ECO3121

Introductory Econometrics

许炜源

Quiz: Quiz 3 - 4
Quiz 6 - 3
Quiz 10 - 4

21-3 Introduction and Statistics Review

1. Basics of econometrics

① Definition

Analysis of economic data with statistical methods

② Steps of doing econometric research

计量→实证分析

- **Steps in an empirical analysis**
  - Step 1: Carefully pose a question.
  - Step 2: Specify an economic or conceptual model.
  - Step 3: Turn the economic model into an econometric model.
  - Step 4: Collect data on the variables and use statistical methods to estimate the parameters, construct confidence intervals for the parameters, and test hypotheses.

③ Economic data sets

- Cross-sectional data 横向数据 represent a random sample
- Time series data 时间序列数据
- Pooled cross sections 混合横向数据 cross sections are drawn independently
- Panel/Longitudinal data 纵向数据

Cross Sectional Data

2003

Time Series Data

1993 2003 2013

Pooled Cross Sectional Data

1993 2003 2013

Panel Data

④ Causality & Ceteris paribus

Relationship between x & y ⊗⇒ Causality

Relationship between x & y + cp ✓⇒ Causality

Control other relavant factors the same ⇒ Make x independent with other relavant factors ⇒ Random sampling

(ΔU = 0) (Zero conditional mean)

2. Statistics review

① Variance

$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$

Recall: Risk $= \sqrt{Var(w_1X_1+\cdots+w_nX_n)} = Std(w_1X_1+\cdots+w_nX_n)$

② Covariance

a. Sample covariance

$$\text{Sample } Cov(X,Y) = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{n-1} = \frac{1}{n-1}\left[\sum X_iY_i - \frac{\sum X_i \sum Y_i}{n}\right]$$

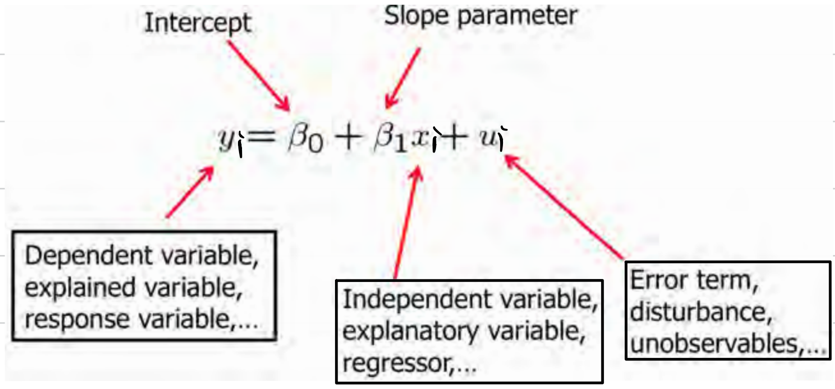
b. Two properties

$Cov(aX,bY) = ab\,Cov(X,Y)$

$Cov(aX+bY, cW+dZ) = ac\,Cov(X,W) + ad\,Cov(X,Z) + bc\,Cov(Y,W) + bd\,Cov(Y,Z)$

L4-6 Simple Linear Regression Model (≙ Bivariate Linear Regression Model)

1. Actual case (散点图) ⟵ 同个 x_i 有多个 y_i



2. Average: Population regression function (PRF)

① Zero conditional mean assumption

u & y: $E(u)=0$ (by defining β_0) 可通过调整来达到

u & x: $E(u|x)=E(u)$ (why not just '$\rho \neq 0$'? : $\rho=0$ 只能说明线性不相关 (u & x'))

$\forall x_i$, $E(u_i|x_i)=0$ ⇒ $E(u|x)=0$ 在实际社科中很难实现

↓

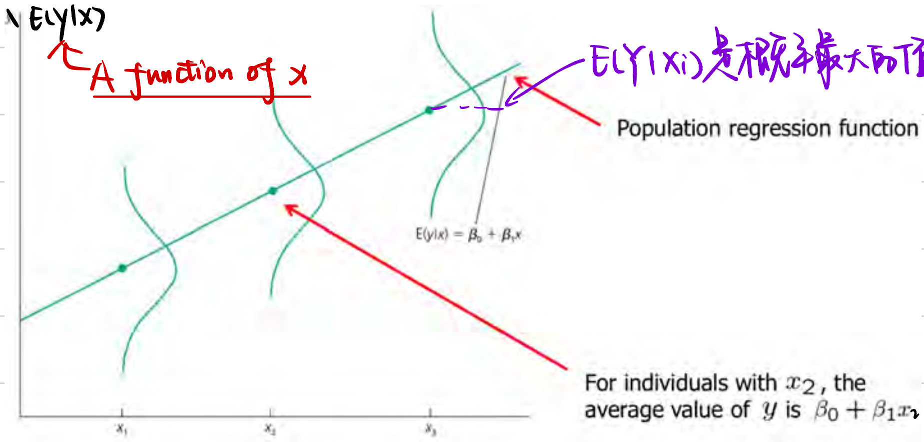
导致：我们很难用 SLR 解释 Causality relationship

② PRF

$E(\beta_1 x|x)=\beta_1 x$

$E(y|x)=E(\beta_0+\beta_1 x+u|x)=\beta_0+\beta_1 x+E(u|x)=\beta_0+\beta_1 x$ ← $y|x \sim N(\beta_0+\beta_1 x, \sigma^2)$

Significance: The average value of y given x can be expressed as a linear function of x.



3. Prediction: Sample regression function / Ordinary Least Squares regression function (OLS regression function)

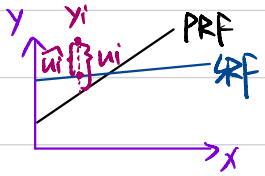
① Overview

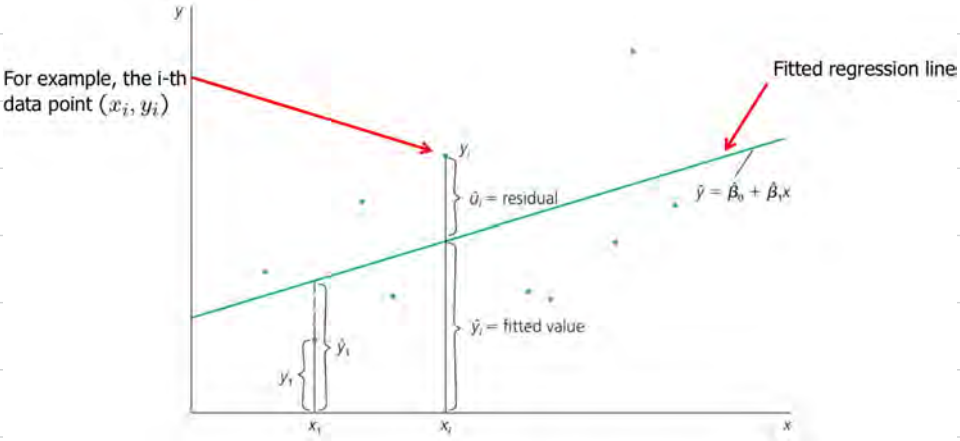
$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ ← fitted value

The mistake is called "residual": $\hat{u}_i = y_i - \hat{y}_i$

$u_i = y_i - E(y|x_i)$

$\hat{u}_i$ (Residual) $= y_i - \hat{y}_i$





解释 $\hat{\beta}_1$: Holding other factors fixed, ...

② OLS (Point estimation)

a. Process

$$\min \sum_{i=1}^{n} \hat{u}_i^2 \rightarrow \min \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \rightarrow \text{FOC}: \begin{cases} \hat{\beta}_0: \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \hat{\beta}_1: \sum x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \end{cases}$$

$$\rightarrow \begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \dfrac{\sum x_i y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2} = Corr(x,y)\dfrac{\sigma_y}{\sigma_x}, \text{ where } \begin{cases} \sigma_x = \sqrt{\sum (x_i - \bar{x})^2} \\ \sigma_y = \sqrt{\sum (y_i - \bar{y})^2} \end{cases} \\ \hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-2} \end{cases}$$

$n \rightarrow n-2$: Biased → Unbiased

易错一个方法 (MM, 课本 P49)

$$E(u) = 0$$

$$Cov(x,u) = E(xu) = 0$$

b. Basic properties

i) $E(\hat{u}_i) = 0$ $\sum_{i=1}^{n} \hat{u}_i = 0$

ii) $E(\hat{u}_i | x_i) = 0$ $\sum_{i=1}^{n} x_i \hat{u}_i = 0$

iii) $(\bar{x}, \bar{y})$ 在 SRF 上 $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

FOC of $\hat{\beta}_0$ $\rightarrow \sum_{i=1}^{n} (y_i - \hat{y}_i) = 0 \rightarrow \sum_{i=1}^{n} \hat{u}_i = 0$

FOC of $\hat{\beta}_1$ $\rightarrow \sum_{i=1}^{n} x_i (y_i - \hat{y}_i) = 0 \rightarrow \sum_{i=1}^{n} x_i \hat{u}_i = 0$

$E(f(x)\hat{u}) = 0$

下面所有都 $\times x_i$

Implications:

i) $\bar{y} = \bar{\hat{y}}$ $y_i = \hat{y}_i + \hat{u}_i \rightarrow \frac{1}{n}\sum y_i = \frac{1}{n}\sum \hat{y}_i + \frac{1}{n}\sum \hat{u}_i \rightarrow \bar{y} = \bar{\hat{y}}$

ii) $\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$ $\sum x_i \hat{u}_i = 0 \rightarrow \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i)\hat{u}_i = \sum \hat{y}_i \hat{u}_i = 0$

区分 $\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$

$\sum_{i=1}^{n} y_i \hat{u}_i \neq 0$

$\sum_{i=1}^{n} y_i \hat{u}_i = \sum_{i=1}^{n} (\hat{y}_i + \hat{u}_i)\hat{u}_i = \sum_{i=1}^{n} \hat{u}_i^2$

c. Statistical properties

5 Assumptions (Gauss-Markov assumptions)

a. Linear relation : $y_i = \beta_0 + \beta_1 x_i + u_i$

b. Random sampling

c. Existence of sample variation : $\sum (x_i - \bar{x})^2 > 0 \Rightarrow \hat{\beta}_1$ 才可求: $\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

d. Zero conditional mean : $\forall i, E(u_i | x_i) = 0$

e. Constant variance: $Var(u_i | x_i) = \sigma^2$

(Homoskedasticity) $\longleftrightarrow$ (Heteroskedasticity)

求Exp时

证明 Unbiasedness 时

求Var时.

