Lecture 1 (Continued from Slides) 13/10/2022

→ Derivation of OLS Estimators

Let $\varepsilon_i = y_i - b_1 x_i - b_0$ we search for values $b_0, b_1 \in \mathbb{R}$ s.t the quantity

$$S(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$ is minimized.

That is we need to solve the problem

$$\min_{b_0, b_1} S(b_0, b_1)$$

To do so we rely on the first order conditions (FOCs):

$$\left.\begin{aligned} \frac{\partial S(b_0, b_1)}{\partial b_0} &= 0 \\ \frac{\partial S(b_0, b_1)}{\partial b_1} &= 0 \end{aligned}\right\} \iff \left.\begin{aligned} -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) &= 0 \\ -2\sum(y_i - b_0 - b_1 x_i) x_i &= 0 \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \begin{aligned} \sum_{i=1}^{n} y_i &= na + b\sum_{i=1}^{n} x_i \quad (1) \\ \sum_{i=1}^{n} x_i y_i &= b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 \quad (2) \end{aligned}$$

Known as normal equations of the model

Then, from (1) we have that ②

$$\hat{b}_0 = \bar{y} - b_1\bar{x} \quad (3), \text{ where } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Then we plug (3) into (2) and we have that

$$\sum_{i=1}^{n} x_i y_i - (\bar{y} - b_1\bar{x})n\bar{x} - b_1\sum_{i=1}^{n} x_i^2 = 0 \iff$$

$$\iff b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

By noting that $Cov(X,Y) = EXY - EXEY$ and $Var X = EX^2 - (EX)^2$ it's easy to see that

$$\hat{b}_1 = \frac{Cov(X,Y)}{Var(X)}$$

Moreover, since

$$\sum x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + n\bar{x}\bar{y} - n\bar{x}\bar{y} =$$

$$= \sum x_i y_i - \bar{x}\sum_{i=1}^{n} y_i + n\bar{x}\bar{y} - n\bar{y}\sum_{i=1}^{n} x_i + n\bar{x}\bar{y}$$

$$= \sum (x_i - \bar{x})(y_i - \bar{y})$$

we conclude that

$$\hat{b}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

- Goodness of fit

let $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$ ← variability of the observations

$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ ← variability from the regression model

$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ ← sum of squared errors

Then $SST = SSR + SSE$

(Hint: $(y_i - \bar{y}) = y_i - \hat{y}_i + \hat{y}_i - \bar{y}$ → take squares and sums and work out the algebra

Then, R^2 is the coefficient of determination given by

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \in (0, 1)$$

↳ high values desired

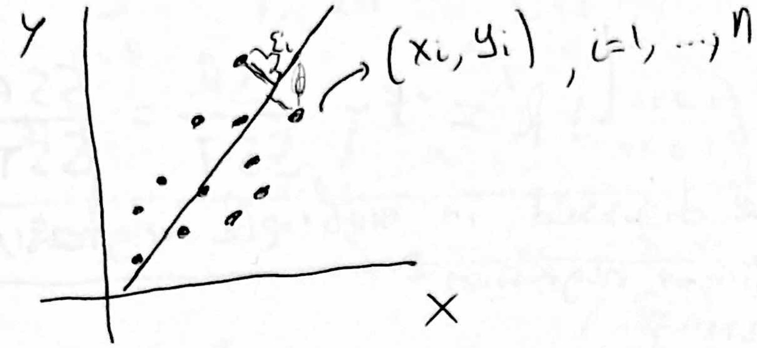
↳ very high danger for overfitting

Econometrics: Lecture 2

20/10/2022

①

- Recap from Lecture 1



- We assume that $y = \beta_0 + \beta_1 x$ and we aim to use the data $\{x_i, y_i\}$ to estimate β_0, β_1 by $\hat{\beta}_0, \hat{\beta}_1$

Since $\{x_i, y_i\}$ are samples (prone to eg measurement error) we also include an error term in the model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{Dist}(\theta)$$

