ECON 640 Ordinary Least Squares (Univariate)

Dr. Yang Liang

1 Derive OLS Estimator

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

```
convert a minimization problem into a maximization problem, (my habit)
```

$$\max_{\hat{\beta}_0,\hat{\beta}_1} - \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

First-Order Conditions (F.O.C) with respect to (w.r.t.) $\hat{\beta}_0$:

$$\begin{aligned}
-2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)(-1) &= 0 \\
\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) &= 0 \\
\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\beta}_0 - \sum_{i=1}^{n}\hat{\beta}_1 x_i &= 0 \\
n\bar{y} - n\hat{\beta}_0 - \hat{\beta}_1 n\bar{x} &= 0 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} \qquad (1)
\end{aligned}$$

F.O.C w.r.t $\hat{\beta}_1$:

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

$$\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n}\left(x_i y_i - \hat{\beta}_0 x_i - \hat{\beta}_1 x_i^2\right) = 0$$

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0 \quad \text{substitute} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i + \hat{\beta}_1\bar{x}\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0$$

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

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

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

2 Assumptions of OLS Estimator

2.1 6 Assumptions

OLS estimation relies on several key assumptions to ensure unbiased, efficient, and consistent estimates. These assumptions are:

1. **Linearity:** The relationship between the independent variables and the dependent variable is linear. $E(Y|u) = \mathbf{X}\beta$

2. **No Perfect Multicollinearity**: full rank of $\mathbf{X}$, $rank(X) = k$ if k variables, so $\mathbf{X}^\mathbf{T}\mathbf{X}$ is invertible, intuition: no redundant information in X.

3. ***Strict Exogeneity***: $\mathbb{E}(u|X) = 0$; $\mathbb{E}(u_i|X_i) = 0$

4. Homoskedasticity: The variance of the error term is constant across observations, $Var(u_i|X) = \sigma^2$

5. No Autocorrelation / No Serial Correlation; $Cov(u_i, u_j) = 0$, for any $i \neq j$

   - Covariance-Variance Matrix ($Var(\mathbf{u})$): $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identify matrix.

6. i.i.d and Normality of the error term, $u_i \sim N(0, \sigma^2)$, this assumption is not not strictly necessary, but super helpful for hypothesis testing and deriving asymptotic distribution.
   fun fact: The theorem was named after Carl Friedrich Gauss and Andrey Markov, although Gauss' work significantly predates Markov's. But while Gauss derived the result under the assumption of independence and normality, Markov reduced these assumptions and only used the above 5.

2.2 Gauss-Markov Theorem:

The Gauss-Markov Theorem states that under assumptions 1-5, the OLS estimator is the **Best Linear Unbiased Estimator (BLUE)**. Specifically, the OLS estimator has the smallest variance among all linear and unbiased estimators.

3 Statistical Properties of OLS

3.1 Unbiasedness of $\hat{\beta}$

Recall eqn (1) *and* (2),

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

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

$$\mathbb{E}(\hat{\beta}_1) = \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$= \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \epsilon_i - \beta_0 - \beta_1\bar{x} - \bar{\epsilon})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$= \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_1(x_i - \bar{x}) + (\epsilon_i - \bar{\epsilon}))}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$= \mathbb{E}\left(\beta_1 \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

recall the previous algebra trick

$$= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})\epsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

use LIE and strict exogeneity assumption

$$= \beta_1 + \mathbb{E}\left(\mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})\epsilon_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \mid \mathbf{X}\right)\right)$$

$$= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\mathbb{E}\left(\epsilon_i \mid \mathbf{X}\right)\right)$$

$$= \beta_1 + \mathbb{E}\left(\frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\left(\mathbf{0}\right)\right)$$

$$= \beta_1$$

then,

$$\mathbb{E}(\hat{\beta}_0) = \mathbb{E}(\bar{y} - \hat{\beta}_1\bar{x})$$

$$= \mathbb{E}(\bar{y}) - \mathbb{E}(\hat{\beta}_1\bar{x})$$

$$= \mathbb{E}(\beta_0 + \beta_1\bar{x}) - \mathbb{E}(\hat{\beta}_1\bar{x}) = \mathbb{E}(\beta_0) + \mathbb{E}(\beta_1\bar{x}) - \bar{x}\mathbb{E}(\hat{\beta}_1) = \beta_0$$

Q.E.D

3.2 Variance of $\hat{\beta}_1$

(drop 1 subscript for the ease of type)

$$Var(\hat{\beta}) = Var\left(\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

Recall again:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}) = \sum_{i=1}^{n} Y_i(X_i - \bar{X})$$

$$Var(\hat{\beta}) = Var\left(\frac{\sum_{i=1}^{n} Y_i(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

Note: The denominator of the expression is a constant conditioning on X, and therefore, it will get squared when we take it out of variance expression.

