Simple Regression 2

Class 5

Wonmun Shin

(wonmun.shin@sejong.ac.kr)

Department of Economics, Sejong University

* This lecture note is written based on Professor Chang Sik Kim's lecture notes.

Classical Assumptions

Classical Assumptions

- **(A1)** Independent variable (X) is not random but deterministic.
- **(A2)** $E\left(e_i\right) = 0$
- **(A3)** $Var\left(e_i\right) = \sigma^2$ *Homoskedasticity*
- **(A4)** $Cov\left(e_i,\, e_j\right) = 0$ for $i \neq j$ *No Autocorrelation*

Classical Assumptions [cont'd]

- **(A1)** Independent variable (X) is not random but deterministic.
  - This assumption is clearly not realistic in economic model, but here we just assume that for mathematical simplicity.
  - However, even when X is random, most results will be the same if we assume **(A2)-(A4)** in a conditional version given X.
- **(A2)** $E(e_i) = 0$
  - Recall, we have already learned this assumption, $E(e_i \mid X_i) = 0$.

$$E(e_i \mid X_i) = 0 \;\Leftrightarrow\; E(Y_i \mid X_i) = \beta_1 + \beta_2 X_i$$

  - Likewise,

$$E(e_i) = 0 \;\Leftrightarrow\; E(Y_i) = \beta_1 + \beta_2 X_i$$

Classical Assumptions [cont'd]

- **(A3)** $Var(e_i) = \sigma^2$ for $i = 1, 2, \cdots, n$
  - For simplicity, we just assume equal variance of all error e_i, which will be relaxed later in this course.
  - Homeskedasticity: $\underbrace{homo}_{\text{same}} + \underbrace{skedasis}_{\text{dispersion}}$
    - Some materials use "homoscedasticity" (with same pronunciation).
- **(A4)** $Cov(e_i, e_j) = 0$ for $i \neq j$
  - For simplicity, we just assume no covariance between different error terms.
  - Hopefully, we will discuss the cases with autocorrelation in the later part of this course. (If we do not have time to learn it, it will be discussed in the "Advanced Econometrics" in the next semester!)

Properties of OLS Estimators

Recall: OLS Estimators

$$Y_i = \beta_1 + \beta_2 X_i + e_i$$

$$\hat{\beta}_2 = \frac{\sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum \left(X_i - \bar{X}\right)^2} = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Unbiasedness

1. Unbiasedness

$$E\left(\hat{\beta}_2\right) = \beta_2$$

$$E\left(\hat{\beta}_1\right) = \beta_1$$

Proof of unbiasedness of $\hat{\beta}_2$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum \underbrace{\left(\frac{x_i}{\sum x_i^2} \right)}_{\equiv \omega_i} Y_i$$

$$= \sum \omega_i Y_i \quad \textbf{(Linear Estimator)}$$

Unbiasedness [cont'd]

- (i) $\sum \omega_i = 0$
  - Why? $\sum \omega_i = \sum \left(\frac{x_i}{\sum x_i^2} \right) = \frac{\sum x_i}{\sum x_i^2} = 0$
- (ii) $\sum \omega_i X_i = 1$
  - Why? $\sum \omega_i X_i = \sum \left(\frac{x_i}{\sum x_i^2} \right) X_i = \frac{\sum x_i X_i}{\sum x_i^2} = 1$
- (iii) $\sum \omega_i^2 = \frac{1}{\sum x_i^2}$
  - Why? $\sum \omega_i^2 = \sum \left(\frac{x_i}{\sum x_i^2} \right)^2 = \frac{\sum x_i^2}{\sum x_i^2 \sum x_i^2} = \frac{1}{\sum x_i^2}$

Unbiasedness [cont'd]

$$\begin{aligned}
\hat{\beta}_2 &= \sum \omega_i Y_i \\
&= \sum \omega_i \left(\beta_1 + \beta_2 X_i + e_i\right) \\
&= \beta_1 \underbrace{\sum \omega_i}_{=0} + \beta_2 \underbrace{\sum \omega_i X_i}_{=1} + \sum \omega_i e_i \\
&= \beta_2 + \sum \omega_i e_i
\end{aligned}$$

$$\begin{aligned}
\therefore E\left(\hat{\beta}_2\right) &= \beta_2 + E\left(\sum \omega_i e_i\right) \\
&= \beta_2 + \sum \omega_i E\left(e_i\right) \\
&= \beta_2
\end{aligned}$$

