17.800 OLS Estimators - Handout

Yiqing Xu

1 OLS Setup

- $\{X, Y\}$ jointly distributed, characterized by $Y = \beta_0 + \beta_1 X + u$, $E[u|X] = 0$
- Observe i.i.d sample $\{x_i, y_i\}_{i=1}^n$, or $(x_1, y_1),\ (x_2, y_2), \cdots, (x_n, y_n)$
- OLS estimators: $\hat{\beta}_0$ and $\hat{\beta}_1$
- Fitted values: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- Residuals: $\hat{u}_i \equiv y_i - \hat{y}_i$
- $y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$ by definition!

2 Derive OLS Estimators

- Objective function: minimize mean square error (Slides7 p.14)

$$\begin{aligned} S(\tilde{\beta}_0, \tilde{\beta}_1) &= \sum_{i=1}^n (y_i - \tilde{\beta}_0 - x_i\tilde{\beta}_1)^2 \\ &= \sum_{i=1}^n (y_i^2 - 2y_i\tilde{\beta}_0 - 2y_i\tilde{\beta}_1 x_i + \tilde{\beta}_0^2 + 2\tilde{\beta}_0\tilde{\beta}_1 x_i + \tilde{\beta}_1^2 x_i^2) \end{aligned}$$

- First order conditions (FOCs) (Slides7 p.15):

$$\frac{\partial S(\tilde{\beta}_0, \tilde{\beta}_1)}{\partial \tilde{\beta}_0} = \sum_{i=1}^n (-2y_i + 2\tilde{\beta}_0 + 2\tilde{\beta}_1 x_i) \tag{1}$$

$$\frac{\partial S(\tilde{\beta}_0, \tilde{\beta}_1)}{\partial \tilde{\beta}_1} = \sum_{i=1}^n (-2y_i x_i + 2\tilde{\beta}_0 x_i + 2\tilde{\beta}_1 x_i^2) \tag{2}$$

- Solve the first FOC (1) (Slides7 p.16):

$$\begin{aligned} 0 &= \sum_{i=1}^n (-2y_i + 2\hat{\beta}_0 + 2\hat{\beta}_1 x_i) \\ n\hat{\beta}_0 &= \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i \end{aligned}$$

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

- Solve the second FOC (2) (Slides7 p.16)p:

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}(-2y_ix_i + 2\hat{\beta}_0x_i + 2\hat{\beta}_1x_i^2) \\
\hat{\beta}_1\sum_{i=1}^{n}x_1^2 &= \sum_{i=1}^{n}x_iy_i - \hat{\beta}_0\sum_{i=1}^{n}x_i \\
\hat{\beta}_1\sum_{i=1}^{n}x_1^2 &= \sum_{i=1}^{n}x_iy_i - (\bar{y} - \hat{\beta}_1\bar{x})\sum_{i=1}^{n}x_i \\
\hat{\beta}_1\sum_{i=1}^{n}x_1^2 &= \sum_{i=1}^{n}x_iy_i - \bar{y}\sum_{i=1}^{n}x_i + \hat{\beta}_1\bar{x}\sum_{i=1}^{n}x_i \\
\hat{\beta}_1\sum_{i=1}^{n}x_1^2 &= \sum_{i=1}^{n}x_iy_i - \frac{1}{n}\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i + \hat{\beta}_1\frac{1}{n}(\sum_{i=1}^{n}x_i)^2 \\
\hat{\beta}_1(\sum_{i=1}^{n}x_1^2 - \frac{1}{n}(\sum_{i=1}^{n}x_i)^2) &= \sum_{i=1}^{n}x_iy_i - \frac{1}{n}\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}x_iy_i - \frac{1}{n}\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_1^2 - \frac{1}{n}(\sum_{i=1}^{n}x_i)^2}
\end{aligned}
$$

$$
\therefore \hat{\beta}_1 = \frac{n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i}{n\sum_{i=1}^{n}x_1^2 - (\sum_{i=1}^{n}x_i)^2}
$$

- Show $\hat{\beta}_1 = \frac{\widehat{Cov}(x_i,y_i)}{\widehat{Var}(x_i)}$

  We start with the following equation:

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
\Leftrightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_iy_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{\sum_{i=1}^{n}(x_i^2 - 2\bar{x}x_i + \bar{x}_i^2)} \\
\Leftrightarrow \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}x_iy_i - \frac{1}{n}\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}x_i^2 - \frac{1}{n}(\sum_{i=1}^{n}x_i)^2}
\end{aligned}
$$

  Almost by definition:

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\widehat{Cov}(x_i, y_i)}{\widehat{Var}(x_i)}
$$

3 Properties by Construction (Slides7 p.17)

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

$$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \text{ which follows FOC (1)}$$

