\underset{\sim}{\beta} = X^T \underset{\sim}{y} \Rightarrow \underset{\sim}{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y} \quad [\text{If } (X^T X) \text{ is invertable}]$$

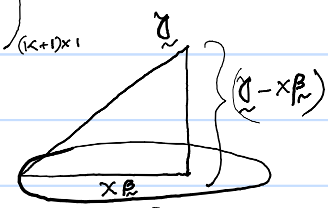
09/08/2023

Multiple Linear Regression:

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \quad \underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \left(\quad \right)_{n\times(k+1)}, \quad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad \underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}_{(k+1)\times 1}$$

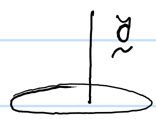
$$\min_{\underset{\sim}{\beta}} \sum_{i=1}^{n} \varepsilon_i^2$$

$$= \min_{\underset{\sim}{\beta}} \underset{\sim}{\varepsilon}^T \underset{\sim}{\varepsilon} = \min_{\underset{\sim}{\beta}} (\underset{\sim}{y} - X\underset{\sim}{\beta})^T (\underset{\sim}{y} - X\underset{\sim}{\beta}) \Rightarrow \underset{\sim}{\hat{\beta}}_{OLS} = (X^T X)^{-1} X^T \underset{\sim}{y}$$



$$(\underset{\sim}{y} - X\underset{\sim}{\beta})^T X\underset{\sim}{\beta} = 0$$

A vector $\underset{\sim}{y}$ is said to be orthogonal to a vector $\underset{\sim}{x}$ if $\underset{\sim}{y}^T \underset{\sim}{x} = 0$



A vector $\underset{\sim}{y}$ is orthogonal to a vector space V if $\underset{\sim}{y}^T \underset{\sim}{x} = 0 \;\forall\, x \in V$

$\Leftrightarrow \underset{\sim}{y}^T \underset{\sim}{x}_i = 0 \;\forall\, i = 1(1)p$, $\{x_1, x_2, \dots, x_p\}$ be the basis vector of V.

$$\therefore (\underset{\sim}{y} - X\underset{\sim}{\beta})^T X\underset{\sim}{\beta} = 0 \;\forall\, \underset{\sim}{\beta} \quad [\text{Orthogonality}]$$

$$\Leftrightarrow (\underset{\sim}{y} - X\underset{\sim}{\beta})^T \underset{\sim}{X}_{\cdot i} = 0 \;\forall\, i = 1(1)\,k+1$$

$$\Leftrightarrow (\underset{\sim}{y} - X\underset{\sim}{\beta})^T (\underset{\sim}{X}_1 \; \underset{\sim}{X}_2 \cdots \underset{\sim}{X}_{k+1}) = 0 \Leftrightarrow (\underset{\sim}{y} - X\underset{\sim}{\beta})^T X = 0 \Leftrightarrow (X^T X)\underset{\sim}{\beta} = X^T \underset{\sim}{y} \quad [\text{Set of Normal Equations}]$$

Properties of $\hat{\beta}_{OLS}$:

① **Unbiasedness:** $\hat{\underset{\sim}{\beta}}_{OLS} = (X^TX)^{-1}X^T\underset{\sim}{y}$, where X is non-stochastic & $\underset{\sim}{y}$ is stochastic.

$$E(\hat{\underset{\sim}{\beta}}_{OLS}) = E\left((X^TX)^{-1}X^T\underset{\sim}{y}\right) = (X^TX)^{-1}X^TE(\underset{\sim}{y}) = (X^TX)^{-1}X^TE(X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}) = (X^TX)^{-1}(X^TX)\underset{\sim}{\beta} + (X^TX)^{-1}X^TE(\underset{\sim}{\varepsilon})$$

$= \underset{\sim}{\beta}$ if we assume $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ (Gauss Markov (1) Assumption)

② **Efficiency:**

μ = Mean of Econometric Course (Parameter)
$\bar{X}$ = Sample Mean
$\bar{X} \longrightarrow \mu$
$Var(\bar{X}) = \sigma^2$
Efficiency $= \frac{1}{\sigma^2}$

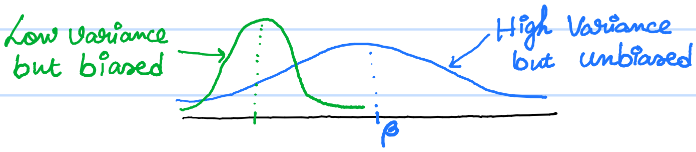
$$\begin{aligned}
Var(\hat{\underset{\sim}{\beta}}_{OLS}) &= Var\left((X^TX)^{-1}X^T\underset{\sim}{y}\right) = Var(A\underset{\sim}{y}) \\
&= E\left([A\underset{\sim}{y} - E(A\underset{\sim}{y})][A\underset{\sim}{y}-E(A\underset{\sim}{y})]^T\right) \\
&= E\left(A(\underset{\sim}{y}-E(\underset{\sim}{y}))(\underset{\sim}{y}-E(\underset{\sim}{y}))^TA^T\right) \\
&= A\,E\left((\underset{\sim}{y}-E(\underset{\sim}{y}))(\underset{\sim}{y}-E(\underset{\sim}{y}))^T\right)A^T = A\,Var(\underset{\sim}{y})A^T \\
&= (X^TX)^{-1}X^T\,Var(\underset{\sim}{y})\,X(X^TX)^{-1} \\
&= (X^TX)X^T\,Var(\underset{\sim}{\varepsilon})\,X(X^TX)^{-1} \quad [\because \underset{\sim}{y} = X\underset{\sim}{\beta}+\underset{\sim}{\varepsilon} \Rightarrow Var(\underset{\sim}{y}) = Var(\underset{\sim}{\varepsilon})] \\
&= (X^TX)^{-1}X^T\,\sigma^2I_n\,X(X^TX)^{-1} \quad [\text{Assume that, } Var(\underset{\sim}{\varepsilon}) = \sigma^2I_n] \\
&= \sigma^2\underset{(k+1)\times(k+1)}{(X^TX)^{-1}} \quad [\text{Variance-Covariance Matrix of } \hat{\underset{\sim}{\beta}}_{OLS}]
\end{aligned}$$