Unbiasedness [cont'd]

Proof of unbiasedness of $\hat{\beta}_1$

$$\begin{aligned}
\hat{\beta}_1 &= \bar{Y} - \hat{\beta}_2 \bar{X} \\
&= \bar{Y} - \bar{X} \cdot \sum \omega_i Y_i \\
&= \frac{1}{n} \sum Y_i - \bar{X} \cdot \sum \omega_i Y_i \\
&= \sum \underbrace{\left(\frac{1}{n} - \bar{X} \omega_i \right)}_{\equiv \bar{\omega}_i} Y_i \\
&= \sum \bar{\omega}_i Y_i \ \textbf{(Linear Estimator)}
\end{aligned}$$

Unbiasedness [cont'd]

- (i) $\sum \bar{\omega}_i = 1$
  - Why? $\sum \bar{\omega}_i = \sum \left(\frac{1}{n} - \bar{X}\omega_i\right) = \sum \frac{1}{n} - \bar{X}\sum \omega_i = 1$
- (ii) $\sum \bar{\omega}_i X_i = 0$
  - Why? $\sum \bar{\omega}_i X_i = \sum \left(\frac{1}{n} - \bar{X}\omega_i\right) X_i = \frac{1}{n}\sum X_i - \bar{X}\sum \omega_i X_i = \bar{X} - \bar{X} = 0$
- (iii) $\sum \bar{\omega}_i^2 = \frac{1}{n}\frac{\sum X_i^2}{\sum x_i^2}$
  - Why?

$$\begin{aligned}
\sum \bar{\omega}_i^2 &= \sum \left(\frac{1}{n} - \bar{X}\omega_i\right)^2 = \sum \left(\frac{1}{n^2} - \frac{2}{n}\bar{X}\omega_i + \bar{X}^2\omega_i^2\right) = \frac{1}{n} - \frac{2}{n}\bar{X}\sum \omega_i + \bar{X}^2 \sum \omega_i^2 \\
&= \frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} = \frac{1}{n} + \frac{1}{n^2}\frac{\left(\sum X_i\right)^2}{\sum x_i^2} = \frac{1}{n}\left(1 + \frac{1}{n}\frac{\left(\sum X_i\right)^2}{\sum x_i^2}\right) = \frac{1}{n}\left(\frac{\sum x_i^2 + \frac{1}{n}\left(\sum X_i\right)^2}{\sum x_i^2}\right) \\
&= \frac{1}{n}\frac{\sum X_i^2}{\sum x_i^2}
\end{aligned}$$

Unbiasedness [cont'd]

$$\begin{aligned}\hat{\beta}_1 &= \sum \bar{\omega}_i Y_i \\ &= \sum \bar{\omega}_i \left(\beta_1 + \beta_2 X_i + e_i\right) \\ &= \beta_1 \underbrace{\sum \bar{\omega}_i}_{=1} + \beta_2 \underbrace{\sum \bar{\omega}_i X_i}_{=0} + \sum \bar{\omega}_i e_i \\ &= \beta_1 + \sum \bar{\omega}_i e_i\end{aligned}$$

$$\begin{aligned}\therefore E\left(\hat{\beta}_1\right) &= \beta_1 + E\left(\sum \bar{\omega}_i e_i\right) \\ &= \beta_1 + \sum \bar{\omega}_i E\left(e_i\right) \\ &= \beta_1\end{aligned}$$

2. Efficiency

$$\begin{aligned} Var\left(\hat{\beta}_2\right) &= Var\left(\hat{\beta}_2 - \beta_2\right) \\ &= Var\left(\sum \omega_i e_i\right) \\ &= Var\left(\omega_1 e_1 + \omega_2 e_2 + \cdots + \omega_n e_n\right) \\ &= \omega_1^2 Var\left(e_1\right) + \omega_2^2 Var\left(e_2\right) + \cdots + \omega_n^2 Var\left(e_n\right) + Covariances \\ &= \left(\omega_1^2 + \omega_2^2 + \cdots + \omega_n^2\right)\sigma^2 \\ &= \sigma^2 \sum \omega_i^2 = \frac{\sigma^2}{\sum x_i^2} \end{aligned}$$

$$\text{Likewise, } \begin{aligned} Var\left(\hat{\beta}_1\right) &= Var\left(\hat{\beta}_1 - \beta_1\right) \\ &= Var\left(\sum \bar{\omega}_i e_i\right) \\ &= \sigma^2 \sum \bar{\omega}_i^2 = \frac{\sigma^2}{n}\frac{\sum X_i^2}{\sum x_i^2} \end{aligned}$$

