Econometrics II TA Session #1

Ryo Sakamoto

Oct 13, 2021

Contents

1. derivation and small sample properties of the OLS estimator
    - mean (unbiasedness)
    - variance (efficiency, Gauss-Marikov theorem)
    - distribution
2. large sample properties of the OLS estimator
    - consistency
    - asymptotic normality
3. appendix

1. Derivation and Small Sample Properties of OLSE
2. Large Sample Properties of OLSE
3. Appendix

Multiple Regression Model

- Consider the following regression model:

$$y_i = \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i = x_i \beta + u_i,$$

for $i = 1, \cdots, n$, where $u_1, \cdots, u_n$ are assumed to be mutually independently and indentically distributed with mean zero and variance σ^2,

$$x_i = (x_{i1}, \cdots, x_{ik}) \in \mathbb{R}^k$$

and

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \in \mathbb{R}^k.$$

Multiple Regression Model

- The stacked model is given by

$$y = X\beta + u,$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n\times k}(\mathbb{R}), \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^n.$$

- Throughout this section, we assume that the explanatory variables $x_i, \ i = 1, \cdots, k$ are fixed.

Derivation of the OLS Estimator

- Consider the following objective function:

$$S(\beta) = (y - X\beta)'(y - X\beta).$$

- The OLS estimator is

$$\hat{\beta} = \arg\min_{\beta} \; S(\beta).$$

Derivation of the OLS Estimator

- The F.O.C. is

$$\frac{\partial S(\hat{\beta})}{\partial \beta} = -2X'y + 2X'X\hat{\beta} = 0.$$

- Solving the equation above, we have the OLS estimator:

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

Derivation of the OLS Estimator

- The S.O.C. is that

$$\frac{\partial^2 S(\hat{\beta})}{\partial\beta\partial\beta'}$$

  must be a positive definite matrix.

- The differential coefficient is $2X'X \in \mathcal{M}_{k\times k}(\mathbb{R})$ which is positive definite since for any vector $a \in \mathbb{R}^k$ such that $a \neq 0$,

$$a'(X'X)a = (Xa)'Xa = z'z = \sum_{j=1}^{n} z_j^2 > 0,$$

  where $z := Xa$.

Mean of the OLS Estimator (Unbiasedness)

- To obtain the properties of the OLS estimator, we rewrite (1) as follows:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + u) \\ &= \beta + (X'X)^{-1}X'u. \end{aligned} \tag{2}$$

- Taking the expectation on both sides of (2), we have

$$\begin{aligned} E(\hat{\beta}) &= E[\beta + (X'X)^{-1}X'u] \\ &= \beta + (X'X)^{-1}X'E(u) = \beta, \end{aligned}$$

because of $E(u) = 0$ by the assumption.

Variance of the OLS Estimator

- The variance of the OLS estimator is

$$
\begin{aligned}
V(\hat{\beta}) &= E\Big[(\hat{\beta}-\beta)(\hat{\beta}-\beta)\Big] \\
&= E\Big\{(X'X)^{-1}(X'u)[(X'X)^{-1}(X'u)]'\Big\} \\
&= E\Big[(X'X)^{-1}X'uu'X(X'X)^{-1}\Big] \\
&= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}.
\end{aligned}
$$

- The fifth equality comes from the assumption that u_i is mutually independently and identically distributed with mean zero and variance σ^2, which implies that $E(u_i^2)=\sigma^2,\ \forall i$ and $E(u_iu_j)=0,\ \forall i\neq j$.

Gauss-Markov Theorem

- The OLS estimator is a BLUE (best linear unbiased estimator), i.e., minimum variance within the class of linear unbiased estimators.

Proof

- Consider another linear unbiased estimator $\tilde{\beta}$,

$$\tilde{\beta} = Cy = C(X\beta + u) = CX\beta + Cu, \tag{3}$$

  where $C \in \mathcal{M}_{k \times n}(\mathbb{R})$.