$$= \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0,\hat{\beta}_1) & \cdots & Cov(\hat{\beta}_0,\hat{\beta}_k) \\ & Var(\hat{\beta}_1) & \cdots & Cov(\hat{\beta}_1,\hat{\beta}_k) \\ & & & \vdots \\ & & & Var(\hat{\beta}_{k+1}) \end{pmatrix} \Rightarrow Var(\hat{\underset{\sim}{\beta}}_{OLS}) = \sigma^2(X^TX)^{-1}$$

$\boxed{\text{Result:}}$ Under the assumption (i) $E(\underset{\sim}{\varepsilon}) = 0$ & (ii) $Var(\underset{\sim}{\varepsilon}) = \sigma^2I_n$, $\hat{\beta}_{OLS}$ is the best linear unbiased estimator (BLUE).

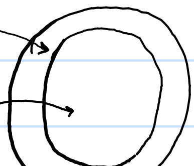
It means that suppose $\tilde{\underset{\sim}{\beta}}$ is another linear unbiased estimator of $\underset{\sim}{\beta}$ with Var-Cov matrix $\tilde{\Sigma}$ then $\tilde{\Sigma} - \sigma^2(X^TX)^{-1}$ will be positive definite. (Proof this, H.W.)

$\boxed{\text{Result:}}$ Suppose $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2I_n)$, $\hat{\beta}_{OLS}$ is the best unbiased estimator. (Think on it)



$\{\tilde{\underset{\sim}{\beta}} : E(\tilde{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}\}$

$\{\tilde{\underset{\sim}{\beta}} : \tilde{\underset{\sim}{\beta}} = B\underset{\sim}{y},\ E(\tilde{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}\}$



$\boxed{\text{Result:}}$ Under the assumption that $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2I_n)$, we can show that

① $\hat{\underset{\sim}{\beta}}_{OLS} \sim N(\underset{\sim}{\beta}, \sigma^2(X^TX)^{-1})$

② RSS = Residual Sum of Squares for OLS $= \hat{\underset{\sim}{\varepsilon}}^T\hat{\underset{\sim}{\varepsilon}} = (\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_{OLS})^T(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_{OLS}) = \underset{\sim}{y}^T(I_n - P_x)\underset{\sim}{y}$

where, $P_x = X(X^TX)^{-1}X^T$, $RSS \sim \chi^2_{n-k-1}$

$\boxed{\text{H.W.}}$ ① Show that P_x is symmetric & idempotent.
② Find the rank of $I - P_x$, $R(I-P_x) = \begin{Bmatrix} k+1 \\ n-k-1 \end{Bmatrix}$ $[I - P_x = PD_\lambda P^T]$.

$\boxed{\text{Result:}}$ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2I_n)$, $\underbrace{\hat{\beta}_{OLS} \sim N(\underset{\sim}{\beta}, \sigma^2(X^TX)^{-1})}$, $\underbrace{\frac{RSS}{\sigma^2} \sim \chi^2_{n-k-1}}$

Independent

$Cov\left(\hat{\beta}_{OLS}, \frac{RSS}{\sigma^2}\right) = 0$

$\Leftrightarrow Cov\left(\hat{\beta}_i, \frac{RSS}{\sigma^2}\right) = 0 \ \forall\ i = 1(1)\ k+1$

Ref: ① Hamilton Ch-8
② Shalabh - iit K

$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ where ① $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$

② $\lim_{n\to\infty} \left(\frac{X^TX}{n}\right) = Q$, finite and non-singular matrix

(Consistency of $\hat{\underset{\sim}{\beta}}_{OLS}$ i.e. $\hat{\underset{\sim}{\beta}}_{OLS} \xrightarrow{p} \underset{\sim}{\beta}$)

$\boxed{\text{Result:}}$ $H_0: \beta_l = 0$ vs $H_1: \beta_l \neq 0$ Test Statistics?

16/08/2023

③ $\hat{\underset{\sim}{\beta}}_{OLS}$ is consistant estimator of $\underset{\sim}{\beta}$.

Proof: $\hat{\underset{\sim}{\beta}}_{OLS} = (X^TX)^{-1}X^TY$

$= (X^TX)^{-1}X^T(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}) = \underset{\sim}{\beta} + \underbrace{\left(\frac{X^TX}{n}\right)^{-1}}_{\parallel\, Q}\underbrace{\left(\frac{X^T\underset{\sim}{\varepsilon}}{n}\right)}_{\downarrow p\ 0} \xrightarrow{p} \underset{\sim}{\beta}$

$E(\hat{\underset{\sim}{\beta}}_{OLS}) = \underset{\sim}{\beta}$

$Var(\hat{\underset{\sim}{\beta}}_{OLS}) \to 0$ then $\hat{\underset{\sim}{\beta}}_{OLS} \xrightarrow{p} \underset{\sim}{\beta}$

Preq: $X_n \xrightarrow{p} X$

$\Leftrightarrow P(|X_n - X| > \varepsilon) \to 0$ as $n\to\infty$

$Var(\hat{\beta}_{OLS}) = \sigma^2(X^TX)^{-1}$

$= \frac{\sigma^2}{n}\left(\frac{X^TX}{n}\right)^{-1} \to 0$ as $n\to\infty$

Assume that, $\lim_{n\to\infty}\left(\frac{X^TX}{n}\right) = Q$, finite and non-singular

