Understanding OLS: Finite Sample Properties

Le Wang

Roadmap

1. Set-up and the goal
2. Assumptions
3. Under these assumptions, what are the properties of OLS estimator? (Can we achieve our goal?)

- Unbiasedness
- Variance
- MSE (combining both features)

Model Set-up

Set-up

As an economist, our interest is the partial effects of x_1 on y. In this course, we focus on the effects of x_1 on the expected value of y.

We are specifically interested in the partial effects of x_1 on the expected value of y, *holding everything else constant*:

$$\mathbb{E}[y|x_1, w]$$

where w denotes *everything else* that also determines y. This is called **structural conditional expectation**.

$$w = \begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ x_{k-1} \\ q \end{pmatrix}$$

Set-up

If we can collect data on y, x_1, w in a random sample from the underlying population of interest, then it is failry straighforward to estimate

$$\mathbb{E}[y|x_1, w]$$

Set-up

If we are willing to assume some functional form

$$\mathbb{E}[y|x_1, w] = \beta_0 + \beta_1 \cdot x_1 + w'\beta_w$$

.

The partial effect of x_1 on y, holding everything fixed, is given by

$$\frac{\partial}{\partial x_1}\mathbb{E}[y|x_1, w] = \beta_1$$

The parameter of interest is β_1.

Set-up

By decomposition property, the structural model is given by

$$y = \beta_0 + \beta_1 \cdot x_1 + w'\beta_w + u$$

where $\mathbb{E}[u|x_1, w] = 0$.

However, we do not observe all the variables in w. Instead, we observe only

$$\begin{pmatrix} x_2 \\ x_3 \\ \vdots \\ x_{k-1} \end{pmatrix}$$

Structural Equation

As a result,

$$y = \beta_0 + \beta_1 \cdot x_1 + w'\beta_w + u$$

Structural Equation

As a result,

$$\begin{aligned} y &= \beta_0 + \beta_1 \cdot x_1 + w'\beta_w + u \\ &= \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_k \cdot x_{k-1} + [q'\beta_q + u] \end{aligned}$$

Structural Equation

As a result,

$$\begin{aligned} y &= \beta_0 + \beta_1 \cdot x_1 + w'\beta_w + u \\ &= \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_k \cdot x_{k-1} + [q'\beta_q + u] \\ &= \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_k \cdot x_{k-1} + \epsilon \end{aligned}$$

Question: Under what **further** conditions (or assumptions), we can obtain $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)'$ using OLS?

Assumptions

Strucutural Equation

We can obtain an independent, identicallly distributed (i.i.d.) random sample of size N from the population $\{(x_i, y_i) : i = 1, 2, \ldots, N\}$.

$$y_i = x_i'\beta + \epsilon_i$$

where $x = (1, x_1, x_2, \ldots, x_{k-1})$.

- 1. The assumption of linearity is our **first assumption**

Additional Asssumptions:

- 2. [**Exogeneity (Mean Independence)**] $\mathbb{E}[\epsilon_i | x_i] = 0 \implies \mathbb{E}[x_i \epsilon_i] = 0$
- 3. [**Full Rank (Invertibility)**] rank$\mathbb{E}[x_i x_i'] = k$ (positive definite matrix)
- 4. [**Homoskeasticity**] $\mathbb{E}[\epsilon_i^2 | x_i] = \sigma^2 \iff Var(\epsilon_i | x_i) = \sigma^2$

On Independence, Mean Independence, and Correlatedness

A fact

1. Independence $\Longrightarrow$ Mean Independence $\Longrightarrow$ Uncorrelatedness
2. The opposite is not true.

On Independence, Mean Independence, and Correlatedness (Formal Definition)

1. Independence $\implies$ mean independence

$$f(\epsilon, X) = f(\epsilon)f(X) \implies \mathbb{E}[\epsilon|X] = \mathbb{E}[\epsilon]$$

2. mean independence $\implies$ uncorrelatedness

$$\mathbb{E}[\epsilon|X] = \mathbb{E}[\epsilon] \implies \mathbb{E}[\epsilon X] = \mathbb{E}[\epsilon]\mathbb{E}[X] \quad \text{or } cov(\epsilon, X) = 0$$

Identification (implied by population assumptions)

Under the first three assumptions on **population relationships**, the parameter of interest, β, is **identified**. In this context, identification simply means that β can be written in terms of population moments in observable variables.

Just premulitiply the structural equation by x_i,

$$y_i \overset{A1}{=} x_i'\beta + \epsilon_i$$

Identification (implied by population assumptions)

Under the first three assumptions on **population relationships**, the parameter of interest, β, is **identified**. In this context, identification simply means that β can be written in terms of population moments in observable variables.

