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: \( \hat{u}_i = y_i - \hat{y}_i \)
- \( y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i \) by definition!

2 Derive OLS Estimators

- Objective function: minimize mean square error (Slides7 p.14)
  \[
  S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - x_i \beta_1)^2
  = \sum_{i=1}^n (y_i^2 - 2y_i \beta_0 - 2y_i \beta_1 x_i + \beta_0^2 + 2\beta_0 \beta_1 x_i + \beta_1^2 x_i^2)
  \]
  - First order conditions (FOCs) (Slides7 p.15):
    \[
    \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^n (-2y_i + 2\hat{\beta}_0 + 2\hat{\beta}_1 x_i) \tag{1}
    \]
    \[
    \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^n (-2y_i x_i + 2\hat{\beta}_0 x_i + 2\hat{\beta}_1 x_i^2) \tag{2}
    \]
  - Solve the first FOC (1) (Slides7 p.16):
    \[
    0 = \sum_{i=1}^n (-2y_i + 2\hat{\beta}_0 + 2\hat{\beta}_1 x_i)
    \]
    \[
    n\hat{\beta}_0 = \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_i
    \]
    \[
    \therefore \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
    \]
• Solve the second FOC (2) (Slides7 p.16):

\[ 0 = \sum_{i=1}^{n} (-2y_i x_i + 2\hat{\beta}_0 x_i + 2\hat{\beta}_1 x_i^2) \]

\[ \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i \]

\[ \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) \sum_{i=1}^{n} x_i \]

\[ \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i \]

\[ \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + \hat{\beta}_1 \frac{1}{n} (\sum_{i=1}^{n} x_i)^2 \]

\[ \hat{\beta}_1 (\sum_{i=1}^{n} x_i^2 - \frac{1}{n} (\sum_{i=1}^{n} x_i)^2) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i \]

\[ \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} (\sum_{i=1}^{n} x_i)^2} \]

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

• Show \( \hat{\beta}_1 = \frac{\text{Cov}(x_i, y_i)}{\text{Var}(x_i)} \)

We start with the following equation:

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

\[ \Leftrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y})}{\sum_{i=1}^{n} (x_i^2 - 2 \bar{x} x_i + \bar{x}^2)} \]

\[ \Leftrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} (\sum_{i=1}^{n} x_i)^2} \]

Almost by definition:

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


3 Properties by Construction (Slides7 p.17)

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

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

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

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

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

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

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

\[
\Leftrightarrow \quad \sum_{i=1}^{n} \hat{y} = \sum_{i=1}^{n} y_i
\]

\[
\Leftrightarrow \quad \sum_{i=1}^{n} (\hat{y} - y_i) = 0
\]

\[
\Leftrightarrow \quad \sum_{i=1}^{n} \hat{u}_i = 0
\]
4 Total Sum of Squares

- \(\text{SST} = \text{SSE} + \text{SSR}\) (Slides7 p.24)

\[\sum_{i}^{n} (y_i - \bar{y})^2 = \sum_{i}^{n} (\hat{y} - \bar{y})^2 + \sum_{i}^{n} (y_i - \hat{y})^2\]

\[\sum_{i}^{n} (y_i - \bar{y})^2 - \sum_{i}^{n} (\bar{y} - \hat{y})^2 = \sum_{i}^{n} (y_i - \hat{y})^2\]

\[\sum_{i}^{n} (y_i - \bar{y} - \hat{y})(y_i - \bar{y} + \hat{y}) = \sum_{i}^{n} (y_i - \hat{y})^2\]

\[\sum_{i}^{n} (y_i - \hat{y})(y_i - 2\bar{y} + \hat{y}) = \sum_{i}^{n} (y_i - \hat{y})^2\]

\[\sum_{i}^{n} (y_i - \hat{y})(y_i - 2\bar{y} + \hat{y}) - \sum_{i}^{n} (y_i - \hat{y})^2 = 0\]

\[\sum_{i}^{n} (y_i - \hat{y})(y_i - 2\bar{y} + \hat{y} - y_i + \hat{y}) = 0\]

\[\sum_{i}^{n} (y_i - \hat{y})(2\hat{y} - 2\bar{y}) = 0\]

\[\sum_{i}^{n} u_i (2\hat{y} - 2\bar{y}) = 0\]

\[\sum_{i}^{n} u_i \hat{y} - \bar{y} \sum_{i}^{n} u_i = 0\]
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 \( \mu \).