Testing of Hypothesis:

$H_0: \beta_l = 0$ vs $H_1: \beta_l \neq 0$

Test Statistic $(T) = \frac{\hat{\beta}_i - 0}{SE(\hat{\beta}_i)}$, $SE(\hat{\beta}_i) = \sqrt{(\sigma^2(X^TX)^{-1})_{(i,i)}} = \sigma\sqrt{\xi_{(i,i)}}$, $i = 0(1)k$

$= \frac{\hat{\beta}_i - 0}{\sigma\sqrt{\xi_{(i,i)}}}$

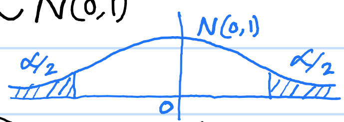
H.W. Why this test statistic is used to test the above hypothesis.

$\hat{\underset{\sim}{\beta}}_{OLS} = \begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \vdots\\ \hat{\beta}_i\\ \vdots\\ \hat{\beta}_k\end{pmatrix}$, $(X^TX)^{-1} = \begin{bmatrix}\Sigma_{00} & \Sigma_{01} & \cdots & \Sigma_{0k}\\ \Sigma_{10} & \Sigma_{11} & \cdots & \Sigma_{1k}\\ \vdots & \vdots & & \vdots\\ \Sigma_{k0} & \Sigma_{k1} & \cdots & \Sigma_{kk}\end{bmatrix}$

$\xi_{(i,i)}$ is the ith diagonal element of $(X^TX)^{-1}$

Dist. of T under H_0:

① Suppose σ^2 is known, $T = \frac{\hat{\beta}_i}{\sigma\sqrt{\xi_{(i,i)}}} \overset{H_0}{\sim} N(0,1)$



Rejection Rule: $\omega_0 : |T| > z_{\alpha/2}$

Critical Region/Rejection Region

upper $\alpha/2$ th point of $N(0,1)$

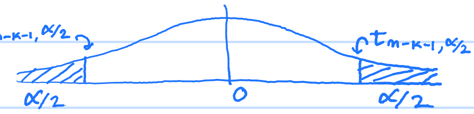
② Suppose σ^2 is unknown, $\hat{\underset{\sim}{\beta}}_{OLS} \sim N(\underset{\sim}{\beta}, \sigma^2(X^TX)^{-1})$, $\frac{RSS}{\sigma^2} \sim \chi^2_{n-k-1}$

Independent

$E\left(\frac{RSS}{n-k-1}\right) = \sigma^2$ (H.W.)

Hence, $\frac{RSS}{n-k-1}$ is unbiased estimator of σ^2

$T = \frac{\hat{\beta}_i - 0}{\hat{\sigma}\sqrt{\xi_{(i,i)}}} \overset{H_0}{\sim} t_{n-k-1}$ (H.W.) where $\hat{\sigma}^2 = \frac{RSS}{n-k-1}$ is an unbiased estimator of σ^2



$E(\hat{\sigma}^2) = \sigma^2$, Rejection Region: $\omega_0 : |T| > t_{n-k-1, \alpha/2}$

$H_0: \beta_1 - 2\beta_2 + \beta_3 = 0$

$-\beta_3 + \beta_4 - \beta_5 = 0$

$\Rightarrow R\underset{\sim}{\beta} = \underset{\sim}{0}$

$R = \begin{bmatrix}0 & 1 & -2 & 1 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & -1 & 1 & -1 & \cdots & 0\end{bmatrix}$

$R\underset{\sim}{\beta} \sim N(R\underset{\sim}{\beta}, \underbrace{\sigma^2 R(X^TX)^{-1}R^T}_{\Sigma})$

$H_0: \beta_1 - \beta_2 = 0$ & $\beta_3 - \beta_4 = 0$ vs $H_1: H_0$ is not true

$\Leftrightarrow H_0: R\underset{\sim}{\beta} = \underset{\sim}{0}$ vs $H_1: H_0$ is not true

where, $R = \begin{bmatrix}0 & 1 & -1 & 0 & 0 & \cdots\\ 0 & 0 & 0 & 1 & -1 & \cdots\end{bmatrix}_{2\times(k+1)}$

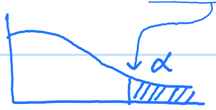
$\Rightarrow (R\hat{\underset{\sim}{\beta}}_{OLS} - R\underset{\sim}{\beta})^T \Sigma^{-1}(R\hat{\underset{\sim}{\beta}}_{OLS} - R\underset{\sim}{\beta}) \sim \chi^2_m$, $m = \text{rank}(\Sigma)$.

(H.W.) (Hamilton)

$\frac{RSS}{\sigma^2} \sim \chi^2_{n-k-1} \Rightarrow T = \frac{(R\hat{\underset{\sim}{\beta}}_{OLS} - R\underset{\sim}{\beta})^T\Sigma^{-1}(R\hat{\underset{\sim}{\beta}}_{OLS} - R\underset{\sim}{\beta})/m}{\frac{RSS}{\sigma^2}/(n-k-1)} \sim F_{m,n-k-1}$

$\left[\because \frac{\chi^2_m/m}{\chi^2_n/n} \sim F_{m,n}\right]$

$\Rightarrow \omega_0 : T > F_{\alpha, m, n-k-1}$



Heteroskedasticity:

$$y_i = x_i^T \underset{\sim}{\beta} + \varepsilon_i \Rightarrow \boxed{\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}}\ ,\ X \text{ is non-stochastic}$$

$$\Rightarrow \hat{\underset{\sim}{\beta}}_{OLS} = (X^TX)^{-1}X^TY$$

① $E(\hat{\underset{\sim}{\beta}}_{OLS}) = \underset{\sim}{\beta}$, under the assumption that $E(\underset{\sim}{\varepsilon}) = 0$

