OLS as a Statistical Estimator

date: 20241007
status: #finished
tags: #econometrics #economic_research #ols

---

0. Introduction

In a parametric linear setting, we have:

$$y = X\beta + \varepsilon$$

where y and $\varepsilon \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times k}$, and $\beta \in \mathbb{R}^k$ is the parameters of interest. For $X\beta = \mathbb{E}[y|X]$ to hold, we assume $\mathbb{E}[\varepsilon|X] = 0$ where ε_i is i.i.d.

> **Corollary**
>
> *If* $\mathbb{E}[\varepsilon|X] = 0$, *then*
>
> $$\mathbb{E}[\varepsilon] = 0.$$

Normally, in this context, we often want to estimate β and $\mathrm{Var}[\varepsilon|X]$.

After deriving the closed form solution for $\hat{\beta}^{\mathrm{OLS}} = (X'X)^{-1}X'y$, which is a **statistical estimator** for β, we want to know its statistical properties, e.g., whether it is unbiased and whether it is asymptotically normal.

---

1. Unbiasedness, Consistency, and Efficiency

When we have an estimator $\hat{\beta}$ for β, we want to know how good it is, and typically there are three kinds of criterion: unbiasedness, consistency, and efficiency. Unbiasedness is to check whether $\mathbb{E}\left[\hat{\beta}\right] = \beta$, consistency aims to see whether $\hat{\beta} \xrightarrow{p} \beta$, and efficiency measures the variance of the estimator.

(i) Unbiasedness

Theorem (OLS is an unbiased estimator)

$$\mathbb{E}\left[\hat{\beta}^{\text{OLS}} \mid X\right] = \beta$$

Proof

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}^{\text{OLS}} \mid X\right] &= \mathbb{E}\left[(X'X)^{-1}X'y \mid X\right] \\
&= \mathbb{E}\left[(X'X)^{-1}X'(X\beta + \varepsilon) \mid X\right] \\
&= \mathbb{E}\left[\beta + (X'X)^{-1}X'\varepsilon \mid X\right] \\
&= \beta + (X'X)^{-1}X'\mathbb{E}[\varepsilon \mid X] \\
&= \beta
\end{aligned}$$

Corollary

$$\mathbb{E}\left[\hat{\beta}^{\text{OLS}}\right] = \beta$$

The intuition for the unbiasedness is that although we have finite sample $\{\varepsilon_i\}$ and we are not sure about whether the samples we observe are representative enough for the population, the estimator is correct on average. That is, if we draw multiple sets of $\{\varepsilon_i\}$ samples, obtaining multiple $\hat{\beta}^{\text{OLS}}$, on average, it's close to the true parameter β.

Note that as $\hat{\beta}^{\text{OLS}} = \beta + (X'X)^{-1}X'\varepsilon$, we only need the condition that $\mathbb{E}\left[(X'X)^{-1}X'\varepsilon\right] = 0$ to ensure the unbiasedness. $\mathbb{E}[\varepsilon|X] = 0$ is a stricter condition and sufficient for that.

(ii) Consistency

Theorem

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}^{\text{OLS}}|X = \beta$$

Proof

$$\begin{aligned}
\hat{\beta}^{\text{OLS}} &= \beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\varepsilon}{n}\right)
\end{aligned}$$

Note that $\frac{X'\varepsilon}{n}$ is actually a kind of mean.

$$\frac{X'\varepsilon}{n} = \frac{1}{n}[x_1, \dots, x_n]\varepsilon$$
$$= \frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i.$$

As $\{\varepsilon_i\}$ are assumed to be i.i.d., $\{x_i\varepsilon_i\}$ are i.i.d. as well, conditional on X. By WLLN,

$$\frac{X'\varepsilon}{n} \xrightarrow{p} \mathbb{E}[x_i\varepsilon_i] = 0$$

as $n \to \infty$.

Hence, $\operatorname*{plim}_{n\to\infty} \hat{\beta}^{\text{OLS}} = \beta$.

(iii) Efficiency

Theorem

$$\begin{aligned} &\operatorname{Var}\left[\hat{\beta}^{\text{OLS}} \mid X\right] \\ &= \left[(X'X)^{-1}X'\right] \operatorname{Var}[\varepsilon \mid X]\left[X(X'X)^{-1}\right]. \end{aligned}$$

Proof

$$\begin{aligned} \operatorname{Var}\left[\hat{\beta}^{\text{OLS}} \mid X\right] &= \operatorname{Var}\left[\beta + (X'X)^{-1}X'\varepsilon \mid X\right] \\ &= \operatorname{Var}\left[(X'X)^{-1}X'\varepsilon \mid X\right] \end{aligned}$$

We know $(X'X)^{-1}X'$ is a constant when the variance is conditional on X, but the question is: how can we take it out from the $\operatorname{Var}[.]$ operator? Let's see the example below:

Suppose $\mathbb{E}[X] = \mu$ and A is a constant matrix.

$$\begin{aligned} \operatorname{Var}[AX] &= \mathbb{E}[(AX - A\mu)(AX - A\mu)'] \\ &= \mathbb{E}[A(X - \mu)(X' - \mu')A'] \\ &= A \operatorname{Var}[X]A' \end{aligned}$$

Therefore,

$$\operatorname{Var}\left[(X'X)^{-1}X'\varepsilon \mid X\right] = \left[(X'X)^{-1}X'\right] \operatorname{Var}[\varepsilon|X]\left[X(X'X)^{-1}\right]$$

Corollary

If we assume $\mathrm{Var}[\varepsilon|X] = \sigma^2 I$,

$$\mathrm{Var}\left[\hat{\beta}^{\mathrm{OLS}} \mid X\right] = \sigma^2 (X'X)^{-1}.$$

Note that the assumption $\mathrm{Var}[\varepsilon|X] = \sigma^2 I$ is called homoscedasticity, and it fulfills the assumption that each sample ε_i is drawn independently and identically.

---

2. Asymptotic Normality

(i) Motivation

Once we obtain the estimator $\hat{\beta}^{\mathrm{OLS}}$, the most common first test is whether the true parameter β differs from zero:

$$H_0 : \beta = 0 \quad \text{against} \quad H_1 : \beta \neq 0.$$

To do this, we need to construct the t-statistic:

$$t = \frac{\hat{\beta}^{\mathrm{OLS}} - 0}{\widehat{\mathrm{SD}}\left[\hat{\beta}^{\mathrm{OLS}}|X\right]}.$$

The intuition is that if we can derive that $\hat{\beta}^{\mathrm{OLS}}$ is asymptotically normal, combining with a consistent estimator of $\mathrm{SD}\left[\hat{\beta}^{\mathrm{OLS}}|X\right]$, the t-statistic is then asymptotically standard normal, thereby facilitating the test.

(ii) Asymptotic Distribution

Theorem

If we assume the homoscedasticity,

$$\hat{\beta}^{\mathrm{OLS}}|X \xrightarrow{d} N\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

Proof

Recall that

$$\hat{\beta}^{\mathrm{OLS}} = \beta + \left(\frac{X'X}{n}\right)^{-1} \left(\frac{X'\varepsilon}{n}\right)$$

As $\{\varepsilon_i\}$ are i.i.d., we can apply the central limit theorem on $\frac{X'\varepsilon}{n} = \frac{1}{n}\sum_i x_i\varepsilon_i$:

$$\frac{X'\varepsilon}{n} \mid X \xrightarrow{d} N\left(\mathbb{E}\left[\frac{X'\varepsilon}{n} \mid X\right], \operatorname{Var}\left[\frac{X'\varepsilon}{n} \mid X\right]\right)$$

where

$$\mathbb{E}\left[\frac{X'\varepsilon}{n} \mid X\right] = \frac{1}{n}X'\mathbb{E}[\varepsilon|X] = 0,$$

and

$$\operatorname{Var}\left[\frac{X'\varepsilon}{n} \mid X\right] = \frac{1}{n^2}X' \operatorname{Var}[\varepsilon|X]X = \frac{\sigma^2}{n^2}(X'X).$$

Therefore,

$$\begin{aligned} \hat{\beta}^{\text{OLS}} \mid X &\xrightarrow{d} N\left(\beta, \left(\frac{X'X}{n}\right)^{-1}\frac{\sigma^2}{n^2}(X'X)\left(\frac{X'X}{n}\right)^{-1}\right) \\ &= N\left(\beta, \sigma^2(X'X)^{-1}\right) \end{aligned}$$

Basically,

$$\hat{\beta}^{\text{OLS}} \mid X \xrightarrow{d} N\left(\mathbb{E}\left[\hat{\beta}^{\text{OLS}}|X\right], \operatorname{Var}\left[\hat{\beta}^{\text{OLS}}|X\right]\right)$$

and

$$\frac{\hat{\beta}^{\text{OLS}} - \beta}{\text{SD}\left[\hat{\beta}^{\text{OLS}}|X\right]} = \frac{1}{\sigma}(X'X)^{\frac{1}{2}}\left(\hat{\beta}^{\text{OLS}} - \beta\right) \xrightarrow{d} N(0, I).$$

---

3. Consistent Estimator for the OLS Estimator's Standard Deviation

As mentioned, we need to derive the consistent estimator $\hat{\sigma}$ so that the t-statistic

$$t = \frac{\hat{\beta}^{\text{OLS}} - 0}{\widehat{\text{SD}}\left[\hat{\beta}^{\text{OLS}}|X\right]} = \frac{1}{\hat{\sigma}}(X'X)^{\frac{1}{2}}\hat{\beta}^{\text{OLS}} \xrightarrow{d} N(0, I)$$

which is ensured by the asymptotic normality and Slutsky's theorem.