$\beta_1 \to \frac{\Delta y}{\Delta x}$

- We are interested in 3 tasks
  - i) estimation and goodness of fit
  - ii) Hypothesis testing
  - iii) Confidence intervals
  - iv) Properties of the estimators (finite- and large-sample size)

i) OLS estimation by minimising the quantity

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

we find $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_x^2}$

and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

②

- Goodness of fit : $SST = SSR + SSE$

$$\underbrace{SST}_{\sum_{i=1}^{n}(y_i-\bar{y})^2} = \underbrace{SSR}_{\sum(\hat{y}_i-\bar{y})^2} + \underbrace{SSE}_{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$$

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

$\boxed{\text{ii) – iv) to be discussed in multiple regression setting}}$

- Briefly for simple linear regression
  - Hypothesis testing

$$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0$$

Calculate $t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$ and use this statistic to decide about rejection for H_0 or not.

$\hookrightarrow$ if $|t| > 1.96$

- Confidence intervals : 95%

$$\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)$$ $\leftarrow$ if 0 is not included equiv. with rejection of H_0 in 5% signif. level

- Properties
  - $\hookrightarrow$ Unbiased estimator $E(\hat{\beta}_1) = \beta_1$
  - $\hookrightarrow$ $Var(\hat{\beta}_1) \propto \frac{1}{n} \xrightarrow[n\to\infty]{} 0$

  $\Rightarrow \hat{\beta}_1 \to \beta_1$

  - $\hookrightarrow$ CLT: $n \to \infty$: $\frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{Var(\hat{\beta}_1)}} \xrightarrow{d} N(0,1)$

• Multiple Linear Regression (more than one Covariates)

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\,\underset{k\times 1}{\beta} + \underset{k\times 1}{\varepsilon} \qquad (1)$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 = x_{11} & x_{12} \dots & x_{1\,k-1} \\ 1 = x_{21} & x_{22} \dots & x_{2\,k-1} \\ \vdots & & \\ 1 = x_{n1} & x_{n2} \dots & x_{n\,k-1} \end{pmatrix}$$

intercept (β_0)

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{k-1} \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$(1) \Rightarrow \quad y_i = \beta_0 \overset{1}{x_{i1}} + \beta_1 x_{i2} + \dots + \beta_{k-1} x_{ik} + \varepsilon_i, \quad i = 1, \dots, n$$

$$(1) \Rightarrow \quad y_i = \underset{k\times 1}{x_i'}\beta + \varepsilon_i$$

→ Gauss-Markov Assumptions : sufficient conditions for the finite sample properties of the OLS estimators

GM1 : Linearity

GM2: No multicollinearity : X is $n\times k$ full column rank matrix → all of its colums are linearly indendent

* Assume $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{2i} + \varepsilon_i$

where $x_{i1} = x_{2i}$

then $y_i = \beta_0 + (\beta_1 + \beta_2) x_{i1} + \varepsilon_i$

and the sum is non-indentifiable

GM3: $E(\varepsilon_i | X) = 0 \quad , i=1,..,n$

Then, • $E(\varepsilon_i) = E\left(E(\varepsilon_i | X) \right) = E(0) = 0$

• $E(\varepsilon_i X_{jk}) = E\left[E(\varepsilon_i x_{jk} | X) \right] =$

$= E\left(x_{jk} \, \underbrace{E(\varepsilon_i | X)}_{\to 0} \right) = 0 \Rightarrow$

$\Rightarrow Cov(\varepsilon_i , x_{jk}) = 0$

* GM1 + GM3 $\Rightarrow E(y | X) = X\beta$

population regression line

GM4: Constant error

$E(\varepsilon_i^2 | X) = \sigma^2 \Rightarrow Var(\varepsilon_i | X) = \sigma^2 \quad \forall i=1,..,n$

and $E(\varepsilon_i \varepsilon_j | X) = 0 \Rightarrow Cov(\varepsilon_i, \varepsilon_j | X) = 0.$

GM4 $\Rightarrow Var(y|X) = Var(X\beta + \varepsilon) = \sigma^2 I_n$

this is why GM4 is also known as spherical covariance matrix for the errors.