② $Var(\hat{\underset{\sim}{\beta}}_{OLS}) = \sigma^2(X^TX)^{-1}$, Under the assumption that $Var(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$ (Homoskedastic Variance)

③ Under the above two assumptions i.e. $E(\underset{\sim}{\varepsilon}) = 0$ and $Var(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$
$\hat{\underset{\sim}{\beta}}_{OLS}$ is the BLUE of $\underset{\sim}{\beta}$

④ Under the normality of $\underset{\sim}{\varepsilon}$ i.e. $\underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I_n)$
$\hat{\underset{\sim}{\beta}}_{OLS}$ is the best unbiased estimator of $\underset{\sim}{\beta}$.

⑤ If we assume that $\lim_{n\to\infty}\left(\frac{X^TX}{n}\right) = Q$ (finite and non-singular) then $\hat{\underset{\sim}{\beta}}_{OLS} \xrightarrow{P} \underset{\sim}{\beta}$

$\boxed{\text{Case I}}$ X is non-stochastic

$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ where X is non-stochastic

And the Gauss-Markov assumptions are - ① $\underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I_n)$, ② $\lim_{n\to\infty}\left(\frac{X^TX}{n}\right) = Q$

finite & non-singular

$\boxed{\text{Case II}}$ X is stochastic

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$$\Rightarrow \hat{\underset{\sim}{\beta}}_{OLS} = \arg\min_{\underset{\sim}{\beta}} (\underset{\sim}{Y} - X\underset{\sim}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim}{\beta}) = (X^TX)^{-1}X\underset{\sim}{Y}$$

$$\left[\begin{array}{l} \because E(X) = E(E(X|Y));\ \because E(XY|X) = XE(Y|X) \\ \because Var(X) = Var(E(X|Y)) + E(Var(X|Y)) \end{array}\right]$$

$$\Rightarrow E(\hat{\underset{\sim}{\beta}}_{OLS}) = E((X^TX)^{-1}X^TY) = E\left(E\left((X^TX)^{-1}X^TY|X\right)\right)$$

$$= E\left((X^TX)^{-1}X^TE(Y|X)\right) = E\left((X^TX)^{-1}X^TE(X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}|X)\right) = E\left(\underset{\sim}{\beta} + (X^TX)^{-1}X^TE(\underset{\sim}{\varepsilon}|X)\right)$$

$$= \underset{\sim}{\beta} + E\left((X^TX)^{-1}X^TE(\underset{\sim}{\varepsilon}|X)\right)$$

If we assume that $E(\underset{\sim}{\varepsilon}|X) = 0\ \forall X$ then $E(\hat{\underset{\sim}{\beta}}_{OLS}) = \underset{\sim}{\beta}$

$Var(\hat{\underset{\sim}{\beta}}_{OLS}|X) = \sigma^2(X^TX)^{-1}$, under the assumption that $Var(\underset{\sim}{\varepsilon}|X) = \sigma^2 I_n$

If we assume the normality of $\underset{\sim}{\varepsilon}$ i.e. $\underset{\sim}{\varepsilon} \sim N(0, \sigma^2 I_n)$ then $\hat{\underset{\sim}{\beta}}_{OLS}|X \sim N(\underset{\sim}{\beta}, \sigma^2(X^TX)^{-1}) \Rightarrow \hat{\underset{\sim}{\beta}}_{OLS} = ??$

$$RSS = (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}_{OLS})^T(\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}_{OLS}) = \underset{\sim}{Y}^T(I - P_X)\underset{\sim}{Y}$$

$$\frac{RSS}{\sigma^2}\Big|X \sim \chi^2_{n-k-1}\ \text{(Does not depend on } X) \Rightarrow \frac{RSS}{\sigma^2} \sim \chi^2_{n-k-1}$$

$\boxed{\text{Result}}$ $\hat{\beta}_{OLS}|X \sim N(\underset{\sim}{\beta}, \sigma^2(X^TX)^{-1})$, $\frac{RSS}{\sigma^2} \sim \chi^2_{n-k-1}$ (Both are Independent)

$\boxed{\text{Result}}$ $H_0: \beta_i = 0$ vs $H_1: \beta_i \neq 0$

$$T = \frac{\hat{\beta}_i - 0}{\sigma\sqrt{\xi_{(i,i)}}}\Big|X \overset{H_0}{\sim} N(0,1) \text{ if } \sigma^2 \text{ is known},\ \omega_0: |T| > z_{\alpha/2}$$

$$T = \frac{\hat{\beta}_i}{\sqrt{\frac{RSS}{n-k-1}}\sqrt{\xi_{(i,i)}}}\Big|X \overset{H_0}{\sim} t_{n-k-1},\ \omega_0: |T| > t_{\alpha/2, n-k-1},\ (X^TX)^{-1} = ((\xi_{ii}))_{i=0}^{k}$$

Heteroskedasticity:

$y_i = X_i^T\underset{\sim}{\beta} + \varepsilon_i$, $i = 1(1)n$

$\Rightarrow \underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

where, ① X is stochastic
② $E(\underset{\sim}{\varepsilon}|X) = \underset{\sim}{0}$
③ $Var(\underset{\sim}{\varepsilon}|X) = \sigma^2 I_n$ $[\because Var(\varepsilon_i|X) = \sigma^2$ and $cov(\varepsilon_i, \varepsilon_j|X) = 0\ \forall i \neq j]$
④ $P\left(\lim_{n\to\infty}\left(\frac{X^TX}{n}\right)\right) = Q$ (finite and stochastic or non-stochastic?)