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} Var\left(\sum_{i=1}^{n} Y_i(X_i - \bar{X})\right)$$

Substituting the value of Y_i,

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} Var\left(\sum_{i=1}^{n}(\alpha + \beta X_i + \epsilon_i)(X_i - \bar{X})\right)$$

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} Var\left(\sum_{i=1}^{n}(\alpha + \beta X_i)(X_i - \bar{X}) + \sum_{i=1}^{n}\epsilon_i(X_i - \bar{X})\right)$$

The above first term in the parenthesis is a constant conditioning on X. So, $Var(\sum_{i=1}^{n}(\alpha + \beta X_i)(X_i - \bar{X})) = 0$. Thus, we arrive at the following equation:

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} Var\left(\sum_{i=1}^{n}\epsilon_i(X_i - \bar{X})\right)$$

Now we shall also use the assumption that $Cov(\epsilon_i, \epsilon_j) = 0$ (For i not equal to j). So, we get:

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} \left(\sum_{i=1}^{n}(X_i - \bar{X})^2 Var(\epsilon_i)\right)$$

By homoskedasticity assumption, $Var(\epsilon_i) = \sigma^2$ (a constant). So we get:

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{1}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} \left(\sum_{i=1}^{n}(X_i - \bar{X})^2\sigma^2\right)$$

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{\sigma^2}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} \sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$Var(\hat{\beta} \mid \mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad \text{(Q.E.D.)}$$

Thus, after intensive mathematics, we got the variance of $\hat{\beta}_1$

3.3 Variance of $\hat{\beta}_0$

Calculating this with scalar notation is extremely tedious, and doesn't gain much insights, so I'll provide the solution directly.

Table 1: The following table sums up all the main final equations we derived:

	$\hat{\beta}_1$	$\hat{\beta}_1$
Expectation: $\mathbb{E}(\hat{\beta} \mid \mathbf{X})$	β_0	β_1
Variance: $\sigma^2_{\hat{\beta}} \mid \mathbf{X}$	$\sigma^2_{\beta_1} = \sigma^2 \cdot \frac{(X_i)^2}{n\sum_{i=1}^n (X_i - \bar{X})^2}$	$\sigma^2_{\beta_1} = \sigma^2 \cdot \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}$

3.4 Estimating σ^2

notice, the variance of beta-hat is depending on σ^2 which is still unknown.

Let,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

2: the loss of d.o.f, if k regressions, d.o.f=n-k

Theorem 1

$$\mathbb{E}(\hat{\sigma}^2) = Var(u_i) = \sigma^2 \qquad \text{(under homoskedasticity)}$$

Theorem 2 if we add normality (Gauss's original assumption) and conditional on X then

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma^2_{\beta_1}\right)$$

or if more general,

$$\hat{\beta}_k \sim N\left(\beta_k, \sigma^2_{\beta_k}\right), k = 1, 2 \qquad \text{(under normality)}$$

and the two are independent. Then, let's plug in $\hat{\sigma}^2$ to estimate the standard errors of $\hat{\beta}$,

$$\hat{\sigma}^2_{\beta_1} = \hat{\sigma}^2 \cdot \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Theorem 3

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\beta_k}} \sim t_{n-2}, k = 1, 2 \qquad \text{(normalized distribution = t-distribution)}$$

implies that 95% confidence intervals for β_k can be constructed by:

$$\hat{\beta}_k \pm t_{n-k,0.975} \cdot \hat{\sigma}_{\beta_k} \tag{1}$$

where you can look up the value of $t_{n-k,0.975}$ in a t-distribution.
Similar results apply approximately in large samples under much weaker assumptions. However, when

$$Var(u_i|\mathbf{X}) = \sigma^2(\mathbf{X})$$

(meaning not a constant) then we have to correct $\hat{\sigma}_{\beta_k}$

3.5 Asymptotic Distribution of OLS

Notice, $\hat{\beta}_1$ collapses to a number (because $Var(\hat{\beta}_1) = \sigma^2 \cdot \frac{1}{n \cdot \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$), it has no distribution! It is $O(1/N)$

Notation for consistency:

$$\begin{aligned} plim(\hat{\beta}) &= \beta; \\ \hat{\beta} &\xrightarrow{p} \beta; \\ \frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\beta_k}} \sim t_{n-k} &\xrightarrow{d} N(0,1) \end{aligned}$$