Just premulitiply the structural equation by x_i,

$$y_i \stackrel{A1}{=} x_i'\beta + \epsilon_i$$
$$x_i y_i = x_i x_i'\beta + x_i \epsilon_i$$

Identification (implied by population assumptions)

Under the first three assumptions on **population relationships**, the parameter of interest, β, is **identified**. In this context, identification simply means that β can be written in terms of population moments in observable variables.

Just premulitiply the structural equation by x_i,

$$
\begin{aligned}
y_i &\overset{A1}{=} x_i'\beta + \epsilon_i \\
x_i y_i &= x_i x_i'\beta + x_i \epsilon_i \\
\mathbb{E}[x_i y_i] &= \mathbb{E}[x_i x_i']\beta + \mathbb{E}[x_i \epsilon_i]
\end{aligned}
$$

Identification (implied by population assumptions)

Under the first three assumptions on **population relationships**, the parameter of interest, β, is **identified**. In this context, identification simply means that β can be written in terms of population moments in observable variables.

Just premulitiply the structural equation by x_i,

$$\begin{aligned} y_i &\overset{A1}{=} x_i'\beta + \epsilon_i \\ x_i y_i &= x_i x_i'\beta + x_i\epsilon_i \\ \mathbb{E}[x_i y_i] &= \mathbb{E}[x_i x_i']\beta + \mathbb{E}[x_i\epsilon_i] \\ \mathbb{E}[x_i y_i] &\overset{A2}{=} \mathbb{E}[x_i x_i']\beta \end{aligned}$$

Identification (implied by population assumptions)

Under the first three assumptions on **population relationships**, the parameter of interest, β, is **identified**. In this context, identification simply means that β can be written in terms of population moments in observable variables.

Just premulitiply the structural equation by x_i,

$$
\begin{aligned}
y_i &\overset{A1}{=} x_i'\beta + \epsilon_i \\
x_i y_i &= x_i x_i'\beta + x_i\epsilon_i \\
\mathbb{E}[x_i y_i] &= \mathbb{E}[x_i x_i']\beta + \mathbb{E}[x_i\epsilon_i] \\
\mathbb{E}[x_i y_i] &\overset{A2}{=} \mathbb{E}[x_i x_i']\beta \\
\beta &\overset{A3}{=} \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i y_i]
\end{aligned}
$$

Assumptions on Samples

The textbook assumes the following conditions, which are implied by the population assumptions and random sampling:

For example,

$$\mathbb{E}[\epsilon_i|X] = \mathbb{E}[\epsilon_i|x_i] = 0$$

$$\mathbb{E}[\epsilon_i^2|X] = \mathbb{E}[\epsilon_i^2|x_i] = \sigma^2$$

$$\mathbb{E}[\epsilon_i\epsilon_j|X] = \mathbb{E}[\epsilon_i|x_i]\mathbb{E}[\epsilon_j|x_j] = 0$$

Assumptions on Samples

2. $\mathbb{E}[\epsilon_i|X] = 0 \implies$

$$\mathbb{E}[\epsilon|X] = \begin{pmatrix} \mathbb{E}[\epsilon_1|X] \\ \mathbb{E}[\epsilon_2|X] \\ \vdots \\ \mathbb{E}[\epsilon_N|X] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[\epsilon_1|\mathbf{x_1}] \\ \mathbb{E}[\epsilon_2|\mathbf{x_2}] \\ \vdots \\ \mathbb{E}[\epsilon_N|\mathbf{x_N}] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

This implies $\mathbb{E}[X\epsilon] = 0$

3. $rank(\mathbb{E}[xx']) = k \implies X'X$ is non-signular and $(X'X)^{-1}$ exists (which is also implied by X is full rank).

Assumptions on Samples

Recall that the unconditional and contional variances in the matrix form are given by

$$Var(Z) = \mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])']$$
$$Var(Z|X) = \mathbb{E}[(Z - \mathbb{E}[Z \mid X])(Z - \mathbb{E}[Z \mid X])' \mid X]$$

4. $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x_i}] = \sigma^2 \implies Var(\epsilon|X) = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$, where I is the identity matrix.

4. $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x_i}] = \sigma^2 \implies Var(\epsilon|X) = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$, where I is the identity matrix.

To see the last assumption holds, recall that

$$Var(\epsilon|X) = \mathbb{E}[(\epsilon - 0)(\epsilon - 0)'|X]$$

4. $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x_i}] = \sigma^2 \implies Var(\epsilon|X) = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$, where I is the identity matrix.

To see the last assumption holds, recall that

$$\begin{aligned} Var(\epsilon|X) &= \mathbb{E}[(\epsilon - 0)(\epsilon - 0)'|X] \\ &= \mathbb{E}[\epsilon\epsilon'|X] \end{aligned}$$

4. $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x_i}] = \sigma^2 \implies Var(\epsilon|X) = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$, where I is the identity matrix.