For simplicity, we will derive the consistent estimator for variance, $\widehat{\sigma^2}$, and then take the square root of it to obtain the consistent estimator for standard deviation, which is guaranteed by the continuous mapping theorem.

(i) Unbiased Estimator for Variance

A possible estimator $\widehat{\sigma^2}$ is:

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}e'e$$

where $e = y - X\hat{\beta}^{\text{OLS}}$.

However, the problem is $\mathbb{E}\left[\frac{1}{n}e'e \mid X\right] \neq \sigma^2$, i.e., a biased estimator. Although we care more about the consistency, the following variance estimator is commonly used as it is not only consistent but also unbiased.

Theorem

$\frac{1}{n-k}e'e$ *is an unbiased estimator for* σ^2.

Proof

We should first connect $\mathbb{E}[e'e]$ with ε, and y is closely related to ε, so we can start from here:

$$\begin{aligned}
e &= y - X\hat{\beta}^{\text{OLS}} \\
&= y - X(X'X)^{-1}X'y \\
&= (I - P)y \\
&= My \\
&= M(X\beta + \varepsilon) \\
&= M\varepsilon
\end{aligned}$$

Note that M is the residual maker spanned from X, so $MX = 0$

$$\begin{aligned}
\mathbb{E}[e'e \mid X] &= \mathbb{E}[\varepsilon' M' M\varepsilon \mid X] \\
&= \mathbb{E}[\varepsilon' M\varepsilon \mid X] \\
&= \mathbb{E}[\text{tr}(\varepsilon' M\varepsilon) \mid X] \\
&= \mathbb{E}[\text{tr}(M\varepsilon\varepsilon') \mid X] \\
&= \text{tr}(M\mathbb{E}[\varepsilon\varepsilon' \mid X]) \\
&= \text{tr}(M)\sigma^2
\end{aligned}$$

The second equality is because M is symmetric and idempotent and the third one is due to the scalar value of $\varepsilon' M \varepsilon$ so $\varepsilon' M \varepsilon = \operatorname{tr}(\varepsilon' M \varepsilon)$. The second-last equality is because both tr and $\mathbb{E}$ are linear operators.

$$\begin{aligned}
\operatorname{tr}(M) &= \operatorname{tr}(I - P) \\
&= n - \operatorname{tr}\left(X(X'X)^{-1}X'\right) \\
&= n - \operatorname{tr}\left((X'X)^{-1}X'X\right) \\
&= n - k \\
\Rightarrow \mathbb{E}[e'e \mid X] &= \operatorname{tr}(M)\sigma^2 = (n-k)\sigma^2
\end{aligned}$$

Therefore, the unbiased estimator of σ^2 is $\frac{1}{n-k}e'e$.

(ii) Consistent Estimator for Variance

Theorem

Assuming homoscedasticity, $\frac{1}{n-k}e'e$ is a consistent estimator for σ^2:

$$\frac{1}{n-k}e'e \xrightarrow{p} \sigma^2$$

Proof

Starting again from $e = M\varepsilon$ with $M = I - P$,

$$e'e = \varepsilon' M \varepsilon = \varepsilon'\varepsilon - \varepsilon' P \varepsilon,$$

so

$$\frac{1}{n-k}e'e = \frac{n}{n-k}\left[\frac{1}{n}\varepsilon'\varepsilon - \frac{1}{n}\varepsilon' P \varepsilon\right].$$

The factor $\frac{n}{n-k} \to 1$ because k is fixed.

For the first term, $\{\varepsilon_i^2\}$ are i.i.d. with $\mathbb{E}[\varepsilon_i^2] = \sigma^2$ under homoscedasticity, so by WLLN,

$$\frac{1}{n}\varepsilon'\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 \xrightarrow{p} \sigma^2.$$

For the second term,

$$\frac{1}{n}\varepsilon' P \varepsilon = \left(\frac{X'\varepsilon}{n}\right)'\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\varepsilon}{n}\right),$$

where $\frac{X'\varepsilon}{n} \xrightarrow{p} 0$ is already established. Hence

$$\frac{1}{n}\varepsilon' P \varepsilon \xrightarrow{p} 0.$$

Therefore,

$$\frac{1}{n-k} e'e \xrightarrow{p} \sigma^2 - 0 = \sigma^2.$$

Finally, with $H_0 : \beta = 0$,

$$t = \left(\frac{1}{n-k} e'e\right)^{-\frac{1}{2}} (X'X)^{\frac{1}{2}} \hat{\beta}^{\text{OLS}} \xrightarrow{d} N(0, I).$$