- $\sum_{i=1}^{n} x_i \hat{u}_i = 0$

$$\sum_{i=1}^{n} x_i \hat{u}_i = \sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \text{ which follows FOC (2)}$$

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

$$\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 x_i)\hat{u}_i = \beta_0 \sum_{i=1}^{n} \hat{u}_i + \beta_1 \sum_{i=1}^{n} \hat{x}_i \hat{u}_i = 0$$

- $\bar{\hat{y}} = \bar{y}$

$$\begin{aligned}
\Leftrightarrow \quad & \sum_{i=1}^{n} \hat{y} = \sum_{i=1}^{n} y_i \\
\Leftrightarrow \quad & \sum_{i=1}^{n} (\hat{y} - y_i) = 0 \\
\Leftrightarrow \quad & \sum_{i=1}^{n} \hat{u}_i = 0
\end{aligned}$$

4 Total Sum of Squares

- SST = SSE + SSR (Slides7 p.24)

$$\sum_i^n (y_i - \bar{y})^2 = \sum_i^n (\hat{y} - \bar{y})^2 + \sum_i^n (y_i - \hat{y})^2$$

$$\sum_i^n (y_i - \bar{y})^2 - \sum_i^n (\hat{y} - \bar{y})^2 = \sum_i^n (y_i - \hat{y})^2$$

$$\sum_i^n (y_i - \bar{y} - \hat{y} + \bar{y})(y_i - \bar{y} + \hat{y} - \bar{y}) = \sum_i^n (y_i - \hat{y})^2$$

$$\sum_i^n (y_i - \hat{y})(y_i - 2\bar{y} + \hat{y}) = \sum_i^n (y_i - \hat{y})^2$$

$$\sum_i^n (y_i - \hat{y})(y_i - 2\bar{y} + \hat{y}) - \sum_i^n (y_i - \hat{y})^2 = 0$$

$$\sum_i^n (y_i - \hat{y})(y_i - 2\bar{y} + \hat{y} - y_i + \hat{y}) = 0$$

$$\sum_i^n (y_i - \hat{y})(2\hat{y} - 2\bar{y}) = 0$$

$$\sum_i^n u_i(2\hat{y} - 2\bar{y}) = 0$$

$$\sum_i^n u_i\hat{y} - \bar{y}\sum_i^n u_i = 0$$

5 Preparation

- **The Weak Law of Large Numbers (LLN):** Suppose $\{x_i\}$ is an infinite sequence of i.i.d. random variables with finite expected value $E[x_i] = \mu < \infty$. Then as n approaches infinity, the random variable $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$ converges in probability to μ.

$$\bar{x}_n \xrightarrow{p} \mu, \text{ as } n \to \infty.$$

- **Lindeberg-Levy Central Limit Theorem (CLT)**: Suppose $\{x_i\}$ is an infinite sequence of i.i.d. random variables with finite expected value and variance, $E[x_i] = \mu < \infty$ and $Var(x_i) = \sigma^2 < \infty$. Then as n approaches infinity, the random variable $\sqrt{n}(\bar{x}_n - \mu)$ converges in distribution to a normal distribution $N(0, \sigma^2)$:

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \sigma^2), \text{ as } n \to \infty.$$

- **Slutskty's Theorem:** Let $\{x_n\}, \{y_n\}$ be two sequences of random variables (or matrices), if $\{x_n\}$ converges in probability to a constant c and $\{y_n\}$ converges in distribution to a random variable (or matrix) y, then $x_n y_n \xrightarrow{d} cy$.

6 OLS Setup in Matrix

Let $\mathbf{X}$ be our $n \times (k+1)$ data matrix for k predictors (plus the constant), $\mathbf{y}$ be the $n \times 1$ vector of the outcome variable, $\boldsymbol{\beta}$ be the $(k+1) \times 1$ vector of true coefficient (intercept and k slopes), and $\mathbf{u}$ be the $n \times 1$ vector of error terms:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}_{n\times(k+1)} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}_{n\times 1}$$

Therefore, a linear model of n observations and k regressors can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Let the OLS estimates be the $(k+1) \times 1$ vector of estimated coefficients:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{(k+1)\times 1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

We assume no perfect collinearity when we write down the above equation (symmetric $(k+1)\times(k+1)$ matrix $\mathbf{X'X}$ is therefore invertible). When we aslo have the linearity assumption, we can write:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\boldsymbol{\beta}+\mathbf{u}) = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'u}$$