To see the last assumption holds, recall that

$$\begin{aligned} Var(\epsilon|X) &= \mathbb{E}[(\epsilon - 0)(\epsilon - 0)'|X] \\ &= \mathbb{E}[\epsilon\epsilon'|X] \\ &= \mathbb{E}\left[\begin{pmatrix} \epsilon_1^2, & \epsilon_1\epsilon_2, & \ldots, & \epsilon_1\epsilon_N \\ \epsilon_2\epsilon_1, & \epsilon_2^2, & \ldots, & \epsilon_2\epsilon_N \\ \vdots, & \vdots, & \ddots, & \vdots \\ \epsilon_N\epsilon_1, & \epsilon_N\epsilon_2, & \vdots, & \epsilon_N^2 \end{pmatrix} \mid X\right] \end{aligned}$$

4. $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x_i}] = \sigma^2 \implies Var(\epsilon|X) = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$, where I is the identity matrix.

To see the last assumption holds, recall that

$$
\begin{aligned}
Var(\epsilon|X) &= \mathbb{E}[(\epsilon - 0)(\epsilon - 0)'|X] \\
&= \mathbb{E}[\epsilon\epsilon'|X] \\
&= \mathbb{E}\left[\begin{pmatrix} \epsilon_1^2, & \epsilon_1\epsilon_2, & \dots, & \epsilon_1\epsilon_N \\ \epsilon_2\epsilon_1, & \epsilon_2^2, & \dots, & \epsilon_2\epsilon_N \\ \vdots, & \vdots, & \ddots, & \vdots \\ \epsilon_N\epsilon_1, & \epsilon_N\epsilon_2, & \vdots, & \epsilon_N^2 \end{pmatrix} \mid X\right] \\
&= \mathbb{E}\left[\begin{pmatrix} \sigma^2, & 0, & \dots, & 0 \\ 0, & \sigma^2, & \dots, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots \\ 0, & 0, & \dots, & \sigma^2 \end{pmatrix}\right]
\end{aligned}
$$

4. $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x_i}] = \sigma^2 \implies Var(\epsilon|X) = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$, where I is the identity matrix.

To see the last assumption holds, recall that

$$
\begin{aligned}
Var(\epsilon|X) &= \mathbb{E}[(\epsilon - 0)(\epsilon - 0)'|X] \\
&= \mathbb{E}[\epsilon\epsilon'|X] \\
&= \mathbb{E}\left[\begin{pmatrix} \epsilon_1^2, & \epsilon_1\epsilon_2, & \dots, & \epsilon_1\epsilon_N \\ \epsilon_2\epsilon_1, & \epsilon_2^2, & \dots, & \epsilon_2\epsilon_N \\ \vdots, & \vdots, & \ddots, & \vdots \\ \epsilon_N\epsilon_1, & \epsilon_N\epsilon_2, & \vdots, & \epsilon_N^2 \end{pmatrix} \mid X\right] \\
&= \mathbb{E}\left[\begin{pmatrix} \sigma^2, & 0, & \dots, & 0 \\ 0, & \sigma^2, & \dots, & 0 \\ \vdots, & \vdots, & \ddots, & \vdots \\ 0, & 0, & \dots, & \sigma^2 \end{pmatrix}\right] \\
&= \sigma^2 \cdot I
\end{aligned}
$$

Assumptions and Properties of OLS

We will discuss the meanings of these assumptions in more detail, then what are the consequences of violations, and how we should resolve these issues. Here, we will first focus on what these assumptions will buy us. Specifically, we would like to examine how these assumptions affect the properties of OLS estimator.

Roadmap

It is important to note that

An estimator (or any statistic), a function of random variables, is a random varaible!

We thus need to study the distribution of it.

1. **Finite-sample Distribution** (often only the first two moments): Sample size N
2. **Asymptotic Distribution (Approximation)** (often the complete distribution): $N \to \infty$

Finite Sample Properties

Summary of Finite Sample Properties

Estimation typically involves a random sample from a population; thus, re-sampling yields different values of $\widehat{\beta}$.

```
set seed 123456



forvalues i = 1/3{
    clear
    set obs 1000

    g epsilon = runiform()
    g x = rnormal()

    g y = 1 + 2*x + epsilon

    reg y x

}
```