① $E(\hat{\beta_1})$

$E(\hat{\beta_1}) = \beta_1$ (Unbiased)

proof:

$$E(\hat{\beta_1}) = E\left(\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\right) = E\left(\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}\right) = E\left(\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}\right) = E\left(\frac{\sum(x_i-\bar{x})(\beta_0+\beta_1x_i+u_i)}{\sum(x_i-\bar{x})^2}\right) = E\left(\frac{\beta_0\sum(x_i-\bar{x})+\beta_1\sum x_i(x_i-\bar{x})+\sum u_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)$$

$$= E\left(\frac{0+\beta_1\sum x_i(x_i-\bar{x})+\sum u_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right) = E\left(\beta_1\frac{\sum x_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right) + E\left(\frac{\sum u_i(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right) = E\left(\beta_1\frac{\sum x_i(x_i-\bar{x})}{\sum(x_i-\bar{x})(x_i-\bar{x})}\right) + E\left(\frac{\sum u_ix_i-\sum u_i\bar{x}}{\sum(x_i-\bar{x})^2}\right)$$

$$= E\left(\beta_1\frac{\sum x_i(x_i-\bar{x})}{\sum x_i(x_i-\bar{x})-\sum\bar{x}(x_i-\bar{x})}\right) + E\left(\frac{\sum u_ix_i}{\sum(x_i-\bar{x})^2}\right) = \beta_1 + 0 = \beta_1$$

② $E(\hat{\beta_0})$

$E(\hat{\beta_0}) = \beta_0$ (Unbiased)

proof:

$$E(\hat{\beta_0}) = E(\bar{y}-\hat{\beta_1}\bar{x}) = y-\beta_1x = \beta_0$$

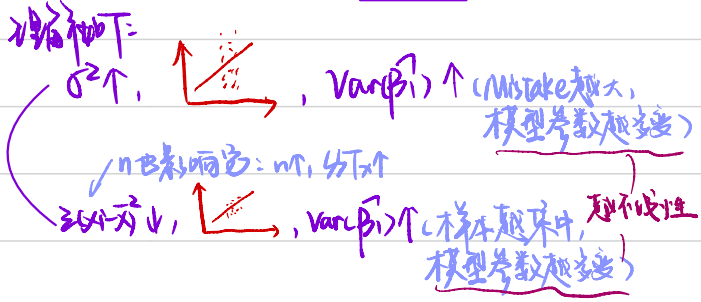
③ $Var(\hat{\beta_1})$

$$Var(\hat{\beta_1}) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$$

Logic: 在"Conditioned on x_i's"的前提下算出 E、Var，
后发现与 x_i 是 x_1 还是 … x_n 无关，所以去除前提也成立。
"Conditioned on x_i's" ⇒ x_i's are constant

proof:

$$Var(\hat{\beta_1}) = Var\left(\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\right) = Var\left(\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}\right) = \frac{\sum(x_i-\bar{x})^2Var(y_i)}{\left(\sum(x_i-\bar{x})^2\right)^2} = \frac{1}{\sum(x_i-\bar{x})^2}Var(y_i) = \boxed{\frac{\sigma^2}{\sum(x_i-\bar{x})^2}}$$



④ $Var(\hat{\beta_0})$

$$Var(\hat{\beta_0}) = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)$$

proof:

$$Var(\hat{\beta_0}) = Var(\bar{y}-\hat{\beta_1}\bar{x}) = Var(\bar{y}) + \bar{x}^2\frac{\sigma^2}{\sum(x_i-\bar{x})^2} - 2Cov(\bar{y},\hat{\beta_1}\bar{x})$$

$$= Var\left(\frac{\sum y_i}{n}\right) + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\sigma^2 - 2\bar{x}\,Cov\left(\frac{\sum y_i}{n}, \frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}\right) = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right) + \frac{2\bar{x}}{n\sum(x_i-\bar{x})^2}Cov\left(\sum y_i, \sum(x_i-\bar{x})y_i\right)$$

$$= \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right) + \frac{2\bar{x}}{n\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})Cov(y_i,y_i) = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right) + \frac{2\bar{x}}{n\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})\sigma_i = \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)$$

⑤ $E(\hat{\sigma}^2) = \sigma^2$ (当 n 变成 n-2 时，就是 unbiased)

Standard error of the regression: $\hat{\sigma}$ (一次) ⟶ Standard error of $\hat{\beta}$

$sd(\hat{\beta}) \Rightarrow se(\hat{\beta})$

$E(\hat{\sigma}^2) = \sigma^2$ ✓ Unbiased

$E(\hat{\sigma}) = \sigma$ ✗ Biased

Standard error
Standard deviation
→ Estimator

③ Goodness of fit

a. Variations

$$y_i = \hat{y}_i + \hat{u}_i$$

$$\longrightarrow y_i - \bar{y} = \hat{y}_i - \bar{y} + \hat{u}_i$$

$$\longrightarrow \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y}+\hat{u}_i)^2$$

$$= \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(\hat{u}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i-\bar{y})\hat{u}_i$$

$$= \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(\hat{u}_i)^2 + 2\sum_{i=1}^{n}\hat{y}_i\hat{u}_i - 2\sum_{i=1}^{n}\bar{y}\hat{u}_i$$

$$= \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(\hat{u}_i)^2$$

SST (y_i , $\hat{y}_i$, $\bar{y}$): y_i to $\hat{y}_i$ = SSR; $\hat{y}_i$ to $\bar{y}$ = SSE — 接近 y_i 的程度

$\longrightarrow SST = SSE + SSR$, where

$SST = \sum_{i=1}^{n}(y_i-\bar{y})^2$	Total variation	y
$SSE = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	Explained part	x
$SSR = \sum_{i=1}^{n}(\hat{u}_i)^2$	Unexpected part	u

总体误差 SST:

- 被预测的 = SSE — Expected — 来自 x_i
- 未被预测的 = SSR — Error — 来自 $\hat{u}_i$

b. R-square

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$R^2 \in [0,1]$, $R^2\uparrow$, Goodness of fit $\uparrow$

(注意，仅适用于相同model下)

- R^2小，说明线性关系弱
- R^2小，不能说明线性模型差 (判断线性模型好坏的依据是是否满足5个assumptions)

g. Alternative expression of R-squared

$$R^2 = \frac{SSE}{SST}$$

$$= \frac{\Sigma(\hat{y}_i-\bar{y})^2}{\Sigma(y_i-\bar{y})^2}$$

$$= \frac{[\Sigma(\hat{y}_i-\bar{y})^2]^2}{\Sigma(y_i-\bar{y})^2\cdot\Sigma(\hat{y}_i-\bar{y})^2}$$

$$\Sigma(\hat{y}_i-\bar{y})^2$$

$$= \Sigma(y_i+\hat{u}_i-\bar{y})(\hat{y}_i-\bar{y})$$

$$= \Sigma(y_i+\hat{u}_i)(\hat{y}_i-\bar{y}) - \Sigma\bar{y}(\hat{y}_i-\bar{y})$$

$$= \Sigma(y_i-\bar{y})(\hat{y}_i-\bar{y}) + \underbrace{\Sigma\hat{u}_i(\hat{y}_i-\bar{y})}_{=0}$$

$$\therefore R^2 = \frac{[\Sigma(y_i-\bar{y})(\hat{y}_i-\bar{y})]^2}{\Sigma(y_i-\bar{y})^2\cdot\Sigma(\hat{y}_i-\bar{y})^2}$$

$$= corr(y_i, \hat{y}_i)^2$$

④ Change of units

a. $\Delta x \longrightarrow \%\Delta x$ (Change in x(dec))

$\Delta x = \frac{\%\Delta x}{100}$ ↑ percentage change

- Intercept 不变
- Slope 变
- R^2 不变

b. $\%\Delta x \longrightarrow \Delta \ln(x)$

$\Delta \ln(x) \approx \frac{\%\Delta x}{100}$

- Intercept 变
- Slope 变
- R^2 变

proof: Since $x \to 0$, $\ln(x+1) \to x$, then $\Delta \ln(x) = \ln(x_0 + \Delta x) - \ln(x_0) = \ln\left(1 + \frac{\Delta x}{x_0}\right) \approx \frac{\Delta x}{x_0} = \frac{\%\Delta x}{100}$

c. Summary of forms of variables (因为都是"y的 = $\beta_0 + \beta_1 x$的"的形式，故都是以R为)

Model	Dep. Var.	Indep. Var.	β_1
Level-Level	y	x	$\Delta y = \beta_1 \Delta x$
Level-Log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-Level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-Log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

← "Constant elasticity model"

4. Regression through the origin ($\beta_0 = 0$)

① Regression model (Assumption)

$$y = \beta_0 + \beta_1 x + u$$

② Sample regression function

a. Overview

$$\hat{y} = \tilde{\beta}_1 x$$

b. Point estimate

$$\tilde{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

c. Statistical properties

△ $E(\tilde{\beta}_1)$

$E(\tilde{\beta}_1) = \beta_0 \frac{\sum x_i}{\sum x_i^2} + \beta_1$ (biased)

proof:

$$E(\tilde{\beta}_1) = E\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = E\left(\frac{\sum x_i(\beta_0 + \beta_1 x_i + u_i)}{\sum x_i^2}\right) = E\left(\frac{\sum x_i \beta_0}{\sum x_i^2}\right) + E\left(\frac{\beta_1 \sum x_i^2}{\sum x_i^2}\right) + E\left(\frac{\sum x_i u_i}{\sum x_i^2}\right) = \beta_0\left(\frac{\sum x_i}{\sum x_i^2}\right) + \beta_1 + 0$$

$$= \beta_0 \frac{\sum x_i}{\sum x_i^2} + \beta_1 \neq \beta_1$$

△ $Var(\tilde{\beta}_1)$

$Var(\tilde{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$

proof:

$$Var(\tilde{\beta}_1) = Var\left(\frac{\sum x_i y_i}{\sum x_i^2}\right) = \frac{\sum x_i^2}{(\sum x_i^2)^2}\sigma^2 = \frac{\sigma^2}{\sum x_i^2}$$

d. Goodness of fit

$$R^2 = 1 - \frac{\sum (y_i - \tilde{\beta}_1 x_i)^2}{\sum (y_i - \bar{y})^2}$$

$\bar{y} = 0$, and when y_i & $\tilde{y}_i$ have different signs, then $R^2 < 0$.