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

- **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 \( \text{Var}(x_i) = \sigma^2 < \infty \). Then as \( n \) approaches infinity, the random variable \( \sqrt{n}(\bar{x}_n - \mu) \) converges in distribution to a normal distribution \( N(0, \sigma^2) \):

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

- **Slutsky’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 \( x_ny_n \xrightarrow{d} cy \).

6 OLS Setup in Matrix

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

\[
y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix}_{n \times (k+1)}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1}, \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}_{n \times 1}
\]

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

\[
y = X\beta + u
\]

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

\[
\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{(k+1) \times 1} = (X'X)^{-1}X'y
\]
We assume no perfect collinearity when we write down the above equation (symmetric \((k + 1) \times (k + 1)\) matrix \(X'X\) is therefore invertible). When we also have the linearity assumption, we can write:

\[
\hat{\beta} = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u
\]

7 Properties of the OLS Estimator

7.1 Unbiasedness (Slides 9.1-48)

\[
E[\hat{\beta}|X] = E[\beta + (X'X)^{-1}X'u|X] \quad \text{(linearity and no perfect collinearity)}
\]

\[
= E[\beta|X] + E[(X'X)^{-1}X'u|X]
\]

\[
= \beta + (X'X)^{-1}X'E[u|X]
\]

\[
\therefore E[\hat{\beta}] = \beta \quad \text{(zero conditional mean)}
\]

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

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

7.2 Consistency (Slides 9.1-54/56)

\[
\hat{\beta} = \beta + (X'X)^{-1}X'u \quad \text{(linearity and no perfect collinearity)}
\]

\[
= \beta + \left(\sum_{i=1}^{n} x'_i x_i\right)^{-1}\left(\sum_{i=1}^{n} x'_i u_i\right)
\]

\[
= \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x'_i x_i\right)^{-1}\left(\frac{1}{n} \sum_{i=1}^{n} x'_i u_i\right)
\]

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

\[
\left(\frac{1}{n} \sum_{i=1}^{n} x'_i x_i\right) \xrightarrow{p} E[x'_i x_i] \equiv Q_{(k+1) \times (k+1)} \quad \text{(a positive definite matrix \(Q\) exist)}
\]

\[
\left(\frac{1}{n} \sum_{i=1}^{n} x'_i u_i\right) \xrightarrow{p} E[x'_i u_i] = 0 \quad \text{(zero conditional mean)}
\]

We have \(\text{plim}(\hat{\beta}) = \beta + Q^{-1} \cdot 0 = \beta\). 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 $x_i$ exists.

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

7.3 Asymptotic Normality

$$\hat{\beta} = \beta + \left( \frac{1}{n} \sum_{i=1}^{n} x'_i x_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} x'_i u_i \right)$$
(linearity and no perfect collinearity)

$$\sqrt{n} (\hat{\beta} - \beta) = \left( \frac{1}{n} \sum_{i=1}^{n} x'_i x_i \right)^{-1} \left[ \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} x'_i u_i \right) \right]$$

$$E[x'_i x_i] = Q$$
$$E[x'_i u_i] = 0 \quad \text{(zero conditional mean)}$$

$$\text{Var}(x'_i u_i) = E_x[\text{Var}(x'_i u_i|x_i)]$$
$$= E_x[E[x'_i u_i | x_i'] x_i']$$
$$= \sigma^2 E_x[x'_i x_i] \quad \text{(homoskedasticity)}$$
$$= \sigma^2 \quad \text{(a positive definite matrix $Q$ exist)}$$

$$E[Q^{-1} x'_i u_i] = 0$$
$$\text{Var}(Q^{-1} x'_i u_i) = \sigma^2 Q^{-1} Q Q^{-1}$$
$$= \sigma^2 Q^{-1}$$

Applying the LLN to the sample mean of $x'_i x_i$ (random sampling):

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

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

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

Finally, applying the Slutsky’s Theorem:

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

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 $x_i$ exists and (6) homoskedasticity.