OLS Estimator is a RV

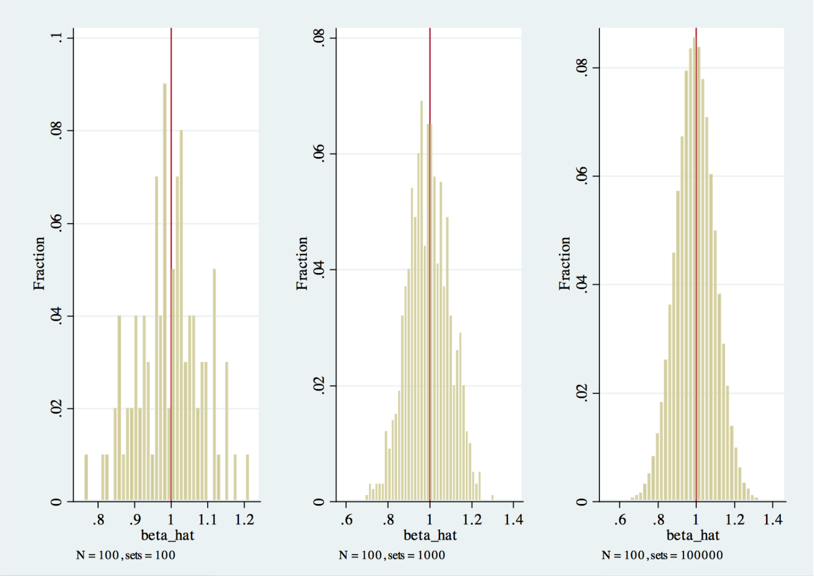


Figure 1: Random Variable

Finite Sample Properties

We will examine three features of the finite-sample distribution of OLS

1. Conditional and Unconditional Mean
2. Conditional and Unconditional Variance
3. Complete Distribution (by imposing further assumptions)

The features inform three common measures of estimator **quality** in finite samples

1. Bias
2. Variance
3. Mean Square Error (MSE)

Important!

To derive these properties, we will invoke the assumptions that we impose earlier. You need to pay close attention to where these assumptions are used, and understand what assumptions are required for derivations of what properties.

Then, you will understand what will happen when the assumptions are violated.

Summary of The Main Results

1. $\mathbb{E}[\widehat{\beta}] = \beta$
2. $Var(\widehat{\beta}|X) = (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1}$.
   a. Special case: $Var(\widehat{\beta}|X) = \sigma^2(X'X)^{-1}$.
3. OLS is the best linear unbiased estimator (BLUE)
4. Under some further assumptions, OLS estimator is normally distributed.

Finite Sample Properties (Unbiasedness)

Unbiasedness

$$\mathbb{E}[\widehat{\beta}] = \beta$$

An estimator is **unbiased** if it yields a correct estimate of β **on average**.

Finite Sample Properties (Unbiasedness)

To investigate the properties of our OLS estimator, it is always useful to re-write it as follows

$$\widehat{\beta}^{OLS} = (X'X)^{-1}X'Y$$

Finite Sample Properties (Unbiasedness)

To investigate the properties of our OLS estimator, it is always useful to re-write it as follows

$$\begin{aligned}\widehat{\beta}^{OLS} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X'\beta + \epsilon)\end{aligned}$$

Finite Sample Properties (Unbiasedness)

To investigate the properties of our OLS estimator, it is always useful to re-write it as follows

$$
\begin{aligned}
\widehat{\beta}^{OLS} &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(X'\beta + \epsilon) \\
&= \beta + (X'X)^{-1}X'\epsilon
\end{aligned}
$$

Finite Sample Properties (Unbiasedness)

Under our assumptions, OLS estimator is unbiased.

$$\mathbb{E}[\widehat{\beta}^{OLS}|X] = \mathbb{E}[(X'X)^{-1}X'Y|X]$$

Finite Sample Properties (Unbiasedness)

Under our assumptions, OLS estimator is unbiased.

$$
\begin{aligned}
\mathbb{E}[\widehat{\beta}^{OLS}|X] &= \mathbb{E}[(X'X)^{-1}X'Y|X] \\
&= \mathbb{E}[\beta + (X'X)^{-1}X'\epsilon|X]
\end{aligned}
$$

Finite Sample Properties (Unbiasedness)

Under our assumptions, OLS estimator is unbiased.

$$\begin{aligned}
\mathbb{E}[\widehat{\beta}^{OLS}|X] &= \mathbb{E}[(X'X)^{-1}X'Y|X] \\
&= \mathbb{E}[\beta + (X'X)^{-1}X'\epsilon|X] \\
&= \beta + \mathbb{E}[(X'X)^{-1}X'\epsilon|X] \\
&= \beta + (X'X)^{-1}X'\mathbb{E}[\epsilon|X] \\
&= \beta
\end{aligned}$$

Finite Sample Properties (Unbiasedness)

Under our assumptions, OLS estimator is unbiased.

$$\begin{aligned}
\mathbb{E}[\widehat{\beta}^{OLS}|X] &= \mathbb{E}[(X'X)^{-1}X'Y|X] \\
&= \mathbb{E}[\beta + (X'X)^{-1}X'\epsilon|X] \\
&= \beta + \mathbb{E}[(X'X)^{-1}X'\epsilon|X] \\
&= \beta + (X'X)^{-1}X'\mathbb{E}[\epsilon|X] \\
&= \beta
\end{aligned}$$

It implies, by the law of iterated expectations,

$$\mathbb{E}[\widehat{\beta}^{OLS}] = \mathbb{E}[\mathbb{E}[\widehat{\beta}^{OLS}|X]] = \mathbb{E}[\beta|X]] = \beta$$

Finite Sample Properties (Unbiasedness)

We need to point out that

1. $\widehat{\beta}^{OLS}$ is not always correct, just correct on average!
2. Our coefficient does not vary with X and identifies only the constant effect.