At this time, it means that regression effectiveness: $\hat{y} > \bar{y} > \tilde{y}$

拟合 y_i 的程度

L7-10 Multiple Linear Regression Model — Estimation & Properties

1. Motivation

① Ceteris paribus

SLR 很难证明 x 与 u 不相关 / CP 成立，i.e., E(u|x) = E(u)，

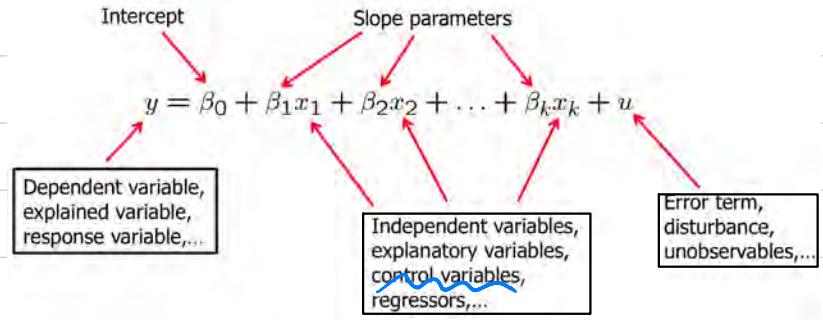
因此，我们引进更多变量，从而使得 x_i 与 u 不相关 / CP 成立，进行研究 Causality $x_1 \to x_1, x_2$ 注意观察 β_1 的变化

② Function form

MLR 使得函数形式更为丰富，e.g. $\begin{cases} x = x \\ y = x^2 \end{cases}$ （单元 / 多元是取决于 β_i 前面的指数）

线性指的是 β_i

2. Actual case



3. Prediction = Sample regression function

① Overview

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ → For i $\begin{cases} x_{i1} \\ x_{i2} \end{cases}$, $\hat{u}_i = y_i - \hat{y}_i$

（当然，不止于此形式）

② OLS (Point estimation)

a. Process

$$\min \sum_{i=1}^{n} \hat{u}_i^2 \to \min \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2 \to \text{FOC:} \begin{cases} \hat{\beta}_0: \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) = 0 \\ \hat{\beta}_1: \sum x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) = 0 \\ \hat{\beta}_2: \sum x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}) = 0 \end{cases}$$

$\hat{\beta}_j$ 下代表的是 Partial effect

(Holding other independent variables fixed)

We don't need to compute the exact form of $\hat{\beta}_j$, but we need to master "Partialling out" Interpretation form of $\hat{\beta}_j$

困于高阶矩，不作计算

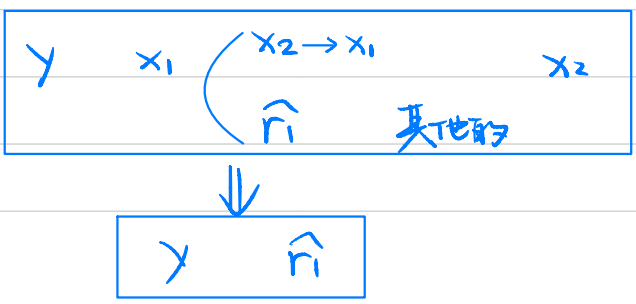
① Reg x_{i1} on x_{i2}, find residual $\hat{r}_{i1}$

② Reg y_i on $\hat{r}_{i1}$ → find $\hat{\beta}_1$

$$\to \hat{\beta}_1 = \frac{\sum (\hat{r}_{i1} - \bar{\hat{r}}_{i1})(y_i - \bar{y}_i)}{\sum (\hat{r}_{i1} - \bar{\hat{r}}_{i1})^2} = \frac{\sum (\hat{r}_{i1} y_i)}{\sum \hat{r}_{i1}^2}$$

$$\hat{\beta}_j = \frac{\sum (\hat{r}_{ij} y_i)}{\sum \hat{r}_{ij}^2} \quad \left(= \frac{\sum (\hat{r}_{i1} - 0)(y_i - \bar{y})}{\sum (\hat{r}_{i1} - 0)^2} \right)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - (k+1)}$$ （因为我们之前估计了 (k+1) 个参数）



b. Basic properties

i) $E(\hat{u}_i)=0$ $\sum_{i=1}^{n}\hat{u}_i=0$

ii) $E(\hat{u}_i|x_i)=0$ $\sum_{i=1}^{n}x_{ij}\hat{u}_i=0$

iii) $(\bar{x}_1,\bar{x}_2,\cdots,\bar{y})$ 满足 $\bar{y}=\hat{\beta}_0+\hat{\beta}_1\bar{x}_1+\cdots+\hat{\beta}_k\bar{x}_k$

Implications

i) $\bar{y}=\bar{\hat{y}}$ $y_i=\hat{y}_i+\hat{u}_i \rightarrow \frac{1}{n}\sum y_i=\frac{1}{n}\sum\hat{y}_i+\frac{1}{n}\sum\hat{u}_i \rightarrow \bar{y}=\bar{\hat{y}}$

$\Rightarrow$ ii) $\sum_{i=1}^{n}\hat{y}_i\hat{u}_i=0$ $\sum_{i=1}^{n}(\hat{\beta}_0+\hat{\beta}_1x_{i1}+\cdots+\hat{\beta}_kx_{ik})\hat{u}_i=0+0+\cdots+0=0$

c. Statistical properties

5 Assumptions (Gauss-Markov assumptions)

a. Linear relation : $y_i=\beta_0+\beta_1x_{i1}+\cdots+\beta_kx_{ik}+u_i$

b. Random sampling

c. No perfect collinearity 完美共线性 :

① Existence of sample variation : For $\forall j$, $\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2>0$

② No exact linear relationship between independent variables (e.g. $x_1=2x_2+\frac{1}{3}x_3$ ×; $x_1=2x_2^2+\frac{1}{3}x_3$ √)

d. Zero conditional mean : $\forall i, E(u_i|x_{i1},\cdots,x_{ik})=0$ 相较于SLR更加宽松

(Exogenity assumption)

Endogenous 内生 : x_j is correlated with u_i

Exogenous 外生 : x_j is uncorrelated with u_i

e. Constant variance : $\forall i, Var(u_i|x_{i1},\cdots,x_{ik})=\sigma^2$

(Homoskedasticity)

① $E(\hat{\beta}_j)$ (Normal case)

$E(\hat{\beta}_j) = \beta_j$

Proof:

$$E(\hat{\beta}_j) = E\left(\frac{\sum_{i=1}^n (\hat{r}_{ij} y_i)}{\sum_{i=1}^n \hat{r}_{ij}^2}\right) = E\left(\frac{\sum_{i=1}^n \hat{r}_{ij}(\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}+u_i)}{\sum_{i=1}^n \hat{r}_{ij}^2}\right) = E\left(\frac{0+0+\cdots+\sum_{i=1}^n \hat{r}_{ij}\beta_j x_{ij}+\cdots+0+\sum_{i=1}^n \hat{r}_{ij}u_i}{\sum_{i=1}^n \hat{r}_{ij}^2}\right)$$

($\hat{r}_{ij}$ is uncorrelated with ···)

$$= E\left(\frac{\sum_{i=1}^n \hat{r}_{ij}\beta_j x_{ij} + \sum_{i=1}^n \hat{r}_{ij}u_i}{\sum_{i=1}^n \hat{r}_{ij}^2}\right) = E\left(\frac{\sum_{i=1}^n \hat{r}_{ij} x_{ij}}{\sum_{i=1}^n \hat{r}_{ij}^2}\beta_j\right) + E\left(\frac{\sum_{i=1}^n \hat{r}_{ij}u_i}{\sum_{i=1}^n \hat{r}_{ij}^2}\right) = E\left(\frac{\sum_{i=1}^n \hat{r}_{ij}(x_{ij}+\hat{r}_{ij})\beta_j}{\sum_{i=1}^n \hat{r}_{ij}^2}\right) + 0$$

(把 x_{ij} 拿出去了)

(A Function of x_i)

$$= E\left(\frac{\sum_{i=1}^n \hat{r}_{ij}^2}{\sum_{i=1}^n \hat{r}_{ij}^2}\beta_j\right) + E\left(\frac{\sum_{i=1}^n \hat{r}_{ij}x_{ij}}{\sum_{i=1}^n \hat{r}_{ij}^2}\beta_j\right) = \beta_j + 0 = \beta_j$$

($\sum \hat{r}_{ij}x_{ij} = 0$)

② $E(\hat{\beta}_j)$ (Overspecifying case) 即：包含了无关变量 x_j

$E(\hat{\beta}_j) = 0$

proof:

$E(\hat{\beta}_j) = \beta_j = 0$ 注意，($\beta_j = 0$; $\hat{\beta}_j \neq 0$ (不一定))

③ $E(\hat{\beta}_j)$ (Underspecifying/Omitting cases)

Exp1: Relevant variable X_2 is omitted

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$

$\hat{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{i1}$

$E(\tilde{\beta}_1) = \beta_1 + \beta_2\tilde{\delta}_1$, 关系不上, $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$

where $\tilde{\delta}_1$ is the estimated slope in SLR where reg x_2 on x_1

$E(\tilde{\beta}_1)$ 来自 β_1

来自 β_2 (Bias) → Type of Bias: Upward / Positive; Downward / Negative

proof:

$$E(\tilde{\beta}_1) = E\left(\frac{\sum_{i=1}^n (x_{i1}-\bar{x}_1)(y_i-\bar{y}_i)}{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}\right) = E\left(\frac{\sum_{i=1}^n (x_{i1}-\bar{x}_1)(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+u_i)}{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}\right)$$

$$= E\left(\beta_1 + \frac{\beta_2\sum_{i=1}^n (x_{i1}-\bar{x}_1)x_{i2}}{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}\right) = \beta_1 + \beta_2\frac{\sum_{i=1}^n (x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}$$

($\frac{\sum_{i=1}^n (x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2} = \tilde{\delta}_1$)

$= \beta_1 + \beta_2\tilde{\delta}_1$

Bias 的正负：β_2 x_2 与 y 的关系；$\tilde{\delta}_1$ x_2 与 x_1 的关系

Exp2: Relevant variable X_3 is omitted

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$

$\hat{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2}$

Similarly, we get

$E(\tilde{\beta}_1) = \beta_1 + \beta_3\tilde{\delta}_1$,

where $\tilde{\delta}_1 = \frac{\sum_{i=1}^n (x_{i1}-\bar{x}_1)(x_{i3}-\bar{x}_3)}{\sum_{i=1}^n (x_{i1}-\bar{x}_1)^2}$

Proof:

$$E(\tilde{\beta}_1) = E\left(\frac{\sum_{i=1}^n (\hat{r}_{i1} y_i)}{\sum_{i=1}^n (\hat{r}_{i1})^2}\right) = E\left(\frac{\sum_{i=1}^n \hat{r}_{i1}(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 x_{i3}+u_i)}{\sum_{i=1}^n \hat{r}_{i1}^2}\right),$$

$$= \beta_0\frac{\sum_{i=1}^n \hat{r}_{i1}}{\sum_{i=1}^n \hat{r}_{i1}^2} + E\left(\frac{\sum_{i=1}^n \hat{r}_{i1}x_{i1}\beta_1}{\sum_{i=1}^n \hat{r}_{i1}^2}\right) + E\left(\frac{\sum_{i=1}^n \hat{r}_{i1}\beta_3 x_{i3}}{\sum_{i=1}^n \hat{r}_{i1}^2}\right)$$