H.W.

① Draw the pdf of $\hat{\beta}_{1,OLS}$ for $n = 100$
② Draw the pdf of $\hat{\beta}_{1,OLS}$ for $n = 1000$
③ Compute the bias and variance of $\hat{\beta}_{1,OLS}$ from ① & ②
④ Draw the pdf of $\hat{\beta}_{1,OLS} + 2$ for $n = 100$ and $n = 1000$

23/08/2023

Heteroskedasticity:

$$y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \varepsilon_i \quad , \quad x_i^T \ \forall i \text{ is stochastic}$$

where, (i) $E(\varepsilon_i | \underset{\sim}{x}_i) = 0$
(ii) $Var(\varepsilon_i | \underset{\sim}{x}_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j | \underset{\sim}{x}_i, \underset{\sim}{x}_j) = 0 \ \forall i \neq j$
(iii) $\varepsilon_i | \underset{\sim}{x}_i \overset{iid}{\sim}$ Normal
(iv) $P\left(\lim_{n\to\infty} \frac{X^T X}{n}\right) = Q$, a finite & non-singular

Under the above assumption,
$\hat{\beta}_{OLS}$ is unbiased, consistent and best linear estimator.

So, we can use $\hat{\beta}_{OLS}$ for — (i) Prediction (Prediction of y_i given $\underset{\sim}{x}_i$) [i.e. $\hat{y}_i = \underset{\sim}{x}_i^T \hat{\beta}_{OLS}$]
(ii) Testing of hypothesis

Confidence Interval of y_i is $\hat{y}_i \pm 1.96 \sqrt{SE(\hat{y}_i)}$

If we assume that σ_i^2's are known, $\frac{y_i}{\sigma_i} = \frac{\underset{\sim}{x}_i^T \underset{\sim}{\beta}}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \Rightarrow y_i^* = \underset{\sim}{x}_i^{*T} \underset{\sim}{\beta} + \varepsilon_i^*$

Here all Gauss-Markov assumptions for ε_i^* are satisfied.

$$\hat{\underset{\sim}{\beta}}_{WLS} = (X^{*T} X^*) X^{*T} Y^* \quad , \quad X^*_{n\times(k+1)} = \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix}_{n\times n} X = \Sigma^{-1/2} X \quad , \quad \underset{\sim}{y}^* = \begin{pmatrix} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_n} \end{pmatrix} \underset{\sim}{y}$$

$$= (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \underset{\sim}{y} \longleftarrow \text{Unbiased, Consistent \& BLUE}$$

Where, $\Sigma = diag(\sigma_1^2, \ldots, \sigma_n^2)$

However, in practice, σ_i^2's are not known.

# of parameters $= K+1+n = (\beta_0, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_n^2)$
# of observations $= n = (y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$

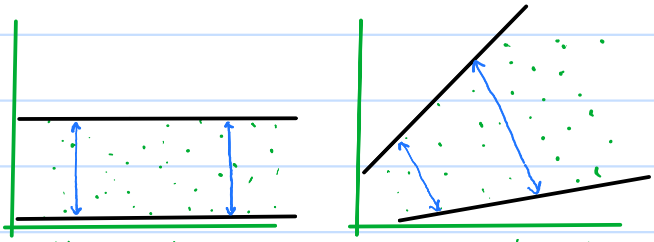
Q1: Given a data $\{(y_1, x_1), \ldots, (y_n, x_n)\}$. How do we know that the regression model has heteroskedastic error variance.

Q2: If we know that the error variances are heteroskedastic then what are the solutions or what are the alternatives of $\hat{\beta}_{OLS}$.

Q3: What are the sources for heteroskedastic error variance? (Just for knowledge)

Ans 1: $y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \varepsilon_i \Rightarrow \hat{\beta}_{OLS} = (X^T X)^{-1} X \underset{\sim}{y} \Rightarrow \hat{\varepsilon}_i = y_i - \underset{\sim}{x}_i^T \hat{\beta}_{OLS}$

$(\hat{\varepsilon}_i, y_i)$
$\forall i = 1(1)n$



Homoskedasticity   Heteroskedasticity

Graphical Representation to check the heteroskedasticity.

Analytical Way to find the heteroskedasticity:

$n+1+K > n$ [ie. we are trying to estimate $n+1+K$ parameters using n data points]

Then the remedy is to reduce the number of parameter.

$\sigma_i^2 = \delta_0 + \delta_1 x_{1i} + \dots + \delta_p x_{pi} \quad \forall i = 1(1)n \quad \Rightarrow \#$ of parameters $= K+1+p+1$

$y_i = \alpha + \beta x_i$
We can estimate α & β using values of y_i & x_i and using least square of that.
If any one of y_i & x_i is not known, OLS is not effective.

$\boxed{\text{eg.}}$ Let, $\sigma_i^2 = \delta_0 + \delta_1 x_{1i}$ Unknown Known

$\Rightarrow \#$ of parameters $= K+1+2 = K+3$
$\#$ of observation $= n$

Idea: $\hat{\varepsilon}_i = y_i - \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}_{OLS}$, $\hat{\varepsilon}_i^2$ can be used as an estimator of σ_i^2.
or $|\hat{\varepsilon}_i|$ " " " " " " σ_i^2.