Finite Sample Properties (Variance)

Recall that the unconditional and contional variances in the matrix form are given by

$$\begin{aligned} Var(Z) &= \mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])'] \\ &= \mathbb{E}[ZZ'] - (\mathbb{E}[Z])(\mathbb{E}[Z])' \\ Var(Z|X) &= \mathbb{E}[(Z - \mathbb{E}[Z|X])(Z - \mathbb{E}[Z|X])'|X] \end{aligned}$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$Var(\widehat{\beta}^{OLS}|X) = \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X]$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$\begin{aligned} Var(\widehat{\beta}^{OLS}|X) &= \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X] \\ &= \mathbb{E}[(\widehat{\beta}^{OLS} - \beta)(\widehat{\beta}^{OLS} - \beta)'|X] \end{aligned}$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$\begin{aligned} Var(\widehat{\beta}^{OLS}|X) &= \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X] \\ &= \mathbb{E}[(\widehat{\beta}^{OLS} - \beta)(\widehat{\beta}^{OLS} - \beta)'|X] \\ &= \mathbb{E}[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'|X] \end{aligned}$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X] \\
&= \mathbb{E}[(\widehat{\beta}^{OLS} - \beta)(\widehat{\beta}^{OLS} - \beta)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)(\epsilon'X(X'X)^{-1}|X]
\end{aligned}$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X] \\
&= \mathbb{E}[(\widehat{\beta}^{OLS} - \beta)(\widehat{\beta}^{OLS} - \beta)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)(\epsilon'X(X'X)^{-1}|X] \\
&= \mathbb{E}[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}|X]
\end{aligned}$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X] \\
&= \mathbb{E}[(\widehat{\beta}^{OLS} - \beta)(\widehat{\beta}^{OLS} - \beta)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)(\epsilon'X(X'X)^{-1}|X] \\
&= \mathbb{E}[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}|X] \\
&= (X'X)^{-1}X'\mathbb{E}[\epsilon\epsilon' \mid X]X(X'X)^{-1}
\end{aligned}$$

Finite Sample Properties (Variance)

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$
$$\widehat{\beta}^{OLS} - \beta = (X'X)^{-1}X'\epsilon$$

$$\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= \mathbb{E}[(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])(\widehat{\beta}^{OLS} - \mathbb{E}[\widehat{\beta}^{OLS}|X])'|X] \\
&= \mathbb{E}[(\widehat{\beta}^{OLS} - \beta)(\widehat{\beta}^{OLS} - \beta)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)((X'X)^{-1}X'\epsilon)'|X] \\
&= \mathbb{E}[((X'X)^{-1}X'\epsilon)(\epsilon'X(X'X)^{-1}|X] \\
&= \mathbb{E}[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}|X] \\
&= (X'X)^{-1}X'\mathbb{E}[\epsilon\epsilon' \mid X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1}
\end{aligned}$$

where $\mathbf{D} = \mathbb{E}[\epsilon\epsilon' \mid X]$.

Finite Sample Properties (Variance)

A special case of the variance, when $D = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$ (**Assumption 4**), is given by

$$Var(\widehat{\beta}^{OLS}|X) = (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1}$$

Finite Sample Properties (Variance)

A special case of the variance, when $D = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$ (**Assumption 4**), is given by

$$
\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1} \\
&\overset{A4}{=} (X'X)^{-1}X'\sigma^2 \cdot I \cdot X(X'X)^{-1}
\end{aligned}
$$

Finite Sample Properties (Variance)

A special case of the variance, when $D = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$ (**Assumption 4**), is given by

$$
\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1} \\
&\overset{A4}{=} (X'X)^{-1}X'\sigma^2 \cdot I \cdot X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 X(X'X)^{-1}
\end{aligned}
$$

Finite Sample Properties (Variance)

A special case of the variance, when $D = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$ (**Assumption 4**), is given by

$$\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1} \\
&\overset{A4}{=} (X'X)^{-1}X'\sigma^2 \cdot I \cdot X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}[X'X(X'X)^{-1}]
\end{aligned}$$

Finite Sample Properties (Variance)

A special case of the variance, when $D = \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 \cdot I$ (**Assumption 4**), is given by

$$\begin{aligned}
Var(\widehat{\beta}^{OLS}|X) &= (X'X)^{-1}X'\mathbf{D}X(X'X)^{-1} \\
&\overset{A4}{=} (X'X)^{-1}X'\sigma^2 \cdot I \cdot X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}[X'X(X'X)^{-1}] \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

As we can see, this is probably the easiest assumption to relax. Instead of estimating $D = \sigma^2 \cdot I$, we will allow for arbitrary form of D, which is called **robust standard errors**.