($\sum \hat{r}_{i1} = 0$)

$$= \beta_0\frac{\sum_{i=1}^n \hat{r}_{i1}}{\sum_{i=1}^n \hat{r}_{i1}^2} + E\left(\frac{\sum_{i=1}^n \hat{r}_{i1}(\hat{x}_{i1}+\hat{r}_{i1})\beta_1}{\sum_{i=1}^n \hat{r}_{i1}^2}\right) + \beta_3\frac{\sum_{i=1}^n \hat{r}_{i1}x_{i3}}{\sum_{i=1}^n \hat{r}_{i1}^2}$$

$$= \beta_1 + \beta_3\frac{\sum_{i=1}^n \hat{r}_{i1}x_{i3}}{\sum_{i=1}^n \hat{r}_{i1}^2}$$

④ Var($\hat{\beta}_j$) (Normal case)

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2(1-R_j^2)}$$

此时称SST_j　R_j^2 即 reg x_j on other independent variables for Goodness of fit

proof:

$$\text{Var}(\hat{\beta}_j) = \text{Var}\left(\frac{\sum_{i=1}^{n}(\hat{r}_{ij}\,y_i)}{\sum_{i=1}^{n}\hat{r}_{ij}^2}\right) = \text{Var}\left(\frac{\sum_{i=1}^{n}\hat{r}_{ij}(\beta_0+\beta_1 x_{i1}+\cdots+\beta_n x_{in}+u_i)}{\sum_{i=1}^{n}\hat{r}_{ij}^2}\right) = \text{Var}\left(\frac{\sum_{i=1}^{n}\hat{r}_{ij}\beta_j x_{ij}+\sum_{i=1}^{n}\hat{r}_{ij}u_i}{\sum_{i=1}^{n}\hat{r}_{ij}^2}\right)$$

$$= \text{Var}\left(\frac{\sum_{i=1}^{n}\hat{r}_{ij}\beta_j(\hat{x}_{ij}+\hat{r}_{ij})+\sum_{i=1}^{n}\hat{r}_{ij}u_i}{\sum_{i=1}^{n}\hat{r}_{ij}^2}\right) = \text{Var}\left(\beta_j+\frac{\sum_{i=1}^{n}\hat{r}_{ij}u_i}{\sum_{i=1}^{n}\hat{r}_{ij}^2}\right)$$

$$= \text{Var}\left(\frac{\sum_{i=1}^{n}\hat{r}_{ij}u_i}{\sum_{i=1}^{n}\hat{r}_{ij}^2}\right) = \frac{\sum_{i=1}^{n}(\hat{r}_{ij})^2}{\left(\sum_{i=1}^{n}(\hat{r}_{ij})^2\right)^2}\text{Var}(u_i) = \frac{1}{\sum_{i=1}^{n}(\hat{r}_{ij})^2}\sigma^2 = \frac{\sigma^2}{SSR_j}$$

Since $R_j^2 = 1-\frac{SSR_j}{SST_j}$, then $SSR_j = (1-R_j^2)SST_j$

then $\frac{\sigma^2}{SSR_j} = \frac{\sigma^2}{SST_j(1-R_j^2)} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2(1-R_j^2)}$

$SST_j = \sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2 \approx n\sigma_{x_j}^2$

↳ 是因为sample里面n没有很多

(i) Multicollinearity: $R_j^2 \to 1$ for some j

(ii) $R_j^2 \to 1 \Rightarrow \text{Var}(\hat{\beta}_j) \to \infty$ 即：多重共线性越强，Ass 3越快违背，MLR以此时模型越差，参数的variance越大

(iii) $\text{se}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2(1-R_j^2)}}$ 注意：此时是se，故用$\hat{\sigma}$代替σ，$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-(k+1)}$

T检验，如果$R^2 \neq 1$，这时就不算Violate Ass 3

④ $Var(\hat{\beta}_j)$ (Special cases)

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ 2

$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$ 1

$Var(\tilde{\beta}_1) = \frac{\sigma^2}{\sum_i (x_{i1} - \bar{x}_1)^2}$ (无论overspecifying / underspecifying, 我们都有: $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$)

总结如下:

当 $\beta_2 = 0$ (说明MLR overspecifying) →
- $E(\hat{\beta}_1) = \beta_1$
- $E(\tilde{\beta}_1) = \beta_1$
- $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$ (当 $\beta_2 \neq 0$ 时)

当 $\beta_2 \neq 0$ (说明MLR underspecifying) →
- $E(\hat{\beta}_1) = \beta_1$
- $E(\tilde{\beta}_1) \neq \beta_1$
- $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$ (当 $\beta_2 \neq 0$ 时)

→ Trade off between bias & variance
- Bias 上升
- Variance 下降

⇒ 由于 (Bias不可变, Variance可通过找高n来减小), 所以当underspecifying时, 我们可选择 MLR

③ Goodness of fit

R^2 可以较大小的前提是相同 Sample

$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$ (Definition)

$= Corr(y_i, \hat{y}_i)^2 = \left(\frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})\sum_{i=1}^{n}(\hat{y}_i-\bar{y})}} \right)^2$ (Intuitive explanation)

当我们引进更多 independent variables, R^2 一/个 ⟶ R^2 代表明很好用目前已知的 x_i 来解释 y_i

如果是无关变量，则 R^2 -

④ Variance inflation factor (即衡量 Multicollinearity 的一个指标)

$VIF_j = \frac{1}{1-R_j^2}$ (多重共线性越强，VIF_j 值越大)

↳ 最好不要大于 10 (≥10 strong multicollinearity; <10 little multicollinearity)

$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)} = \frac{\sigma^2}{SST_j} VIF_j$

相比于 VIF_j，我们更在意 $Var(\hat{\beta}_j)$

4. Regression through the origin

① Regression model (Assumption)

$y = \beta_1 x_1 + \cdots + \beta_k x_n + u$

② Sample regression function

At this time, similar to SLR, $\tilde{\beta}_1, \cdots, \tilde{\beta}_k$ are biased

(忘掉 β_0 / 忘掉 relevant x_i ⟶ 都会造成 Bias)

5. Efficiency of OLS: The Gauss-Markov Theorem

The Gauss-Markov Theorem (当满足5个assumptions时) $\triangleq$ BLUE

- B = Best (with smallest variance) 对于每个 $\hat{\beta}$ 都成立
- L = Linear (每个estimator都是 y_i 的 Linear combination)
- U = Unbiased
- E = Estimator (Data → Estimator)

① Under OLS, parameters are unbiased linear estimators

SLR: $\hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2}$

MLR: $\hat{\beta}_j = \frac{\sum \hat{r}_{ij} y_i}{\sum \hat{r}_{ij}^2}$

都是 linear combination of y_i

② Under OLS, the variances of unbiased estimators are smallest

注意: Smallest variance ≠ Efficiency !!!

Smallest variance

Efficiency: 在 Unbiasedness 的前提下, Smallest variance

L11-14 Multiple Linear Regression Model —— Statistical inference

1. Classical linear model (CLM) assumptions

Classical linear model = Gauss-Markov assumption + ass 6:

$u_i \sim N(0, \sigma^2)$

(其实就是 combine ass 4 & ass 5，并进一步假设一个 Distribution)

→ $\hat{\beta_j} \sim N(\beta_j, \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ij}-\bar{x_j})^2(1-R_j^2)})$ ($\hat{\beta_j} = \beta_j + \sum w_{ij}u_i$)

$y|x \sim N(\beta_0+\beta_1x_1+\cdots+\beta_kx_k, \sigma^2)$

$E(y|x) = \beta_0+\beta_1x_1+\cdots+\beta_kx_k$

2. T-statistic

① Definition (t-statistic)

$\frac{\hat{\beta_j}-\beta_j}{se(\hat{\beta_j})} \sim t(n-(k+1))$

之所以这个 t 的统计量的自由度 = $\beta_0, \cdots, \beta_k$ ($k+1$ 个)

(这是由于 $\frac{\hat{\beta_j}-\beta_j}{sd(\hat{\beta_j})} \sim N(0,1)$)

② Properties

可以看成标准化

△ t distribution is more spread out than standard normal distribution

△ n↑, df↑, the difference between t-distribution & standard normal distribution ↓ (especially for df > 120)

(t distribution 的 df)

3. T-HT

① Step 1 (写假设检验)

- H_0
- H_1
  - One-sided
  - Two-sided
- Significance level (Type I error)
  - $\alpha = P(\text{Type 1 error}) = P(T \in C | H_0)$
  - $\beta = P(\text{Type 2 error}) = P(T \notin C | H_1)$

② Step 2 (计算统计量)

$t_{\hat{\beta_j}} = \frac{\hat{\beta_j}-\beta_j}{se(\hat{\beta_j})}$ (Based on H_0)

☆注意：STATA 所计算的 t 值都是基于 $H_0: \beta_j = 0$

$t = \frac{(\hat{\beta_1}-\hat{\beta_2})-(\beta_1-\beta_2)}{se(\hat{\beta_1}-\hat{\beta_2})}$

in STATA: vce

T 大 → H_0 is rejected → $\beta_j \neq 0$ → x_j is statistically significant

③Step 3 (判断)

a. Critical region　不能说"Accept"

$|t_{\hat{\beta}_j}| \le |C|$: Fail to reject

$|t_{\hat{\beta}_j}| > |C|$: Reject

$t_{\hat{\beta}_j}$ ↓ 计算　C ↓ 查表

☆注意：C 的下标代表 Significance level

One-sided

Two-sided　无需减半

b. p-value

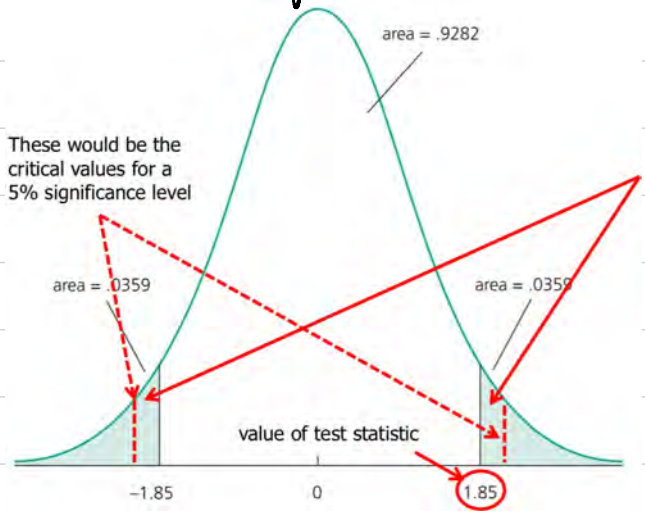
p-value = $P(|T| > t)$

p-value $\ge \alpha$: Fail to reject

p-value $< \alpha$: Reject

☆注意：STATA 所计算的 p-value 都是 two-sided

(因为 STATA 默认 Two-sided HT)



c. Confidence intervals

$CI: [\hat{\beta}_j - z_\alpha se(\hat{\beta}_j), \hat{\beta}_j + z_\alpha se(\hat{\beta}_j)]$

4. F-statistic

① Motivation

- T-statistic : Simple HT
- F-statistic : Multiple HT / Joint HT　　Unrestricted model → Restricted model

当发现 Strong multicollinearity 时，我们可以把研究对象从一个"提升"为一组.