So, $\hat{\sigma}_i^2 = \hat{\varepsilon}_i^2 \Rightarrow \hat{\sigma}_i^2 = \delta_0 + \delta_1 x_{1i} = \hat{\varepsilon}_i^2$ ie. $\begin{pmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{pmatrix} = \arg\min_{\delta_0, \delta_1} \sum (\hat{\varepsilon}_i - \delta_0 - \delta_1 x_{1i})^2$
Unknown Known

$\hat{\sigma}_i^2 = \hat{\delta}_0 + \hat{\delta}_1 x_{1i} \quad \forall i = 1(1)n$

$\varepsilon_i | X \overset{iid}{\sim} N(0, \sigma^2)$, $\hat{\sigma}^2 = \frac{1}{n-K-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \longrightarrow \sigma^2$ ie. $\hat{\varepsilon}_i^2 \longrightarrow \sigma^2$ for large n.

$\hat{\sigma}_i^2 = \hat{\delta}_0 + \hat{\delta}_1 x_{1i} \quad \forall i = 1(1)n$

H_0: Error Variance is Homoskedastic , H_1: Error Variance is Heteroskedastic

$\Leftrightarrow H_0: \delta_1 = 0$, $H_1: \delta_1 \neq 0$

(1) LM Test: $\sigma_i^2 = \delta_0 + \delta_1 z_i^2$

$H_0: \delta_1 = 0$ vs $H_1: \delta_1 \neq 0$

Algorithm: Step 1: Fit $y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \varepsilon_i$ using OLS & $\varepsilon_i = y_i - \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}_{OLS}$

Step 2: Fit $\hat{\varepsilon}_i^2 = \delta_0 + \delta_1 z_i^2 + u_i$ using OLS
& Compute $\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (\hat{\varepsilon}_i - \hat{\delta}_0 - \hat{\delta}_1 z_i^2)^2 = R_u^2$ Redundent Step $\xrightarrow{H_0}$ $\hat{\varepsilon}_i^2 = \delta_0 + u_i \Rightarrow \hat{\delta}_0 = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \hat{\bar{\varepsilon}}^2$

Step 3: Under H_0, $n R_u^2 \sim \chi_{n-1}^2 \Rightarrow n \sum_{i=1}^{n} (\hat{\varepsilon}_i^2 - \hat{\bar{\varepsilon}}^2)^2 \sim \chi_{n-1}^2$

Step 4: $n R_u^2 > \chi_{\alpha, n-1}^2 \longleftarrow$ Rejection Rule
If $n R_u^2 > \chi_{\alpha, n-1}^2$ then reject H_0, otherwise accept H_0.

(2) Breuch - Pagan Test:

Assumption: $\sigma_i^2 = \delta_0 + \delta_1 z_{1i} + \delta_2 z_{2i} + \dots + \delta_p z_{pi}$

$H_0: \delta_1 = \delta_2 = \dots = \delta_p = 0$ VS H_1: H_0 is not true [ie. $\exists\, \delta_i \neq 0$ for some $i = 1(1)p$]

(3) Park Test:

$\ln \sigma_i^2 = \delta_0 + \delta_1 z_{1i} + \dots + \delta_p z_{pi}$

$\Leftrightarrow \sigma_i^2 = e^{\delta_0 + \delta_1 z_{1i} + \dots + \delta_p z_{pi}}$

Advantage: $e^{\delta_0 + \dots + \delta_p z_{pi}} > 0$ and $\sigma_i^2 > 0$
For all other tests, $\delta_0 + \dots + \delta_p z_{pi}$ may less than 0 and $\sigma_i^2 > 0$.

$H_0: \delta_1 = \delta_2 = \dots = \delta_p = 0$ VS H_1: H_0 is not true

Multiple Linear Regression:

30/08/2023

- ⓐ Gauss Markov Assumption & properties of OLS
- ⓑ Test based on $\hat{\beta}_{OLS}$, $H_0: R\beta = 0$ vs $H_1: R\beta \neq 0$
- ⓒ GM - Homoskedasticity of error variance.
  - (i) Consequences if the heteroskedasticity assumption is not true
  - (ii) How to know that the error variances are heteroskedastic.
  - (iii) Suppose the error variances are heteroskedastic. Then how to improve the estimator of β over $\hat{\beta}_{OLS}$?

Ans: $y_i = \underset{\sim}{x}_i^{(k+1)\times 1} \underset{\sim}{\beta} + \varepsilon_i$, $\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2(\underset{\sim}{x}_i))$, where $\sigma^2(\underset{\sim}{x}_i) = \alpha_0 + \sum_{j=1}^{m} \alpha_j x_{ij}$, $\alpha_0 > 0$, $\alpha_j \geq 0$

MLE: $y_i | \underset{\sim}{x}_i \sim N(\underset{\sim}{x}_i^T \underset{\sim}{\beta}, \sigma^2(x_i))$ $\forall i = 1(1)n$

$$f(y_i|\underset{\sim}{x}_i) = \frac{1}{\sqrt{2\pi}\,\sigma(x_i)} e^{-\frac{1}{2\sigma^2(x_i)}(y_i - \underset{\sim}{x}_i^T \underset{\sim}{\beta})^2}$$

Likelihood: $L(\theta) = \prod_{i=1}^{n} f(y_i|\underset{\sim}{x}_i)$ where $\underset{\sim}{\theta} = \begin{pmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\alpha} \end{pmatrix}$

Log-Likelihood: $l(\theta) = \log(L(\underset{\sim}{\theta})) = \sum_{i=1}^{n} \log f(y_i|\underset{\sim}{x}_i) = -\frac{1}{2}\sum_{i=1}^{n}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\sigma^2(\underset{\sim}{x}_i) - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - \underset{\sim}{x}_i^T\underset{\sim}{\beta})^2}{\sigma^2(\underset{\sim}{x}_i)}$

$$= -\sum_{i=1}^{n} \log\sqrt{2\pi} - \frac{1}{2}\sum_{i=1}^{n}\log(\underset{\sim}{\alpha}^T \underset{\sim}{z}_i) - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\underset{\sim}{\alpha}^T z_i}(y_i - \underset{\sim}{x}_i^T\underset{\sim}{\beta})^2$$

$\sum_{i=1}^{n}(y_i - \underset{\sim}{x}_i^T \underset{\sim}{\beta})^2 \Rightarrow \hat{\underset{\sim}{\beta}}_{OLS} = (X^TX)^{-1}X^Ty$ | $\frac{\partial}{\partial\underset{\sim}{\theta}}(l(\theta)) = l'(\theta) = 0$ | $\hat{\theta}_{MLE} = \arg\max_{\theta} l(\theta)$

$l'(\underset{\sim}{\theta})$ be a vector, $l''(\underset{\sim}{\theta}) = \frac{\partial^2}{\partial\theta\partial\theta^T}(l(\underset{\sim}{\theta}))$ be a matrix.