Efficiency [cont'd]

- The greater the variance σ^2 is, the greater the variances of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are.
  - Note that $\sigma^2 = Var(e_i) = Var(Y_i)$
  - That is, if the uncertainty about the value of Y will be greater, then the relationship between Y_i and $\beta_1 + \beta_2 X_i$ will be weaker.
  - Clearly, the better information (data), the better your estimates will be (the smaller variances).

- Obviously, if the number of observations (n) is larger, the variances of the estimators are generally smaller.

- $\sum x_i^2 \left(= \sum (X_i - \bar{X})^2 \right)$ is inversely related to the variances of the OLS estimators.
  - That is, the more variation (more information to estimate the regression model) you have, the better (reliable) estimates you get.

Efficiency [cont'd]

- **Gauss-Markov Theorem (GMT)**
  - Given the classical assumptions, the OLS estimators are **BLUE (Best Linear Unbiased Estimator)**.
  - That is, they are **most efficient** in the class of linear unbiased estimators.
    - GMT says that the OLS estimators have minimum variance among linear and unbiased estimators. *(proof: Gujarati Ch. 3 Appendix 3A.6)*
    - It does not say that OLS estimators are the best of all possible estimators.
  - GMT requires (A2)-(A4). It does not depend on any distributional assumptions like normality.
    - Normality assumption $\rightarrow$ OLS estimators are **MVUE (Minimum Variance Unbiased Estimator).**

Consistency

3. Consistency

- Recall, $\hat{\theta}$ is a consistent estimator of θ

$$\iff \hat{\theta} \longrightarrow \theta \quad \left(\text{or } \lim_{n\to\infty} \hat{\theta} = \theta\right)$$

$$\Longleftarrow \lim_{n\to\infty} MSE\left(\hat{\theta}\right) = 0$$

$$\iff \lim_{n\to\infty} Bias\left(\hat{\theta}\right) = 0, \ \lim_{n\to\infty} Var\left(\hat{\theta}\right) = 0$$

- $\lim_{n\to\infty} Bias\left(\hat{\beta}_2\right) = 0, \quad \lim_{n\to\infty} Bias\left(\hat{\beta}_1\right) = 0$
- $\lim_{n\to\infty} Var\left(\hat{\beta}_2\right) = \lim_{n\to\infty} \frac{\sigma^2}{\Sigma x_i^2} = 0$
- $\lim_{n\to\infty} Var\left(\hat{\beta}_1\right) = \lim_{n\to\infty} \frac{\sigma^2}{n} \frac{\sum X_i^2}{\sum x_i^2} = 0$

Variance of Error Term

Variance of Error Term

- Another unknown parameter of the simple regression model is: σ^2
  - We need to estimate σ^2 to get the variance of OLS estimators.

$$Var\left(e_i\right) = E\left(e_i^2\right) - \left[E\left(e_i\right)\right]^2 = E\left(e_i^2\right) \rightarrow \frac{\sum e_i^2}{n}$$

- But, e_i is unobservable, so we should use observable $\hat{e}_i$.

$$\therefore \hat{\sigma^2} = \frac{\sum \hat{e}_i^2}{n}$$

  - However, the above estimator is not unbiased. So we modify it to produce an unbiased estimator. *(proof: Gujarati Ch. 3 Appendix 3A.5)*

$$\hat{\sigma^2} = \frac{\sum \hat{e}_i^2}{n-2}$$

- <u>Note</u> $\hat{\sigma^2}$ is an unbiased and consistent estimator.

Variance of Error Term

- Then, we can get the estimates of variances of OLS estimators.

$$\widehat{Var\left(\hat{\beta_2}\right)} = \frac{\hat{\sigma^2}}{\sum x_i^2}$$

$$\widehat{Var\left(\hat{\beta_1}\right)} = \frac{\hat{\sigma^2}}{n}\frac{\sum X_i^2}{\sum x_i^2}$$