② Definition

$$F = \frac{(SSR_r - SSR_{ur})/(df_r - df_{ur})}{SSR_{ur}/df_{ur}} \sim F(df_r - df_{ur},\ df_{ur})$$

($F > 0$；$df_r - df_{ur}$：也就是我们 test 的变量个数)

(在STATA中，用 "test" 指令)

5. F-HT

① Step 1

- H_0
  - Multiple exclusion（即 $\beta_j = 0$）
  - Overall significant (Overall exclusion)
  - General linear restriction：$\beta_1 = 1,\ \beta_2 = 2,\ \beta_3 = 3$；$\beta_1 = \beta_2,\ \beta_3 = 1$ (此时 H_0 均存在 linear combination)
- H_1 : H_0 is not true (F-HT 只能测双侧)
- Significance level

不能用 R^2 来表示 F-stat：$SST_r \neq SST_{ur}$ (Restricted model 左侧不是 y)

② Step 2

计算 F-statistic

③ Step 3

a. Critical region

- $F \leq C$: Fail to reject
- $F > C$: Reject

b. P-value

$P\text{-value} = P(F > f)$

- p-value $\geq \alpha$: Fail to reject
- p-value $< \alpha$: Reject

VARIABLES	Dependent Variable: log(salary) (1)	(2)	(3)	(4)
years	0.069*** (5.684)	0.068*** (5.592)	0.070*** (5.776)	0.071*** (5.703)
gamesyr	0.013*** (4.742)	0.016*** (10.079)	0.011*** (5.202)	0.020*** (15.023)
bavg	0.001 (0.887)	0.001 (1.331)	0.001 (0.691)	---
hrunsyr	0.014 (0.899)	0.036*** (4.964)	---	---
rbisyr	0.011 (1.500)	---	0.017*** (5.117)	---
Constant	11.192*** (38.752)	11.021*** (41.476)	11.275*** (41.197)	11.224*** (103.625)
Observations	353	353	353	353
R-squared	0.628	0.625	0.627	0.597

△ 10% → 5% → 1%

6. Relationship between F & T statistics

① F-T

$$F_{1,df_{ur}} = t^2_{df_{ur}}$$

② F-R^2

$$F = \frac{(R_{ur}^2 - R_r^2)/(df_r - df_{ur})}{(1-R_{ur}^2)/(df_{ur})}$$

proof:

$$F = \frac{(SSR_r - SSR_{ur})/(df_r - df_{ur})}{SSR_{ur}/df_{ur}}$$

Since $R^2 = 1 - \frac{SSR}{SST}$, then $SSR = (1-R^2)(SST)$

then $$F = \frac{[(1-R_r^2)SST - (1-R_{ur}^2)SST]/(df_r - df_{ur})}{(1-R_{ur}^2)SST/df_{ur}} = \frac{(R_{ur}^2 - R_r^2)/(df_r - df_{ur})}{(1-R_{ur}^2)/df_{ur}}$$

7. Change of units (不包括转变ln x)

Units change ⟶ Estimators change ⟶ R^2 stays unchanged

T stat & F stat stay unchanged

L15-16 Multiple Linear Regression Model — Further issues

1. Large sample vs Small sample

① Motivation

In CLM assumptions,

in practice, some of them cannot be satisfied. Especially, the 6th one : Normality assumption.

② Consistency

a. Definition

$\lim P(|\theta_n - \theta| < \varepsilon) = 1, \quad \forall \varepsilon > 0$ and $n \to \infty \triangleq \operatorname{plim} \theta_n = \theta$

Ass 1 ~ Ass 4 ⟶ Unbiasedness & Consistency

proof

In MLR, $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)u_i}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}$

$\operatorname{plim} \hat{\beta}_1 = \beta_1 + \frac{Cov(x_1, u)}{Var(x_1)}$

For OLS, (Ass 4 ✓)

By Ass 4 (Zero conditional mean), we get $Cov(x_1, u) = 0$, (Weaker than Ass 4)

So, $\operatorname{plim} \hat{\beta}_1 = \beta_1$ (Consistent)

For underspecified OLS, (Ass 4 ×)

For example, $\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v$ ⟶ $\tilde{y} = \beta_0 + \beta_1 x_1 + u$

$Cov(x_1, u) = Cov(x_1, \beta_2 x_2 + v) = \beta_2 Cov(x_1, x_2) + Cov(x_1, v)$

$= \beta_2 Cov(x_1, x_2)$

⟶ $\operatorname{plim} \hat{\beta}_1 = \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{Var(x_1)} = \beta_1 + \beta_2 \delta_1$ (Inconsistent)

b. Interpretation

$n \uparrow, \ \theta_n \to \theta$

Make the estimator get closer to the true value

Although not all useful estimators are unbiased, virtually all economists agree that **consistency** is a minimal requirement for an estimator. The Nobel Prize–winning econometrician Clive W. J. Granger once remarked, "If you can't get it right as n goes to infinity, you shouldn't be in this business." The implication is that, if your estimator of a particular population parameter is not consistent, then you are wasting your time.

c. Example

$\operatorname{plim}(\bar{x}_n) = \mu$

③ Asymptotic normality

Normality ⟶ Asymptotic normality

CLT, Large sample

2. Standardized effect

① Derivation

a. Simple linear regression

$$\begin{cases} y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i \\ \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \end{cases} \longrightarrow \frac{y_i - \bar{y}}{\hat{\sigma}_y} = \hat{\beta}_1 \frac{\hat{\sigma}_x}{\hat{\sigma}_y} \frac{x_i - \bar{x}}{\hat{\sigma}_x} + \frac{\hat{u}_i}{\hat{\sigma}_y}$$

标准化变量 → 标准化影响

$$\longrightarrow z_y = \hat{b}_1 z_x + \frac{\hat{u}_i}{\hat{\sigma}_y}, \quad \hat{b}_1 = \hat{\beta}_1 \frac{\hat{\sigma}_x}{\hat{\sigma}_y}$$

b. Multiple linear regression ← 标准化时没有 intercept

Beta coefficients / Standardized coefficients

$$z_y = \hat{b}_1 z_1 + \cdots + \hat{b}_K z_K + \text{error}, \quad \hat{b}_j = \hat{\beta}_j \frac{\hat{\sigma}_j}{\hat{\sigma}_y}$$

② Significance

The significance here is to standardize the effect, then to compare the economic significance / effect

从而解决单位不同这类的无法比较的问题

3. More on Functional form

① Natural log

a. Example

$\Delta \ln x = \frac{\%\Delta x}{100}$

When estimated using the data in HPRICE2, we obtain

$$\widehat{\log(price)} = 9.23 - .718 \log(nox) + .306\ rooms$$
$$(0.19)\ \ (.066)\qquad\qquad (.019) \qquad [6.7]$$
$$n = 506,\ R^2 = .514.$$

Thus, when *nox* increases by 1%, *price* falls by .718%, holding only *rooms* fixed. When *rooms* increases by one, *price* increases by approximately 100(.306) = 30.6%.

When $\hat{\beta}$ is not too large, then $\log(\hat{y})$ is a good estimate.

b. Vantages & Drawbacks

△ Vantages: (Mitigate heteroskedasticity ⟶ Better fit the assumption; Narrow down the range of dependent variable ⟶ Allay the outlier influence)

△ Drawbacks: (The variables taking natural log must be positive, and the variable had better be away from 0; We can get $\log(y)$'s, but not the original one y)

c. Applications

△ Often used for dollar amounts (wage, salary…)

△ Not used for (years; percentage-point variables)

② Quadratic form

$$\widehat{wage} = 3.73 + .298\ exper - .0061\ exper^2$$
$$(.35)\ \ (.041)\qquad\ (.0009)$$

当同一个effect同时以(一次型)与(二次型)出现在regression model中时，
通过求导判断其实际影响，通过FOC求出分界点。

③ Interaction term

Original: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$

Transformed: $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 (x_1 - u_1)(x_2 - u_2) + u$ ⟶ 可以用于判定当 x_1 | x_2 为 Sample mean 时的 partial effect

4. More on R^2

① Adjusted / Corrected R^2

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{SSR/n}{SST/n} = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_y^2}$$

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_y^2} = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

② Relationship between R^2 & $\bar{R}^2$

$$\bar{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$$

($\bar{R}^2$ 可能是负的)

③ Pros & Cons

Pro: 比较多个 x independent variables 时，Goodness-of-fit 也可能下降 (用于比较不同 model，同一 sample)

Con: The ratio of two unbiased estimators is not an unbiased estimator (我们)不认为 $\bar{R}^2$ 优于 R^2)

注意：R^2 和 $\bar{R}^2$ 都不能下降为 Goodness-of-fit 百分比来衡量 dependent variable 的形式

- 次数 ind variable 形式 ✓
- 次数 dep variable 形式 ×

17-18 Dummy variables

1. Motivation

When we want to incorporate some qualitative-information variables (e.g. Gender, region…) into our regression model, then at this time, they are called dummy variables.