Newton-Raphson: $\underset{\sim}{\theta}^{(t+1)} = \underset{\sim}{\theta}^{(t)} + (-l''(\underset{\sim}{\theta}^{(t)}))^{-1} l'(\underset{\sim}{\theta}^{(t)})$

Initial guess of $\underset{\sim}{\theta}$ is (say) $\underset{\sim}{\theta}^{(0)}$ if $|\underset{\sim}{\theta}^{(t+1)} - \underset{\sim}{\theta}^{(t)}| < \varepsilon$ then $\hat{\underset{\sim}{\theta}}_{MLE} = \underset{\sim}{\theta}^{(t+1)}$

When errors are auto-correlated:

Note: One of the Gauss Markov assumptions is that errors are correlated. But in practice it may happen that they are correlated, e.g., $y_t = \underset{\sim}{x}_t^T \underset{\sim}{\beta} + \varepsilon_t = \beta_0 + \beta_1 x_{t-1} + \dots + \beta_K x_{t-K} + \varepsilon_t$, where $\varepsilon_t = \phi\varepsilon_{t-1} + u_t$, $u_t \overset{iid}{\sim} N(0, \sigma^2)$

(i) Consequences: $\hat{\beta}_{OLS}$
- Unbiased ✓
- Consistent ✓ (H.W.)
- Efficient ✗

(ii) How to test whether errors are correlated or not?

$H_0: \phi = 0$ vs $H_1: \phi \neq 0$

Durbin Watson Test: $d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})}{\sum_{t=1}^{n}\varepsilon_t^2} = \frac{\frac{1}{n}\left(\sum_{t=2}^{n}\varepsilon_t^2 - 2\sum_t \varepsilon_t\varepsilon_{t-1} + \sum_t \varepsilon_{t-1}^2\right)}{\frac{1}{n}\left(\sum_t \varepsilon_t^2\right)} = \frac{Var(\varepsilon_t) - 2Cov(\varepsilon_t, \varepsilon_{t-1}) + Var(\varepsilon_{t-1})}{Var(\varepsilon_t)}$

$= 2\left(1 - \frac{Cov(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{Var(\varepsilon_t)Var(\varepsilon_{t-1})}}\right) = 2(1 - Cor(\varepsilon_t, \varepsilon_{t-1})) = 2(1-\rho)$ [Using Weak Law of Large Number, $|\phi| < 1$ i.e. ε_t is stationary]

ρ	0	1	-1
d	2	0	4

When H_0 is true, then the value of d should be ρ.

(iii) If error part is correlated, then how can we estimate $\underset{\sim}{\beta}$ efficiently?

Ref: Last Chap. of Dougherty (Auto Correlation), $y_t = \underset{\sim}{x}_t^T \underset{\sim}{\beta} + \varepsilon_t$, $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, $\rho y_{t-1} = \underset{\sim}{x}_{t-1}^T(\rho\underset{\sim}{\beta}) + \rho\varepsilon_{t-1}$

$y_t - \rho y_{t-1} = \underset{\sim}{x}_t^T \underset{\sim}{\beta} - \underset{\sim}{x}_{t-1}^T(\rho\underset{\sim}{\beta}) + \varepsilon_t - \rho\varepsilon_{t-1} \Rightarrow y_t = \rho y_{t-1} + (\underset{\sim}{x}_t - \rho \underset{\sim}{x}_{t-1})^T \underset{\sim}{\beta} + u_t$, $u_t \overset{iid}{\sim} N(0, \sigma^2)$

$\tilde{\beta}_{K+1} = \arg\min_{\rho, \underset{\sim}{\beta}} \sum_{t=2}^{n}\left(y_t - \rho y_{t-1} - (\underset{\sim}{x}_t - \rho\underset{\sim}{x}_{t-1})^T\underset{\sim}{\beta}\right)^2$, $\hat{\beta}_{OLS} = \sum(y_t - \underset{\sim}{x}_t^T\underset{\sim}{\beta})^2$

Endogeneity:

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{Cov(x_i, \varepsilon_i)}{Var(x_i)}$

$E(\varepsilon_i|x_i) = 0 \Rightarrow Cov(x_i, \varepsilon_i) = 0$ [x_i is exogenous], If $E(\varepsilon_i|x_i) \neq 0 \Rightarrow x_i$ is endogenous.