- Taking the expectation of $\tilde{\beta}$, we obtain:

$$\begin{aligned} E(\tilde{\beta}) &= CX\beta + CE(u) \\ &= CX\beta. \end{aligned}$$

- Since $\tilde{\beta}$ is unbiased, $E(\tilde{\beta}) = \beta$ must be met, which implies

$$CX = I_k.$$

- Substituting this into (3), we have

$$\tilde{\beta} = CX\beta + Cu = \beta + Cu.$$

- The variance of $\tilde{\beta}$ is

$$\begin{aligned} V(\tilde{\beta}) &= E\left[(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'\right] \\ &= E(Cuu'C') \\ &= CE(uu')C' \\ &= \sigma^2 CC'. \end{aligned}$$

- Now, let $C = D + (X'X)^{-1}X'$, then

$$\begin{aligned} V(\tilde{\beta}) &= \sigma^2 CC' \\ &= \sigma^2 \left[D + (X'X)^{-1}X'\right]\left[D + (X'X)^{-1}X'\right]' \end{aligned}$$

- Calculating CX, we have

$$\begin{aligned} CX &= \left[D + (X'X)^{-1}X'\right]X \\ &= DX + I_k. \end{aligned}$$

- Since $CX = I_k$ must be met ($\because$ unbiasedness of $\tilde{\beta}$), we have the following condition:

$$DX = 0 \in \mathcal{M}_{k\times k}(\mathbb{R}).$$

- Substituting this condition into the variance of $\tilde{\beta}$, we have

$$
\begin{aligned}
V(\tilde{\beta}) &= \sigma^2 \left[D + (X'X)^{-1}X'\right]\left[D + (X'X)^{-1}X'\right]' \\
&= \sigma^2 (X'X)^{-1} + \sigma^2 DD' + \sigma^2 (X'X)^{-1}(DX)' + \sigma^2 DX(X'X)^{-1} \\
&= V(\hat{\beta}) + \sigma^2 DD'.
\end{aligned}
$$

- Thus, $V(\tilde{\beta}) - V(\hat{\beta})$ is a positive definite matrix.
- This implies $V(\tilde{\beta}_j) - V(\hat{\beta}_j) > 0$ for all $j \in \{1, \cdots, k\}$.
- Therefore, the OLS estimator $\hat{\beta}$ is a minimum variance (i.e., best) linear unbiased estimator of β. (QED)

Distribution of the OLS Estimator

- So far, u_i has not been assumed to follow a normal distribution.
- However, we need the normality assumption to obtain the distribution of the OLS estimator in the small sample.
- Then, hereafter we assume

$$u_i \sim \mathcal{N}_{\mathbb{R}}(0, \sigma^2) \Rightarrow u \sim \mathcal{N}_{\mathbb{R}^n}\left(0, \sigma^2 I_n\right).$$

- Under normality assumption on the error term u, it is known that the distribution of the OLS estimator is

$$\hat{\beta} \sim \mathcal{N}_{\mathbb{R}^k}\left(\beta, \sigma^2 (X'X)^{-1}\right).$$

Proof

- The moment generating function of u is

$$\begin{aligned}\phi_u(\theta_u) &:= E\left[\exp(\theta' X)\right] \\ &= E\left(\frac{1}{2}\theta_u' \sigma^2 I_n \theta_u\right) \\ &= E\left(\frac{\sigma^2}{2}\theta_u' \theta_u\right),\end{aligned}$$

which comes from $u \sim \mathcal{N}_{\mathbb{R}^n}(0, \sigma^2 I_n)$.

- The moment generating function of $\hat{\beta}$ is

$$
\begin{aligned}
\phi_\beta(\theta_\beta) &:= E\left[\exp(\theta_\beta'\hat{\beta})\right] \\
&= E\left\{\exp\left[\theta_\beta'\beta + \theta_\beta'(X'X)^{-1}X'u\right]\right\} \\
&= \exp(\theta_\beta'\beta)\cdot E\left\{\exp\left[\theta_\beta'(X'X)^{-1}X'u\right]\right\} \\
&= \exp(\theta_\beta'\beta)\cdot \phi_u\left(\theta_\beta'(X'X)^{-1}X'\right) \\
&= \exp(\theta_\beta'\beta)\cdot \exp\left\{\frac{\sigma^2}{2}\left[\theta_\beta'(X'X)^{-1}X'\right]'\left[\theta_\beta'(X'X)^{-1}X'\right]\right\} \\
&= \exp(\theta_\beta'\beta)\cdot \exp\left[\frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta\right]
\end{aligned}
$$

$$\begin{aligned}\phi_\beta(\theta_\beta) &= \exp(\theta_\beta'\beta)\cdot\exp\left[\frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta\right]\\ &= \exp\left[\theta_\beta'\beta+\frac{1}{2}\theta_\beta'\sigma^2(X'X)^{-1}\theta_\beta\right].\end{aligned}$$