Finite Sample Properties (Variance)

Some more general rule if you like to memorize things:

For any $N \times r$ matrix $A = A(X)$

$$Var(A'y|X) = Var(A'\epsilon|X) = A'DA$$

where $D = \mathbb{E}[\epsilon\epsilon'|X]$. In this special case, $A = X(X'X)^{-1}$.

Finite Sample Properties: MSE

Mean Squared Error:

$$\text{MSE} = \mathbb{E}[(\widehat{\theta} - \theta)^2] = Var(\widehat{\theta}) + \text{Bias}^2$$

where $\text{Bias} = \mathbb{E}[\widehat{\theta}] - \theta$. Show in homework that this is true.

For an unbiased estimator

$$\text{MSE} = \mathbb{E}[(\widehat{\theta} - \theta)^2] = Var(\widehat{\theta})$$

Finite Sample Properties: MSE

There is a trade-off when choosing an estimator

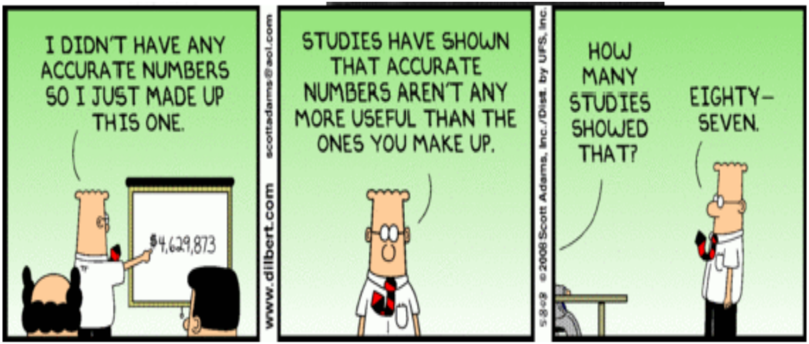


Figure 2: Trade-off

Finite Sample Properties: BLUE OLS

Gauss-Markov Theorem: OLS is **BLUE** (Best Linear Unbiased Estimator)

where **best** menans having the smallest variance, i.e., for any linear and unbiased estimator, b, we have that

$$Var(b|X) - Var(\widehat{\beta}^{OLS}|X) \geq 0$$

(i.e., $Var(b|X) - Var(\widehat{\beta}^{OLS}|X)$ is a **positive semi-definite matrix**)

Finite Sample Properties: BLUE OLS

Sampling variance

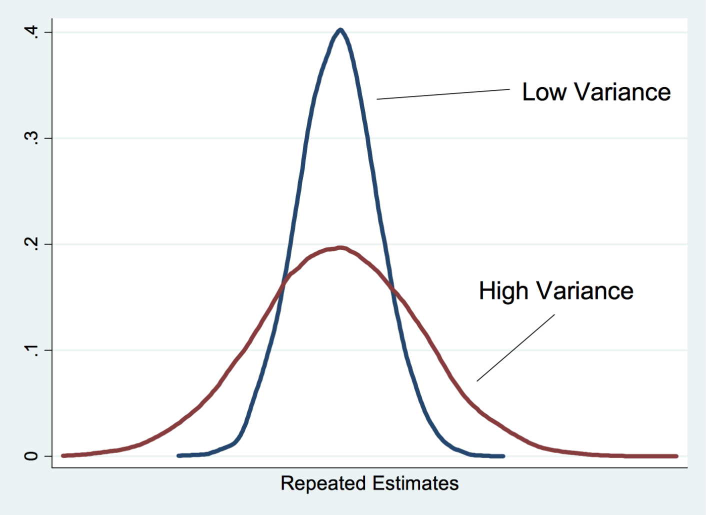
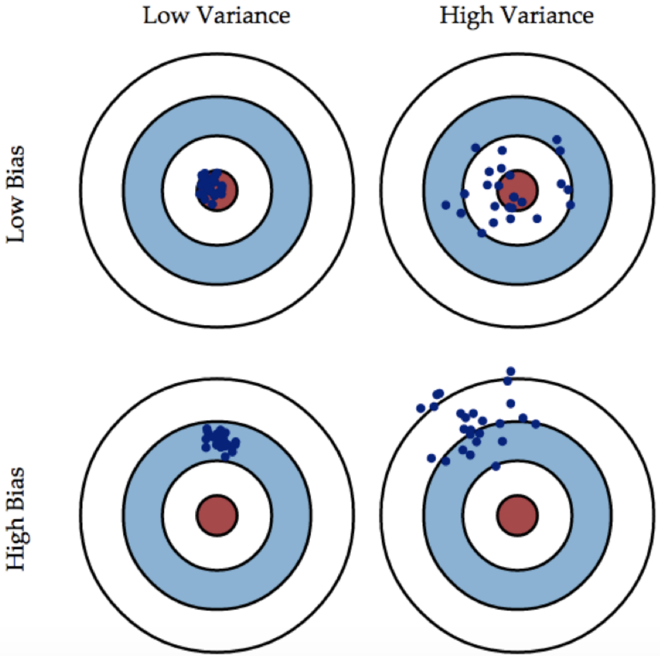