(Binary variable / Zero-one variable)

2. Single dummy variable

①Process

a. Introduce dummy variables

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + u$$

δ_0 = the wage gain/loss if the person is a woman rather than a man (holding other things fixed)

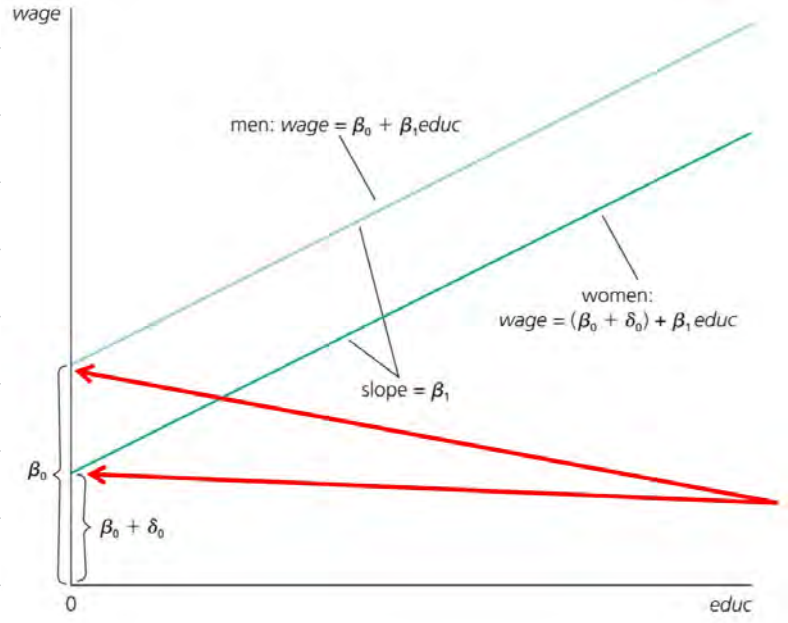
Dummy variable:
female =1 if the person is a woman
不female =0 if the person is man ← The base category is men. (benchmark category)

b. Compute the coefficient of dummy variables

$$\delta_0 = E(wage|female = 1, educ)$$

$$-E(wage|female = 0, educ)$$

②Graphic interpretation



Intercept 变
Slope 不变

③Dummy variable trap

Dummy variables 包含全部categories, 不能在model中同时出现, (i.e. 不要引进base case)

否则: Perfect collinearity

3. Multiple dummy variables

① Process

a. Introduce & transform the dummy variables

- Two qualitative variables: *married* and *female*
- Four categories:

  marrmale (married males): *married* =1, *female* =0

  marrfem (married females): *married* =1, *female* =1

  singmale (single males): *married* =0, *female* =0

  singfem (single females): *married* =0, *female* =1

为何不 "$\widehat{\log wages} = \beta_0 + \delta_0 female + \delta_1 married + \cdots$" ?:

An important limitation of this model is that the marriage premium is assumed to be the same for men and women; this is relaxed in the following example.

$$\widehat{\log(wage)} = \underset{(.100)}{.321} + \underset{(.055)}{.213}\, marrmale - \underset{(.058)}{.198}\, marrfem - \underset{(.056)}{.110}\, singfem + \underset{(.007)}{.079}\, educ + \underset{(.005)}{.027}\, exper - \underset{(.00011)}{.00054}\, exper^2$$

b. Compute the coefficients

→ 以 base category (0=0) 为标准，相互比较 effect

② Deal with ordinal variables with multiple dummy variables

Ordinal variables (0, 1, 2, 3 ···) ⟶ Multiple dummy variables (0, 1)

a. Example 1

Example: City credit ratings and municipal bond interest rates

Municipal bond rate

Credit rating from 0-4 (0=worst, 4=best)

$$MBR = \beta_0 + \beta_1 CR + other\ factors$$

This specification would probably not be appropriate as the credit rating only contains ordinal information. A better way to incorporate this information is to define dummies:

$$MBR = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + other\ factors$$

这样的表达，相比于之前模型，
体现了更多的局部对比

b. Example 2

- **Example: effects of law school rankings on starting salaries** (LAWSCH85.dta)

$$\widehat{\log(salary)} = \underset{(.41)}{9.17} + \underset{(.053)}{.700}\, top10 + \underset{(.039)}{.594}\, r11to25 + \underset{(.034)}{.375}\, r26to40 + \underset{(.028)}{.263}\, r41to60 + \underset{(.021)}{.132}\, r61to100 + \underset{(.0031)}{.0057}\, LSAT + \underset{(.074)}{.014}\, GPA + \underset{(.026)}{.036}\, log(libvol) + \underset{(.0251)}{.0008}\, log(cost)$$

$$n = 136, R^2 = .911, \bar{R}^2 = .905$$

Holding other things fixed, students from top 10 schools earn 70% more than from schools below 100

4. Interactions involving dummy variables

① Motivation

- Interaction of quantitative variable & quantitative variable ⟶ quantitative variable
- Interaction of dummy variable & quantitative variable ⟶ dummy variable

② Example

- **Restricted model (same regression for both groups)**

College grade point average

Standardized aptitude test score

High school rank percentile

$$cumgpa = \beta_0 + \beta_1 sat + \beta_2 hsperc + \beta_3 tothrs + u$$

Total hours spent in college courses

- **Unrestricted model (contains full set of interactions)**

$$cumgpa = \beta_0 + \delta_0 female + \beta_1 sat + \delta_1 female \cdot sat + \beta_2 hsperc + \delta_2 female \cdot hsperc + \beta_3 tothrs + \delta_3 female \cdot tothrs + u$$

③ Hypothesis testing (Whether there is group difference)

a. Method 1: General F-test: R ⟶ UR

- **Null hypothesis**

$H_0 : \delta_0 = 0, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$

All interaction effects are zero, i.e. the same regression coefficients apply to men and women

b. Method 2: Chow-Test

$SSR_{ur} = SSR_1 + SSR_2$

$SSR_r = SSR_p$ pooled SSR

$$\longrightarrow F = \frac{(SSR_r - SSR_{ur})/(df_r - df_{ur})}{SSR_{ur}/df_{ur}} = \frac{(SSR_p - (SSR_1 + SSR_2))/(n-(k+1)) - n+2(k+1))}{(SSR_1 + SSR_2)/(n-2(k+1))} = \frac{(SSR_p - (SSR_1 + SSR_2))/(k+1)}{(SSR_1 + SSR_2)/(n-2(k+1))}$$

这里的 SSR_p、SSR_1、SSR_2 都来自 Restrictive model

5. Binary dependent variable

① Model

Linear regression model ⟶ Linear probability model (LPM)

表以子 classification

② Derivation

Since PRF: $E(y|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K$

Conditional expectation: $E(y|x) = 1 \cdot P(y=1|x) + 0 \cdot P(y=0|x) = P(y=1|x)$

then we get $P(y=1|x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K$

$\beta_j = \frac{\partial P(y=1|x)}{\partial x_j}$

Linear probability model (LPM) 左边从y变成 $P(y=1|x)$

③ Sample regression function

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_K x_K$

④ Downside of LPM

a. $\hat{y}$ can be out of range: $\hat{y} > 1$ & $\hat{y} < 0$ can happen and cannot be explained

b. Constant partial effect

c. Heteroskedasticity: $Var(y|x) = P(y=1|x)[1 - P(y=1|x)]$ (取决于x的取值)

proof:

$Var(y|x) = E(y^2|x) - E(y|x)^2 = E(y|x) - E(y|x)^2 = E(y|x)[1 - E(y|x)]$

$= (\beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K)(1 - \beta_0 - \beta_1 x_1 - \cdots - \beta_K x_K)$

$= P(y=1|x)[1 - P(y=1|x)]$

⑤ Goodness of fit

Rather than R^2, we use "correct percent" to measure goodness-of-fit

First, we define $\tilde{y}_i = \begin{cases} 1, & \text{if } \hat{y}_i \geq 0.5 \\ 0, & \text{if } \hat{y}_i < 0.5 \end{cases}$ (y_i 别不用 predict 因为 $y_i = 0|1$)

Second, we observe the results and compute correct percent

1975		$\widetilde{inlf}_i$ = 0	1	Total
$inlf_i$	0	203	122	325
	1	78	350	428
Total		281	472	753

goodness of fit = correct percent = $\frac{203 + 350}{753}$

L18 Instrumental variables → Ass 4

1. Motivation

被迫omit：因为unobservable (如ability)

Endogeneity problem (Omit some independent variables) → Bias

这里omitted independent variable 被记入 u，则 x_i 与 u correlated，从而使得到的Regression model有偏.

Solutions to deal with Endogeneity problem:

- Proxy variables method : 用 proxy 去补充一些unobservable independent variables
- Instrumental variables method (IV) : (本节课讨论)
- Fixed effect method : (之后讨论)

2. Instrumental variables (IV)

① Definition

可以理解为：把endo中exo那part单独提取下一个variable

x's instrumental variable is z, which satisfies the following three conditions:

a. z does not appear in the regression

b. $Cov(z, x) \neq 0$ 只与这个endogenous variable x 有关

c. $Cov(z, u) = 0$

② IV-estimator

If we assume the existence of IV z, then we have $Cov(z, u) = 0$

$$\rightarrow Cov(z, y - \beta_0 - \beta_1 x) = Cov(z, y) - \beta_1 Cov(z, x) = 0$$

$$\rightarrow \beta_1 = \frac{Cov(z, y)}{Cov(z, x)}$$

$$\rightarrow \hat{\beta}_{IV} = \frac{\widehat{Cov}(z, y)}{\widehat{Cov}(z, x)} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$$

③ Quality of IV-estimator

$$plim\ \hat{\beta}_{1,OLS} = \beta_1 + Corr(x, u)\frac{\sigma_u}{\sigma_x}$$

$$plim\ \hat{\beta}_{1,IV} = \beta_1 + \frac{Corr(z, u)}{Corr(z, x)}\frac{\sigma_u}{\sigma_x}$$

poor / weak instrumental variables : $Corr(z, x)$ is small

④ Two stage least square estimation (2SLS estimation)

We use 2SLS to deal with endogenous variable:

First stage: reg this endogenous variables on other exogenous variables (Reduced form regression)

Second stage: reg y on all the exogenous variables

eg. $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_K z_{K-1} + u_1$ (y_2 endo ; $z_1, \cdots, z_{K-1}$ exo)

1st stage: $\hat{y_2} = \hat{\pi_0} + \hat{\pi_1} z_1 + \cdots + \hat{\pi}_{K-1} z_{K-1} + \hat{\pi}_K z_K$ Additional exogenous variable

2nd stage: $y_1 = \beta_0 + \beta_1 \hat{y_2} + \beta_2 z_1 + \cdots + \beta_K z_{K-1} +$ error

⑤ Test after 2SLS estimation

$H_0: \delta_1 = 0$

$H_1: \delta_1 \neq 0$

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1$$

Variable that is suspected to be endogenous

Reduced form regression:

$$y_2 = \pi_0 + \pi_1 z_1 + \cdots + \pi_k z_{k-1} + \pi_k z_k + v_2$$

Variable y_2 is exogenous if and only if v_2 is uncorrelated with u_1, i.e. if the parameter δ_1 is zero in the regression:

$$u_1 = \delta_1 v_2 + e_1$$

The null hypothesis of exogeneity of y_2 is rejected, if in this regression the parameter δ_1 is significantly different from zero

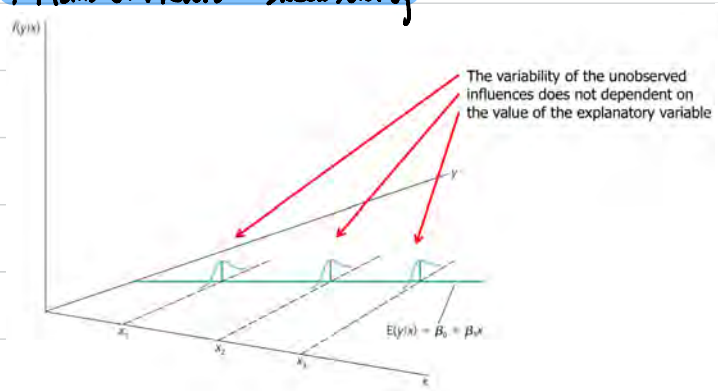
Test equation:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + \delta_1 \hat{v}_2 + e_1$$

y_2 exogenous $\longrightarrow$ $Cov(y_2, u_1) = 0 \longrightarrow Cov(v_2, u_1) = 0 \longrightarrow \delta_1 = 0$