Endogeneity:

06/09/2023

Def: Endogeneity in a multiple linear regression is said to occur if at least one of the regressors or covariate is correlated with error ε_i i.e. $Cov(X_{ji}, \varepsilon_i) \neq 0$ for some $j = 1(1)k$

① $E(\varepsilon_i \mid \underset{\sim}{X}_i) = 0 \Rightarrow Cov(\varepsilon_i, X_{ji}) = 0 \quad \forall j = 1(1)k$

$\boxed{Eg.}$ Consider a simple linear regression:

$y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \quad , \quad i = 1(1)n$ [Suppose that, $Cov(X_{1i}, \varepsilon_i) \neq 0$]

eg. $exp_i = \beta_0 + \beta_1 I_i + \varepsilon_i$ where, exp_i : Expenditure of i^{th} individual in a data

I_i : Income of i^{th} individual in a data

$$\hat{\beta}_{1,OLS} = \frac{\sum (X_{1i} - \bar{X})(Y_i - \bar{Y})}{\sum (X_{1i} - \bar{X})^2} = \underbrace{\beta_1}_{\text{Constant}} + \frac{\sum (X_{1i} - \bar{X})(\varepsilon_i - \bar{\varepsilon})}{\sum (X_{1i} - \bar{X})^2} \leftarrow \text{Ratio of sample moments}$$

As $n \to \infty$, $\hat{\beta}_{1,OLS} \longrightarrow \beta_1 + \dfrac{Cov(X_1, \varepsilon)}{Var(X_1)} \leftarrow$ Ratio of population moments $\equiv \beta_1 + \text{bias}$

Hence, $\hat{\beta}_{1,OLS}$ is biased and inconsistent. [ie. $E(\hat{\beta}_{1,OLS}) = \beta_1 + (\cdot)$]

Source of Endogeneity: There are mainly three sources.

① Omitted variables or covariates.
② Measurement error in covariates.
③ Simultaneity.

① Omitted Variables or Covariates:

Consider a very simple case: $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$: True Model (All the GM assumptions are true)

$\Rightarrow y_i = \beta_0 + \beta X_{1i} + \varepsilon_i^*$ where $\varepsilon_i^* = \beta_2 X_{2i} + \varepsilon_i$ [this is because X_{2i} may be unobserved.]

Note that,

$$Cov(X_{1i}, \varepsilon_i^*) = Cov(X_{1i}, \beta_2 X_{2i} + \varepsilon_i) = Cov(X_{1i}, \beta X_{2i}) + Cov(X_{1i}, \varepsilon_i) = \beta_2 \, Cov(X_{1i}, X_{2i})$$

[$\because Cov(X_{1i}, \varepsilon_i) = 0$, $\because Cov(X_{2i}, \varepsilon_i) = 0$]

If $Cov(X_{1i}, X_{2i}) \neq 0$ ie. X_{1i} and X_{2i} are correlated to some extent) then,

$Cov(X_{1i}, \varepsilon_i) = \beta_2 \, Cov(X_{1i}, X_{2i}) \neq 0 \Rightarrow \hat{\beta}_{1,OLS}$ will be bias and inconsistent.

② Measurement Error in a Covariate:

True Model: $y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$

Measurement error in X_{1i} means, X_{1i}^* (Observed) $= X_{1i}$ (Actual) $+ v_i$ (measurement error) $\Rightarrow X_{1i} = (X_{1i}^* - v_i)$

$y_i = \beta_0 + \beta_1 (X_{1i}^* - v_i) + \varepsilon_i = \beta_0 + \beta_1 X_{1i}^* + (\varepsilon_i - \beta_1 v_i) = \beta_0 + \beta_1 X_{1i}^* + \omega_i$: Working Model

Note that,

$$Cov(X_{1i}^*, \omega_i) = Cov(X_{1i} + v_i, \varepsilon_i - \beta_1 v_i) = Cov(X_{1i}, \varepsilon_i) - \beta_1 Cov(X_{1i}, v_i) + Cov(v_i, \varepsilon_i) - \beta_1 Var(v_i)$$
$$= -\beta_1 \sigma_v^2 \neq 0$$

$\boxed{H.W.}$ Show that if the measurement error occurs only with y_i then $\hat{\beta}_{OLS}$ will be still unbiased and inconsistent.

③ Simultaneity (Both way Causality):

$Y_i \rightleftarrows X_i$

Effecting each other

Consider a very simple example: $y_i = \beta_0 + \beta_1 X_{1i} + u_i$ ——— (i)

$X_{1i} = \alpha_0 + \alpha_1 y_i + v_i$ ——— (ii)

From eq (ii), eq (i) can be written as,

$$y_i = \beta_0 + \beta_1(\alpha_0 + \alpha_1 y_i + v_i) + u_i \Rightarrow (1 - \alpha_1\beta_1) y_i = (\beta_0 + \alpha_0\beta_1) + (u_i + \beta_1 v_i) \Rightarrow y_i = \left(\frac{\beta_0 + \alpha_0\beta_1}{1 - \alpha_1\beta_1}\right) + \left(\frac{u_i + \beta_1 v_i}{1 - \alpha_1\beta_1}\right) \text{ ——— (iii)}$$

Similarly,

$$X_{1i} = \alpha_0 + \alpha_1(\beta_0 + \beta_1 X_{1i} + u_i) + v_i \Rightarrow X_{1i} = \left(\frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1}\right) + \left(\frac{\alpha_1 u_i + v_i}{1 - \alpha_1\beta_1}\right) \text{ ——— (iv)}$$

Eq. iii & iv are called the reduced form of the original models i & ii

$\therefore \mathrm{Cov}(X_{1i}, u_i) = \mathrm{Cov}\left(\frac{\alpha_0+\alpha_1\beta_0}{1-\alpha_1\beta_1} + \frac{\alpha_1 u_i + v_i}{1-\alpha_1\beta_1}, u_i\right) = \mathrm{Cov}\left(\frac{\alpha_1}{1-\alpha_1\beta_1} u_i, u_i\right) = \frac{\alpha_1}{1-\alpha_1\beta_1}\sigma_u^2$

Similarly, $\mathrm{Cov}(Y_i, v_i) = \frac{\beta_1}{1-\alpha_1\beta_1}\sigma_v^2$