Figure 3: Trade-off


Low Variance
High Variance
Low Bias
High Bias

Finite Sample Properties: BLUE OLS

This is an efficiency justification for the OLS estimator. The justifiation is however limited because the class of models is restricted to **homoskedastic linear regression**, and the class of potential estimators is restricted to **linear unbiased estimators**.

$\Longrightarrow$

A nonlinear or biased estimator could have lower mean squared error than the least-squraes estimator. Note that we are discussing mean squared erorr for β not y.

Finite Sample Properties: BLUE OLS

Proof:

Note that $\widehat{\beta} = (X'X)^{-1}X'Y \equiv AY$. Let b be any unbiased, linear estimator of β:

$$b = WY$$
$$\mathbb{E}[b|X] = \beta$$

where $W \equiv= W(X)$ is a function of the data X.

Finite Sample Properties: BLUE OLS

This implies that

$$\begin{aligned}\beta = \mathbb{E}[b|X] &= \mathbb{E}[WY|X] \\ &= \mathbb{E}[W(X\beta + \epsilon)|X] \\ &= WX\beta + W \cdot \mathbb{E}[\epsilon|X] \\ &\overset{A2}{=} WX\beta\end{aligned}$$

In other words

$$WX = I_k$$

Finite Sample Properties: BLUE OLS

Recall that

$$\widehat{\beta} = \beta + (X'X)^{-1}X'\epsilon$$

$$\widehat{\beta} - \beta = (X'X)^{-1}X'\epsilon$$

Similarly, for any unbiased, linear estimator

$$\begin{aligned} b = WY &= W(X\beta + \epsilon) \\ &= \beta + W\epsilon \\ b - \beta &= W\epsilon \end{aligned}$$

Finite Sample Properties: BLUE OLS

$$
\begin{aligned}
Var(\widehat{\beta} - b|X) &= Var(\widehat{\beta}|X) - Cov(\widehat{\beta}, b|X) - Cov(\widehat{\beta}, b|X) + Var(b|X) \\
&= Var(b|X) - Var(\widehat{\beta}|X)
\end{aligned}
$$

We show that $Cov(\widehat{\beta}, b|X) = Var(\widehat{\beta}|X)$.

Fact: the variance-covariance matrix is a semi-positive defninite matrix. Done!

Finite Sample Properties: BLUE OLS

Next, we show that $Cov(\widehat{\beta}, b|X) = Var(\widehat{\beta}|X)$.

$$
\begin{aligned}
Cov(\widehat{\beta}, b|X) &= \mathbb{E}[(\widehat{\beta} - \beta)(b - \beta)'|X] \\
&= \mathbb{E}[(X'X)^{-1}X'\epsilon(W\epsilon)'|X] \\
&= \mathbb{E}[(X'X)^{-1}X'\epsilon\epsilon'W'|X] \\
&= (X'X)^{-1}X'\mathbb{E}[\epsilon\epsilon'|X]W' \\
&\overset{A4}{=} (X'X)^{-1}X'\sigma^2 \cdot IW' \\
&= \sigma^2(X'X)^{-1}X'W' \\
&= \sigma^2(X'X)^{-1}(WX)' \\
&= \sigma^2(X'X)^{-1} \\
&= Var(\widehat{\beta}|X)
\end{aligned}
$$

Finite Sample Properties: BLUE OLS

$$
\begin{aligned}
Var(\widehat{\beta} - b|X) &= Var(\widehat{\beta}|X) - Cov(\widehat{\beta}, b|X) - Cov(\widehat{\beta}, b|X) + Var(b|X) \\
&= Var(b|X) - Var(\widehat{\beta}|X)
\end{aligned}
$$

We know that the variance-covariance matrix is a semi-positive defninite matrix. Done!

Finite Sample Properties (Unconditional Variance)

We can also show (in your homework) that

$$Var(\widehat{\beta}^{OLS}) = \mathbb{E}[Var(\widehat{\beta}^{OLS}|X)] + Var(\mathbb{E}[\widehat{\beta}^{OLS}|X])$$

Finite Sample Properties (Unconditional Variance)

We can also show (in your homework) that

$$Var(\widehat{\beta}^{OLS}) = \mathbb{E}[Var(\widehat{\beta}^{OLS}|X)] + Var(\mathbb{E}[\widehat{\beta}^{OLS}|X])$$

Under the assumption of non-stochastic regressor, the second term is equal to zero. So,

$$\begin{aligned} Var(\widehat{\beta}^{OLS}) &= \mathbb{E}[Var(\widehat{\beta}^{OLS}|X)] \\ &= \sigma^2 \mathbb{E}[(X'X)^{-1}] \end{aligned}$$

As noted in Greene, the unconditional variance of $\widehat{\beta}^{OLS}$ can only be described in terms of the average behavior of X in some sense, it would be necessary to make some assumptions about the variances and covariances of the regressors. We will discuss this later.