- This is equivalent to the normal distribution with mean β and variance $\sigma^2(X'X)^{-1}$. (QED)

Distribution of the OLS Estimator

- Taking the jth element of $\hat{\beta}$, its distribution is

$$\hat{\beta}_j \sim \mathcal{N}_{\mathbb{R}}\left(\beta_j, \sigma^2 (X'X)^{-1}_{jj}\right) \iff \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 (X'X)^{-1}_{jj}}} \sim \mathcal{N}_{\mathbb{R}}(0,1),$$

  where $(X'X)^{-1}_{jj}$ denotes the jth diagonal element of $(X'X)^{-1}$.

- Replacing σ^2 by its estimator s^2, we have the following t distribution:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2 (X'X)^{-1}_{jj}}} \sim t(n-k),$$

  where $t(n-k)$ denotes the t distribution with $n-k$ degrees of freedom.

Distribution of the OLS Estimator

- s^2 is taken as follows:

$$s^2 = \frac{1}{n-k}(y - X\hat{\beta})'(y - X\hat{\beta}),$$

which is an unbiased estimator of σ^2.

Properties of OLS

- So far, we have looked at
  - how to derive the OLS estimator
  - small sample properties
    - how to derive the mean, variance and distribution of the OLS estimator
    - the Gauss-Markov theorem (efficiency)
- In what follows, we will focus on the large sample properties of OLS estimator:
  - consistency
  - asymptotic normality

Consistency and Asymptotic Normality

- In Appendix, we can review the asymptotic theory.
  - Convergence in distribution
  - Convergence in probability
  - Chebyshev's inequality
  - Slutsky's theorem
  - Central limit theorem

Consistency

- In what follows, we assume that the explanatory variables are random.
- In addition, we impose the exogeneity assumption $E[u|X] = 0$.
- The error term is assumed to be

$$u_i|x_i \sim \left(0, \sigma^2\right) \Rightarrow u|X \sim \left(0, \sigma^2 I_n\right).$$

- Note that we do not need the normality assumption.
- The OLS estimator is a consistent estimator, i.e., $\hat{\beta} \xrightarrow[n \to \infty]{p} \beta$.

Proof

- The OLS estimator is

$$
\begin{aligned}
\hat{\beta} &= \beta + (X'X)^{-1}X'u \\
&= \beta + \left(\frac{1}{n}X'X\right)^{-1}\frac{1}{n}X'u \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} x_i' u_i.
\end{aligned}
$$

Proof

- By the law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i \xrightarrow[n\to\infty]{p} E[x_i' x_i] =: M_{xx} \in \mathcal{M}_{k\times k}(\mathbb{R}),$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i' u_i \xrightarrow[n\to\infty]{p} E[x_i' u_i].$$

- Using the law of iterated expectation,

$$E[x_i' u_i] = E[x_i' E(u_i | x_i)] = 0.$$

Proof

- By the Slutsky theorem, we have

$$\hat{\beta} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i' u_i$$

$$\xrightarrow[n \to \infty]{p} \beta + M_{xx}^{-1} \cdot 0 = \beta,$$

which concludes the proof. (QED)

Asymptotic Normality

- The asymptotic normality of the OLS estimator means

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \to \infty]{d} \mathcal{N}_{\mathbb{R}^n}\left(0, \sigma^2 M_{xx}^{-1}\right).$$

Proof

- Rewriting the expression of the OLS estimator yields

$$\begin{aligned} \hat{\beta} &= \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i' u_i \\ \iff \hat{\beta} - \beta &= \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} x_i' u_i \\ \iff \sqrt{n}(\hat{\beta} - \beta) &= \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \frac{\sqrt{n}}{n} \sum_{i=1}^{n} x_i' u_i . \end{aligned}$$

- By the law of large numbers and the Slutsky's theorem, we have

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \xrightarrow[n\to\infty]{p} M_{xx}^{-1}.$$

- By the central limit theorem, we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i' u_i\right) \xrightarrow[n\to\infty]{d} \mathcal{N}_{\mathbb{R}^k}\Big(0, V(x_i' u_i)\Big).$$

$$\begin{aligned} V(x_i' u_i) &= E\left[V(x_i' u_i | x_i)\right] + V\left[E(x_i' u_i | x_i)\right] \\ &= E\left[x_i' V(u_i | x_i) x_i\right] \\ &= \sigma^2 E\left[x_i' x_i\right] \\ &= \sigma^2 M_{xx}. \end{aligned}$$