7 Properties of the OLS Estimator

7.1 Unbiasedness (Slides 9.1-48)

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= E[\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'u}|\mathbf{X}] \quad \text{(linearity and no perfect collinearity)} \\ &= E[\boldsymbol{\beta}|\mathbf{X}] + E[(\mathbf{X'X})^{-1}\mathbf{X'u}|\mathbf{X}] \\ &= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}E[\mathbf{u}|\mathbf{X}] \\ &= \boldsymbol{\beta} \quad \text{(zero conditional mean)} \\ \therefore \quad E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\beta} \quad \text{(random sampling)} \end{aligned}$$

Therefore, in order to have unbiasedness of the OLS estimator, we need four assumptions:

1. linearity $\mathbf{y} = \mathbf{X'}\boldsymbol{\beta} + \mathbf{u}$
2. random sampling of $\{X_i, y_i\}$
3. no perfect collinearity in $\mathbf{X}$ (therefore $(\mathbf{X'X})^{-1}$ exists)
4. zero conditional mean $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$.

7.2 Consistency (Slides 9.1-54/56)

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'u} \quad \text{(linearity and no perfect collinearity)} \\ &= \boldsymbol{\beta} + \left(\sum\nolimits_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\sum\nolimits_{i=1}^{n} \mathbf{x}_i' u_i\right) \\ &= \boldsymbol{\beta} + \left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{x}_i' u_i\right) \end{aligned}$$

Applying the LLN to the sample means (random sampling):

$$\begin{aligned} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right) &\xrightarrow{p} E[\mathbf{x}_i'\mathbf{x}_i] \equiv \mathbf{Q}_{\mathbf{(k+1)\times(k+1)}} \quad \text{(a positive definite matrix } \mathbf{Q} \text{ exist)} \\ \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i' u_i\right) &\xrightarrow{p} E[\mathbf{x}_i' u_i] = \mathbf{0} \quad \text{(zero conditional mean)} \end{aligned}$$

We have $\text{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \mathbf{Q}^{-1}\cdot\mathbf{0} = \boldsymbol{\beta}$. Therefore, in order to have consistency of the OLS estimator, we need five assumptions: (1) linearity, (2) random sampling, (3) no perfect collinearity,

(4) zero conditional mean, and (5) the second moment of $\mathbf{x}_i$ exists.

Note that if $\mathbf{Q}$ is positive definite, it is non-singular; therefore, $\mathbf{Q}^{-1}$ exist. Also note that $E[\mathbf{x}_i' u_i] = 0$ since $E[\mathbf{x}_i' E[u_i|\mathbf{x}_i]] = 0$ by the Law of Iterated Expectations.

7.3 Asymptotic Normality

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta} + \left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{x}_i' u_i\right) \\
&\qquad \text{(linearity and no perfect collinearity)} \\
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left[\sqrt{n}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} \mathbf{x}_i' u_i\right)\right]
\end{aligned}$$

$$\begin{aligned}
E[\mathbf{x}_i'\mathbf{x}_i] &= \mathbf{Q} \\
E[\mathbf{x}_i' u_i] &= \mathbf{0} \quad \text{(zero conditional mean)} \\
Var(\mathbf{x}_i' u_i) &= E_x[Var(\mathbf{x}_i' u_i|\mathbf{x}_i)] \\
&= E_x[E[\mathbf{x}_i' u_i u_i' \mathbf{x}_i'|\mathbf{x}_i]] \\
&= \sigma^2 E_x[\mathbf{x}_i'\mathbf{x}_i] \quad \text{(homoskedasticity)} \\
&= \sigma^2 \mathbf{Q} \quad \text{(a positive definite matrix } \mathbf{Q} \text{ exist)}
\end{aligned}$$

$$\begin{aligned}
E[\mathbf{Q}^{-1}\mathbf{x}_i' u_i] &= \mathbf{0} \\
Var(\mathbf{Q}^{-1}\mathbf{x}_i' u_i) &= \sigma^2 \mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1} \\
&= \sigma^2 \mathbf{Q}^{-1}
\end{aligned}$$

Applying the LLN to the sample mean of $\mathbf{x}_i'\mathbf{x}_i$ (random sampling):

$$\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i'\mathbf{x}_i\right) \xrightarrow{p} E[\mathbf{x}_i'\mathbf{x}_i] = \mathbf{Q}$$

Applying the CLT to the sample mean of $\mathbf{x}_i' u_i$ (random sampling):

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i' u_i - \mathbf{0}\right) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q})$$

Finally, applying the Slutsky's Theorem:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \sigma^2\mathbf{Q}^{-1}).$$

Therefore, in order to have asymptotic normality of the OLS estimator, six assumptions are needed: (1) linearity, (2) random sampling, (3) no perfect collinearity, (4) zero conditional mean, (5) the second moment of $\mathbf{x}_i$ exists and (6) homoscedasticity.