Finite Sample Properties: Complete Distribution

Recall that

$$\widehat{\beta}^{OLS} = \beta + (X'X)^{-1}X'\epsilon$$

Suppose, further, that

$$\epsilon_i | x_i \sim N(0, \sigma^2)$$

$$\epsilon | X \sim N(0, \sigma^2 \cdot I)$$

Finite Sample Properties: Complete Distribution

Recall what we learned about normal distribution earlier in the semester,

Lemma: Let $x \sim \mathcal{N}(\mu, \Sigma)$, and $y = \alpha + \Gamma x$. then

$$y \sim \mathcal{N}(\alpha + \Gamma\mu, \Gamma\Sigma\Gamma')$$

It is trivial to show that

$$\begin{aligned}\widehat{\beta}^{OLS} &= \beta + (X'X)^{-1}X'\epsilon \\ &= \beta + \Gamma\epsilon\end{aligned}$$

Finite Sample Properties: Complete Distribution

Recall what we learned about normal distribution earlier in the semester,

Lemma: Let $x \sim \mathcal{N}(\mu, \Sigma)$, and $y = \alpha + \Gamma x$. then

$$y \sim \mathcal{N}(\alpha + \Gamma\mu, \Gamma\Sigma\Gamma')$$

It is trivial to show that

$$\begin{aligned} \widehat{\beta}^{OLS} &= \beta + (X'X)^{-1}X'\epsilon \\ &= \beta + \Gamma\epsilon \\ \widehat{\beta}|X &\sim N(\beta + \Gamma \cdot 0, \Gamma\sigma^2 \cdot I\Gamma') \end{aligned}$$

Finite Sample Properties: Complete Distribution

Recall what we learned about normal distribution earlier in the semester,

Lemma: Let $x \sim \mathcal{N}(\mu, \Sigma)$, and $y = \alpha + \Gamma x$. then

$$y \sim \mathcal{N}(\alpha + \Gamma\mu, \Gamma\Sigma\Gamma')$$

It is trivial to show that

$$\begin{aligned}
\widehat{\beta}^{OLS} &= \beta + (X'X)^{-1}X'\epsilon \\
&= \beta + \Gamma\epsilon \\
\widehat{\beta}|X &\sim N(\beta + \Gamma \cdot 0, \Gamma\sigma^2 \cdot I\Gamma') \\
&\sim N(\beta, \sigma^2(X'X)^{-1}X'I((X'X)^{-1}X')')
\end{aligned}$$

Finite Sample Properties: Complete Distribution

Recall what we learned about normal distribution earlier in the semester,

Lemma: Let $x \sim \mathcal{N}(\mu, \Sigma)$, and $y = \alpha + \Gamma x$. then

$$y \sim \mathcal{N}(\alpha + \Gamma\mu, \Gamma\Sigma\Gamma')$$

It is trivial to show that

$$\begin{aligned}
\widehat{\beta}^{OLS} &= \beta + (X'X)^{-1}X'\epsilon \\
&= \beta + \Gamma\epsilon \\
\widehat{\beta}|X &\sim N(\beta + \Gamma \cdot 0, \Gamma\sigma^2 \cdot I\Gamma') \\
&\sim N(\beta, \sigma^2(X'X)^{-1}X'I((X'X)^{-1}X')') \\
&\sim N(\beta, \sigma^2(X'X)^{-1}X'X(X'X)^{-1})
\end{aligned}$$

Finite Sample Properties: Complete Distribution

Recall what we learned about normal distribution earlier in the semester,

Lemma: Let $x \sim \mathcal{N}(\mu, \Sigma)$, and $y = \alpha + \Gamma x$. then

$$y \sim \mathcal{N}(\alpha + \Gamma\mu, \Gamma\Sigma\Gamma')$$

It is trivial to show that

$$\begin{aligned}
\widehat{\beta}^{OLS} &= \beta + (X'X)^{-1}X'\epsilon \\
&= \beta + \Gamma\epsilon \\
\widehat{\beta}|X &\sim N(\beta + \Gamma \cdot 0, \Gamma\sigma^2 \cdot I\Gamma') \\
&\sim N(\beta, \sigma^2(X'X)^{-1}X'I((X'X)^{-1}X')') \\
&\sim N(\beta, \sigma^2(X'X)^{-1}X'X(X'X)^{-1}) \\
&\sim N(\beta, \sigma^2(X'X)^{-1})
\end{aligned}$$