- Then, we have

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i' u_i\right) \xrightarrow[n\to\infty]{d} \mathcal{N}_{\mathbb{R}^k}\left(0, \sigma^2 M_{xx}\right).$$

- By the Slutsky's theorem, we have

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i' u_i\right) \xrightarrow[n\to\infty]{d} M_{xx}^{-1} \cdot B,$$

where

$$B \sim \mathcal{N}_{\mathbb{R}^k}\left(0, \sigma^2 M_{xx}\right).$$

- Note that

$$B \sim \mathcal{N}_{\mathbb{R}^k}\left(0, \sigma^2 M_{xx}\right) \Rightarrow M_{xx}^{-1} \cdot B \sim \mathcal{N}_{\mathbb{R}^k}\left(0, \sigma^2 M_{xx}^{-1}\right).$$

- Hence, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \to \infty]{d} \mathcal{N}_{\mathbb{R}^k}\left(0, \sigma^2 M_{xx}^{-1}\right).$$

which concludes the proof. (QED)

Moment Generating Function (積率母関数)

- X is a random variable and $X \sim \mathcal{N}_{\mathbb{R}}(\mu,\ \sigma^2)$.
- Then, the moment generating function is given by

$$
\begin{aligned}
M(\theta) &:= E[\exp(\theta X)] \\
&= \exp\left(\mu\theta + \frac{1}{2}\sigma^2\theta\right).
\end{aligned}
$$

Moment Generating Function (積率母関数)

- X is a random **vector** and $X \sim \mathcal{N}_{\mathbb{R}^n}(\mu, \ \Sigma)$.
- Then, the moment generating function is given by

$$\begin{aligned} \phi(\theta) &:= E[\exp(\theta' X)] \\ &= \exp\left(\theta' \mu + \frac{1}{2}\theta' \Sigma \theta\right). \end{aligned}$$

Convergence in Distribution (分布収束)

- A series of random variables $X_1, \cdots, X_n, \cdots$, have distribution functions $F_1, \cdots, F_n, \cdots$, respectively.
- If

$$\lim_{n \to \infty} F_n = F,$$

then we say that a series of random variables $X_1, X_2, \cdots$ converges to F in distribution.

Convergence in Probability (確率収束)

- Let $\{X_n\}, n = 1, 2, \cdots$ be a sequence of random variables.
- If

$$\lim_{n \to \infty} P(|X_n - \theta| < \epsilon) = 1, \ \forall \epsilon > 0,$$

then we way that $\{X_n\}$ converges to θ in probability.
- We denote as $X_n \xrightarrow[n \to \infty]{p} \theta$.

Chebyshev's Inequality

- Let $g : \mathbb{R}^n \to \mathbb{R}$.
- For $g(X) \geq 0$,

$$P\left[g(X) \geq k\right] \leq \frac{E[g(X)]}{k}$$

where k is a positive constant.

Slutsky's Theorem

- Let X_n and Y_n be random variables such that $X_n \xrightarrow[n\to\infty]{p} c$ and $Y_n \xrightarrow[n\to\infty]{p} d$.
- Then,
  1. $X_n + Y_n \xrightarrow[n\to\infty]{p} c + d$
  2. $X_n Y_n \xrightarrow[n\to\infty]{p} cd$
  3. $X_n / Y_n \xrightarrow[n\to\infty]{p} c/d$ for $d \neq 0$
  4. $g(X_n) \xrightarrow[n\to\infty]{p} g(c)$

Central Limit Theorem: Univariate case

- $X_1, \cdots, X_n \in \mathbb{R}$ are mutually independent and identically distributed as $X_i \sim (\mu, \sigma^2)$.
- Then,

$$\frac{\overline{X} - E(\overline{X})}{\sqrt{V(\overline{X})}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{d} \mathcal{N}_{\mathbb{R}}(0, 1),$$

which implies

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow[n\to\infty]{d} \mathcal{N}_{\mathbb{R}}(0, \sigma^2),$$

where $\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$.

Central Limit Theorem: Multivariate case

- $X_1, \cdots, X_n \in \mathbb{R}^k$ are mutually independently and identically distributed as $X_i \sim (\mu, \Sigma)$.
- Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow[n \to \infty]{d} \mathcal{N}_{\mathbb{R}^k}(0, \Sigma).$$