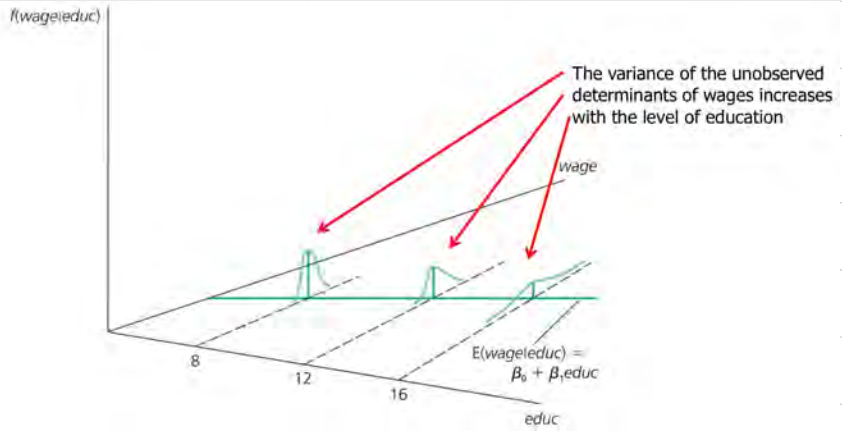
2.19 Heteroskedasticity → Ass 5

1. Homo & Hetero – skedasticity



Homoskedasticity

x变，Var(u)/Var(y)不变：$Var(u|x_i)=\sigma^2$



Heteroskedasticity

x变，Var(u)/Var(y)变：$Var(u|x_i)=\sigma_i^2=\sigma^2 h(x_i)$

2. Consequences of Heteroskedasticity

① OLS is still consistent

② Interpretation of R^2 is not changed

③ F-test & T-test are not valid

④ OLS is no longer BLUE, there may be more efficient estimators

3. Testing for heteroskedasticity

① BP test (u^2与x_i线性不相关)

H_0: Homoskedasticity : $Var(u|x)=\sigma^2$ → $H_0: \delta_1=\cdots=\delta_k=0$

H_1: Heteroskedasticity → H_1: H_0 is not true

Under H_0: $Var(u|x)=E(u^2|x)-[E(u|x)]^2=E(u^2|x)=\sigma^2$

→ $E(u^2|x_1,\cdots,x_k)=E(u^2)=\sigma^2$

→ $\hat{u}^2=\delta_0+\delta_1x_1+\cdots+\delta_kx_k+error$, where $\delta_1=\cdots=\delta_k=0$ (*) 此时用 $\hat{u}_i$ 代替 u_i

→ F test

② White test (u^2与x_i线性、二次型不相关)

a. Original white test

Start from (*), we have $\hat{u}^2=\delta_0+\delta_1x_1+\cdots+\delta_kx_k+\delta_{k+1}x_1^2+\cdots\delta_{2k}x_k^2+\delta_{2k+1}x_1x_2+\cdots\delta_nx_{k-1}x_k+error$, where $\delta_1=\cdots=\delta_n=0$

即：从一阶开至二阶，使得这个inference更准确

b. Alternative White test

$\hat{u}^2=\delta_0+\delta_1\hat{y}+\delta_2\hat{y}^2+error$, where $\delta_1=\delta_2=0$

4. Heteroskedasticity-robust inference estimation → 算Var (没有消除/减轻Heteroskedasticity，只是用了一种合适的方式去计算Var)

① Implementation

In STATA:

reg ···· , robust 加一个"robust"

② Results

Large sample → Closer to homogeneity

- Estimator不变
- Variance of estimator 变大/变小 (更为接近大)

- Robust inference → OLS isn't BLUE
- WLS → OLS is BLUE

5. Weighted least squares estimation (WLS estimation)

① Procedure of WLS estimation

If heteroskedasticity, then we def $Var(u_i|x_i)=\sigma^2 h(x_i)$, $h(x_i)=h_i>0$

$y_i=\beta_0+\beta_1 x_{i1}+\cdots+\beta_k x_{ik}+u_i$

→ Var会减小

$$\rightarrow \left(\frac{y_i}{\sqrt{h_i}}\right)=\beta_0\frac{1}{\sqrt{h_i}}+\beta_1\frac{x_{i1}}{\sqrt{h_i}}+\cdots+\beta_k\frac{x_{ik}}{\sqrt{h_i}}+\frac{u_i}{\sqrt{h_i}}$$

$$\rightarrow y_i^*=\beta_0 x_{i0}^*+\beta_1 x_{i1}^*+\cdots+\beta_k x_{ik}^*+u_i^*$$

(这里还有一项)

② Examples

a. Example 1: 已知 $h(x_i)$

Example: Savings and income

$$sav_i=\beta_0+\beta_1 inc_i+u_i,\ Var(u_i|inc_i)=\sigma^2 inc_i$$

$$\left[\frac{sav_i}{\sqrt{inc_i}}\right]=\beta_0\left[\frac{1}{\sqrt{inc_i}}\right]+\beta_1\left[\frac{inc_i}{\sqrt{inc_i}}\right]+u_i^*$$

Note that t model has

The transformed model is homoskedastic

$$E(u_i^{*2}|\mathbf{x}_i)=E\left[\left(\frac{u_i}{\sqrt{h_i}}\right)^2|\mathbf{x}_i\right]=\frac{E(u_i^2|\mathbf{x})}{h_i}=\frac{\sigma^2 h_i}{h_i}=\sigma^2$$

b. Example 2: 取平均会使得: Homo → Hetero

- **Example: facters affect pension plan contribution (individual-level)**

Contribution to pension plan for employee e in firm i | Earnings and age for employee e in firm i | Percentage firm contributes to plan | Homoskedasticity error term

$$contrib_{i,e}=\beta_0+\beta_1 earns_{i,e}+\beta_2 age_{i,e}+\beta_3 mrate_{i,e}+u_{i,e}$$

- Suppose, however, that we only have firm-level data, i.e., average values of contributions, earnings, and age by firm

取平均后

- **Firm-level equation**

Average contribution to pension plan in firm i | Average earnings and age in firm i | Percentage firm contributes to plan | Average error across all employees in firm

$$\overline{contrib}_i=\beta_0+\beta_1\overline{earns}_i+\beta_2\overline{age}_i+\beta_3 mrate_i+\bar{u}_i$$

Let m_i denote the number of employees at each firm

$$\Rightarrow Var(\bar{u}_i)=Var\left(\frac{1}{m_i}\sum_{i=1}^{m_i}u_{i,e}\right)=\sigma^2/m_i$$

Heteroskedasticity error at the firm-lev

If errors are homoskedastic at the employee level, WLS with weights equal to firm size m_i should be used.

→ 对变量取平均处理后，要使用WLS转变为Homogeneity模型

③ WLS with LPM

In LPM, $Var(y|x) = E(y^2|x) - [E(y|x)]^2 = E(y|x)[1 - E(y|x)] = P(y=1|x)(1 - P(y=1|x)) = y_i(1-y_i)$

$\longrightarrow \sigma^2 \hat{h}_i = \hat{y}_i(1-\hat{y}_i)$

$\longrightarrow \hat{h}_i = \hat{y}_i(1-\hat{y}_i)$ (σ^2 무시해도 됨)

$\longrightarrow$ Taking the above $\hat{h}_i$, we do WLS

$\hat{y}_i < 0 \longrightarrow \hat{y}_i = 0.01$

$\hat{y}_i > 0 \longrightarrow \hat{y}_i = 0.99$

$\longrightarrow$ To guarantee $\hat{h}_i > 0$

6. Feasible generalized least squares estimation (FGLS)

① Setup

If heteroskedasticity, then we def $Var(u|x) = \sigma^2 h(x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_K x_K)$

General form of heteroskedasticity, exp-function is used to ensure positivity

② Procedure of FGLS estimation

$Var(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_K x_K)$

→ $u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_K x_K) v$ (Error term)

→ $\log(u^2) = \alpha_0 + \delta_1 x_1 + \cdots + \delta_K x_K + e$

→ $\log(\hat{u}^2) = \hat{\alpha}_0 + \hat{\delta}_1 x_1 + \cdots + \hat{\delta}_K x_K$

→ $h(x_i) = \exp[\log(\hat{u}^2)]$

→ WLS

- **A Feasible GLS Procedure to Correct for Heteroskedasticity:**
  1) Estimate the model by OLS and obtain the residuals, $\hat{u}_i$.
  2) Create $\log(\hat{u}_i^2)$ by first squaring the OLS residuals and then taking the natural log.
  3) Run the regression of $\log(\hat{u}_i^2)$ on all explanatory variables and obtain the fitted values, $\hat{g}_i$
  4) Exponentiate the fitted values: $\hat{h}_i = exp(\hat{g}_i)$
  5) Find WLS estimation using weights $1/\hat{h}_i$

L20-21 Time series

1. Static model

① Definition

We focus on the variables at the same time t

② Examples

$$inf_t = \beta_0 + \beta_1 unem_t + u_t$$

$$mrdrte_t = \beta_0 + \beta_1 convrte_t + \beta_2 unem_t + \beta_3 yngmle_t + u_t$$

2. Finite distributed Lag model (FDL Models)

① Definition

We focus on the variables with a time lag

② Population model

$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t$

Transitory (Short-run) effect: 某一处/某n处的系数

Permanent (Long-run) effect: 全部系数和

③ Assumptions

a. Ass 1: Linear model

b. Ass 2: No perfect collinearity
- No independent variable is constant
- No perfect linear combination

c. Ass 3: Zero conditional mean: $E(u_t|X)=0$ For any period, u is uncorrelated with the independent variables in all periods

("Strict exogeneity") where $X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \\ x_{t1} & \cdots & x_{tk} \\ \vdots & & \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}$

- Past
- Now
- Future

⚠
- Contemporaneous exogeneity: $E(u_t|x_t)=0$ for the same period
- Strict exogeneity: $E(u_t|X)=0$ for all the periods Stronger

⚠ Ass 3 implies that in $y_t=\beta_0+\beta_1 z_t+u_t$: z has no lagged effect on y

proof: Contemporaneous exogeneity 也对

if there is lagged effect, then $z_{t-1}\neq 0$, z_{t-1} is included in u_t, then u_t is correlated with z_t, then Ass 3 is violated

⚠ Example of violating Ass 3

- **Example: City's murder rate and number of police officers**

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t$$

we assume it is uncorrected with $polpc_t$

- City adjusts the size of its police force based on past values of murder rate. That means $polpct_{t+1}$ might be correlated with $mrdrte_t$ and u_t, which violates TS.3

$u_t \longrightarrow mrdrte_t \longrightarrow polpct_{t+1}$

violate

d. Ass 4: Constant variance: $Var(u_t|X) = Var(u_t) = \sigma^2$

e. Ass 5: No serial correlation: $Corr(u_t, u_s|X)=0$, $\forall t\neq s$

- $Corr(x_t, x_s)\neq 0$
- $Corr(u_t, x_s)=0$
- $Corr(u_t, u_s)=0$

f. Ass 6: Normality: $u_t \sim N(0,\sigma^2)$

→
- Ass 1 — Ass 3: Unbiasedness
- Ass 1 — Ass 5: BLUE (满足前5个TT假设不需要满足Gauss-Markov Assumption in T)
- Ass 1 — Ass 6: The usual T & F tests are valid (非常注意 Normality 是T假设 Testing的前提)

④ Dummy variables

We use dummy variables to do event study.