*

5

- The OLS estimator:

Recall our aim: We observe y_i and x_i and we need to find b such that the difference

$$y_i - x_i' b \text{ is as small as possible}$$

This can be achieved by using the least squares approach which means to minimize wrt b the quantity

$$S(b) = \sum_{i=1}^{n} (y_i - x_i' b)$$

$$\overset{\text{matrix}}{\underset{\text{form}}{=}} (y - Xb)'(y - Xb) \left(= \| y - Xb \| \right)$$

L^2-norm in $\mathbb{R}^n$

That is we need to solve the problem

$$\min_b S(b) = (y - Xb)'(y - Xb) =$$

$$= y'y - 2b'X'y + b'X'Xb$$

Thus minimization is achieved when

$$\frac{\partial S(b)}{\partial b} = 0 \Leftrightarrow -2X'y + 2X'Xb = 0$$

$$\Leftrightarrow X'Xb = X'y \Leftrightarrow$$

$$\Leftrightarrow \hat{b} = \boxed{(X'X)^{-1} X'y} = \boxed{\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i}$$

* $X'X$ is invertible since $\text{rank}(X) = k$ i.e. no multicollinearity

* Easy to check that $\frac{\partial^2 S(\beta)}{\partial \beta^2} = X'X > 0$

and we have a minimum indeed.

(Hint: positive definite $X'X$.)

Theoritical exercise (Proof):

Since $\text{rank}(X) = k$ and X has k columns (X full rank)

we have that there is not any k-dim vector $a \neq 0_{k\times1}$

s.t. $\underset{n\times k}{X}\,\underset{k\times1}{a} = 0_{n\times1}$ (linear independent columns)

This means that exists a n-dim vector $c \neq 0$

such that $Xa = c \Rightarrow$

$\Rightarrow a'X'Xa = c'c = \sum_{i=1}^{n} c_i^2 > 0$

since $c_i \neq 0 \quad \forall i = 1, \dots, n$.

Algebraic Properties of OLS estimators:

a) • OLS is ~~an estimat~~ linear transformation of the covariates since

$$\hat{\beta} = \underbrace{(\underset{k\times n}{X'}\underset{n\times k}{X})^{-1}\underset{k\times n}{X'}}_{\text{depends only on } X}\underset{n\times1}{y} = \underset{k\times n}{A}\,\underset{n\times1}{y} = \begin{cases} k \text{ equations} \\ \text{of the form} \\ \sum_{j=1}^{n} a_{\cdot j}(x) y_j \end{cases}$$

b) • Residuals $\hat{\varepsilon}_i = y_i - x_i'\hat{\beta}$ or in matrix form $\hat{\varepsilon} = y - X\hat{\beta}$

• From the first order conditions we have

$X'(y - X\hat{\beta}) = 0 \Leftrightarrow \boxed{X'\hat{\varepsilon} = 0} \rightarrow$ useful conditio[n] for deriving the OL[S] estimator by using the M[...] Gam[...]

$(X\beta)'\hat{\varepsilon} = \beta' X'\hat{\varepsilon}$

γ) From β) follows that $\boxed{\hat{y}'\hat{\varepsilon} = 0}$

δ) $\boxed{X'\hat{\varepsilon} = 0} \xrightarrow[\text{element}]{\text{first}} \sum x_{i1}\hat{\varepsilon}_i = \sum_{i=1}^{n} \hat{\varepsilon}_i = 0$

ε) $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (\underbrace{x_i'\hat{\beta}}_{\hat{y}_i} + \hat{\varepsilon}_i) =$

$= \sum_{i=1}^{n} \hat{y}_i + \cancel{\sum_{i=1}^{n} \hat{\varepsilon}_i} \Rightarrow \boxed{\bar{y} = \bar{\hat{y}}}$

Geometric interpretation of OLS method:

OLS is an approx. solution (since we have errors) to the system $y \approx X\beta$ $\left(\begin{array}{l} n \text{ equation} \\ k \text{ unknowns} \\ k << n \end{array}\right)$

- We learn $X\beta$ which is a linear combination of the vectors of the covariates by minimizing the L^2-norm of the residuals $\varepsilon_i = y_i - x_i'\beta$ $(\varepsilon = y - X\beta)$ $\min_{\beta} \|\varepsilon\|$

Thus $y - X\beta$ is a vector with minimum length in a linear space spanned by the colums of X (denote by set of all linear combinations of the colums of X.) where y is projected orthongally

$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y =$

$= Py$

projection matrix → see also behind