Examine whether a particular event influences some outcome.

- **Example: Factors effect fertility rates** (FERTIL3.dta)

Children born per 1,000 women in year t 1913-1984 | Tax exemption in year t | Dummy for World War II years (1941-45) | Dummy for availabity of contraceptive pill (1963-present)

$$\widehat{gfr}_t = \underset{(3.68)}{98.68} + \underset{(.030)}{.083}\, pe_t - \underset{(7.46)}{24.24}\, ww2_t - \underset{(4.08)}{31.59}\, pill_t$$

$$n = 72, R^2 = .473, \bar{R}^2 = .450$$

⑤ Trend

a. Method 1: Add a trend variable

A trend variable should be included when (Some independent variables display trending behaviour; Dependent variable displays trending behaviour)

Example: Housing investment and prices (HSEINV.dta)

Per capita housing investment | Housing price index

$$\widehat{\log(invpc)} = \underset{(.043)}{-.550} + \underset{(.382)}{1.241}\, \log(price)$$

$$n = 42, R^2 = .208, \bar{R}^2 = .189$$

It looks as if investment and prices are positively related

- **Example: Housing investment and prices (cont.)**

$$\widehat{\log(invpc)} = \underset{(.136)}{-.913} + \underset{(.679)}{.381}\, \log(price) + \underset{(.0035)}{.0098}\, t$$

$$n = 42, R^2 = .341, \bar{R}^2 = .307$$

There is no significant relationship between price and investment anymore

b. Method 2: Detrending

$y_t = \alpha_0 + \alpha_1 t + u_t \longrightarrow$ let $\ddot{y}_t = \hat{u}_t$

$x_t = \beta_0 + \beta_1 t + v_t \longrightarrow$ let $\ddot{x}_t = \hat{v}_t$

(除去了 trend, constant)

$\longrightarrow$ Reg $\ddot{y}_t$ on $\ddot{x}_t$ (No intercept)

c. Relation between two methods

Adding a trend $\longrightarrow R^2 \uparrow$

Detrending $\longrightarrow R^2$

- R^2 of reg detrended y on x & t
- $= R^2$ of reg detrended y on detrended x
- $\neq R^2$ of reg y on x & t (大小)

⑥ Seasonality

a. Method 1: Adding a set of seasonal dummies

b. Method 2: Deseasonalizing 首先是下为seasons而dummy variables

L22-24 Panel data

1. Pooled OLS regression (POLS regression)

① Definition

Pooling independent cross sections across time:

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \delta_2 d_2 + \cdots \delta_n d_n$

Mean values of variables have changed over time

Include dummy variables for every period (不能不引入 base group)

② How to test whether the βs are constants over time?

Chow Test ⇓ (Variables & Time variables's interaction term 系数都为0 → 时间不影响自variable系数 → Slope coefficients 是 constant)

In Chapter 7, we discussed how the Chow test—which is simply an F test—can be used to determine whether a multiple regression function differs across two groups. We can apply that test to two

a. Compute SSRp (全 interaction term 前系数 = 0)

b. Compute $SSR_1, \cdots, SSR_T$ (全不含 Time variables 前系数 = 0 且只有含对应时间的 interaction term)

c. Compute F: $F = \frac{[SSRp - (SSR_1 + \cdots + SSR_T)] / [(T-1)K]}{(SSR_1 + \cdots + SSR_T) / [n - T(k+1)]}$

$df(SSRp) = n - k - (T-1) - 1 = (T-1)k$

$df(\Sigma SSR_i) = n - T(k+1)$

2. Policy analysis with pooled cross sections

① Motivation

Two or more independent cross sections can be used to evalute the impact of a certain event or policy change

② Difference-in-difference estimator (DiD)

a. Definition

If we found that the slope coefficients change after the policy/event's implementation, we introduce an interaction term, and at this time, the slope of this interaction term is called DiD.

b. Example

After incinerator was built:

$\widehat{rprice} = \underset{(3,093.0)}{101,307.5} - \underset{(5,827.71)}{30,688.27}\, nearinc$

Dummy equals to one if the house is near the incinerator

One has to compare with the situation **before the incinerator was built:**

$\widehat{rprice} = \underset{(2,653.79)}{82,517.23} - \underset{(4,744.59)}{18,824.37}\, nearinc$

The difference in the two coefficents on $nearinc$ is

$\hat{\delta}_1 = -30,688.27 - (-18,824.37) = -11,863.9$

$$rprice = \beta_0 + \delta_0\, after + \beta_1\, nearinc + \delta_1 after \cdot nearinc + u$$

Differential effect of being in the location and after the incinerator was built

3. Panel data analysis

① Panel data regression function

a. Two-period

$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + (a_i + u_{it})$

$d2_t$: Dummy variable, $d2_t = 1$ when $t = 2$; $d2_t = 0$ when $t = 1$

a_i : Unobserved fixed effect (不随t变化)

u_{it} : Unobserved idiosyncratic error (随t变化)

b. T-period

$y_{it} = \beta_0 + \delta_2 d2_t + \cdots + \delta_T dT_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + (a_i + u_{it})$

先i再t后k

② First-difference model (FD model)

a. Definition

We subtract Two-period / T-period regression functions, then we are able to eliminate fixed effect, and then get the true relationship between x & y

b. Example

△FD for Two-period panel data

$$crmrte_{i1987} = \beta_0 + \delta_0 \cdot 1 + \beta_1 unem_{i1987} + a_i + u_{i1987}$$

$$crmrte_{i1982} = \beta_0 + \delta_0 \cdot 0 + \beta_1 unem_{i1982} + a_i + u_{i1982}$$

Subtract: $\Rightarrow \Delta crmrte_i = \delta_0 + \beta_1 \Delta unem_i + \Delta u_i$

Fixed effect drops out! 我们订的这

Δui is uncorrelated

The variance of FD estimator may be large if the independent variable's change over the time is small:

$Var(\beta_1) = \frac{\sigma^2}{SST_x}$ (大 小)

△FD for T-period panel data

- **Example: Effect of Michigan job training program on worker productivity of manufacturing firms** (JTRAIN.dta)

  For years 1987, 1988, 1989:

$$\log(scrap_{it}) = \beta_0 + \beta_1 d88_t + \beta_2 d89_t + \beta_3 grant_{it} + \beta_4 grant_{it-1} + a_i + u_{it}$$

注意把每个时段的"grant"分开，写成各一个新的自变量 (因为系数不一样)

Additional job training in 1988 made workers more productive in 1989.

Time-invariant reasons why one firm is more productive than another are controlled for. The important point is that these may be correlated with the other independent variables.

First differenced equation using the years 1987, 1988, 1989:

$$\widehat{\Delta \log(scrap_{it})} = \underset{(.091)}{-.091} - \underset{(.125)}{.096}\, d89_t - \underset{(.131)}{.223}\, \Delta grant_{it} - \underset{(.258)}{.351}\, \Delta grant_{it-1}$$

③Fixed effects estimation

a. Definition

Since $a_i = \bar{a}_i$, then we can use time-demeaning method to eliminate fixed effect

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}$$

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{i1} + \cdots + \beta_k \bar{x}_{ik} + \bar{a}_i + \bar{u}_i$$

Time demeaning → $(y_{it} - \bar{y}_i) = \beta_1 (x_{it1} - \bar{x}_{i1}) + \cdots + \beta_k (x_{itk} - \bar{x}_{ik}) + (u_{it} - \bar{u}_i)$ 关键在于保证 u_{it} is uncorrelated

At this time, $\beta_1, \cdots, \beta_k$ are called FE estimator

FE estimator is equivalent to introducing a dummy variable in the original regression and using pooled OLS

b. Example

Fixed-effects estimation using the years 1987, 1988, 1989:

$$\widehat{log(scrap_{it})} = \underset{(.109)}{-.080}\, d88_t^* \underset{(.133)}{-.247}\, d89_t^* \underset{(.151)}{-.252}\, grant_{it}^* \underset{(.210)}{-.422}\, grant_{it-1}^*$$

$$n = 162, R^2 = .201$$

Training grants significantly improve productivity (with a time lag)

c. Further discussion

⚠ The R^2 of the demeaned equation is inappropriate

⚠ The effect of time-invariant variables cannot be estimated

⚠ The effect of interactions with time-invariant variables can be estimated

⚠ Degree of freedom have to be adjusted because the N time average are estimated in addition

d. FD & FE

⚠
- T=2 : FD = FE
- T>2 : FE > FD

⚠ FD may be better if there is severe serial correlation in the errors

⚠ Generally, it's a good idea to compute both

④ Random effects model

a. Assumption

$\underline{Cov(x_{itj}, a_i) = 0}$ for $y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + u_{it}$

b. Downside

We set up an assumption to satisfy Zero conditional mean assumption,

but still the Ass 5 "No serial correlation" is not satisfied.

$Cov(a_i + u_{it}, a_i + u_{is}) = Cov(a_i, a_i) = \sigma_a^2$

c. Transformation to correct Ass 5.

We introduce $\theta = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}}$ $\left(\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + T\hat{\sigma}_a^2}}\right)$, and then we have:

$(y_{it} - \theta\bar{y}_i) = \beta_0(1-\theta) + \beta_1(x_{it1} - \theta\bar{x}_{i1}) + \cdots + \beta_K(x_{itK} - \theta\bar{x}_{iK}) + (a_i - \theta\bar{a}_i + u_{it} - \theta\bar{u}_i)$

This equation is called Quasi-demeaned equation, the coefficients are called RE estimator

$\theta \in [0, 1]$ → 0 OLS

→ 1 FE