Finite-Sample Properties of OLS

Raphael Redmer, Arkadiusz Modzelewski,
Burak Balaban

University of Bonn
Research Module in Econometrics and Statistics

November 4, 2019

Overview

Introduction to the Theorem

Notation

1. y_i dependent variable,
2. x_{ik} kth independent variable (or regressor) with $k = 1, ..., K$,
3. ϵ_i stochastic error term,
4. i indexes the ith individual with $i = 1, ..., n$, where n is the sample size.

Assumptions

1. Assumption 1.1 Linearity:
   $$y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i, i = 1, ..., n$$
2. Assumption 1.2 Strict Exogeneity:
   $\mathbb{E}(\epsilon_i|\mathbf{X}) = 0$
3. Assumption 1.3 Rank Condition:
   $rank(\mathbf{X}) = K$
4. Assumption 1.4 Spherical Error:
   $\mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2$
   $\mathbb{E}(\epsilon_i\epsilon_j|\mathbf{X}) = 0, \; i \neq j$

Assumptions

1. Assumption 1.1 Linearity:
   $$y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i, i = 1, ..., n$$
2. Assumption 1.2 Strict Exogeneity:
   $\mathbb{E}(\epsilon_i|\mathbf{X}) = 0$
3. Assumption 1.3 Rank Condition:
   $rank(\mathbf{X}) = K$
4. Assumption 1.4 Spherical Error:
   $\mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2$
   $\mathbb{E}(\epsilon_i \epsilon_j|\mathbf{X}) = 0, i \neq j$

Assumptions

1. Assumption 1.1 Linearity:
   $$y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i, i = 1, ..., n$$
2. Assumption 1.2 Strict Exogeneity:
   $\mathbb{E}(\epsilon_i|\mathbf{X}) = 0$
3. Assumption 1.3 Rank Condition:
   $rank(\mathbf{X}) = K$
4. Assumption 1.4 Spherical Error:
   $\mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2$
   $\mathbb{E}(\epsilon_i\epsilon_j|\mathbf{X}) = 0, i \neq j$

Assumptions

1. Assumption 1.1 Linearity:
   $$y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \epsilon_i, i = 1, ..., n$$
2. Assumption 1.2 Strict Exogeneity:
   $\mathbb{E}(\epsilon_i|\mathbf{X}) = 0$
3. Assumption 1.3 Rank Condition:
   $rank(\mathbf{X}) = K$
4. Assumption 1.4 Spherical Error:
   $\mathbb{E}(\epsilon_i^2|\mathbf{X}) = \sigma^2$
   $\mathbb{E}(\epsilon_i\epsilon_j|\mathbf{X}) = 0, \, i \neq j$

Introduction to the Theorem

Theorem 4.1: Finite Sample Properties

The OLS estimator b

1. is an unbiased estimator: $\mathbb{E}(b|X) = \beta$
2. has variance: $\mathbb{V}(b|X) = \sigma^2(X'X)^{-1}$
3. has the lowest variance in the class of all linear unbiased estimators.

Introduction to the Theorem

Theorem 4.1: Finite Sample Properties

The OLS estimator b

1. is an unbiased estimator: $\mathbb{E}(b|X) = \beta$
2. has variance: $\mathbb{V}(b|X) = \sigma^2(X'X)^{-1}$
3. has the lowest variance in the class of all linear unbiased estimators.

Introduction to the Theorem

Theorem 4.1: Finite Sample Properties

The OLS estimator b

1. is an unbiased estimator: $\mathbb{E}(b|X) = \beta$
2. has variance: $\mathbb{V}(b|X) = \sigma^2(X'X)^{-1}$
3. has the lowest variance in the class of all linear unbiased estimators.

Explanation of (3)

For any unbiased estimator $\tilde{\mathbf{b}}$ that is linear in $\mathbf{y}$, we have:
$\mathbb{V}(\tilde{\mathbf{b}}|\mathbf{X}) \geq \mathbb{V}(\mathbf{b}|\mathbf{X})$ in the matrix sense,
meaning that
$\mathbb{V}(\tilde{\mathbf{b}}|\mathbf{X}) - \mathbb{V}(\mathbf{b}|\mathbf{X}) = \mathbf{D}$, where $\mathbf{D}$ is a positive semidefinite $K \times K$ matrix, i.e., $\mathbf{a}'\mathbf{D}\mathbf{a} \geq 0$ for any K-dimensional vector $\mathbf{a}$.
$\Rightarrow \mathbb{V}(\tilde{\mathrm{b}}_k|\mathbf{X}) \geq \mathbb{V}(\mathrm{b}_k|\mathbf{X})$ for any $k = 1, \ldots, K$.

Proof of (1)

$$
\begin{aligned}
\mathbb{E}(b|X) &= \mathbb{E}((X'X)^{-1}X'y|X) \\
&= \mathbb{E}((X'X)^{-1}X'(X\beta+\varepsilon)|X) \\
&= \mathbb{E}((X'X)^{-1}X'X\beta+(X'X)^{-1}X'\varepsilon)|X) \\
&= \beta+(X'X)^{-1}X'\underbrace{\mathbb{E}(\varepsilon|X)}_{\underset{(A1.2)}{=}0} = \beta
\end{aligned}
$$

$\square$

Proof of (1)

$$
\begin{aligned}
\mathbb{E}(b|X) &= \mathbb{E}((X'X)^{-1}X'y|X) \\
&= \mathbb{E}((X'X)^{-1}X'(X\beta+\varepsilon)|X) \\
&= \mathbb{E}((X'X)^{-1}X'X\beta+(X'X)^{-1}X'\varepsilon)|X) \\
&= \beta + (X'X)^{-1}X'\underbrace{\mathbb{E}(\varepsilon|X)}_{\underset{(A1.2)}{=}0} = \beta
\end{aligned}
$$

$\square$

Proof of (1)

$$
\begin{aligned}
\mathbb{E}(b|X) &= \mathbb{E}((X'X)^{-1}X'y|X) \\
&= \mathbb{E}((X'X)^{-1}X'(X\beta+\varepsilon)|X) \\
&= \mathbb{E}((X'X)^{-1}X'X\beta+(X'X)^{-1}X'\varepsilon)|X) \\
&= \beta+(X'X)^{-1}X'\underbrace{\mathbb{E}(\varepsilon|X)}_{\underset{(A1.2)}{=}0} = \beta
\end{aligned}
$$

$\square$

Proof of (1)

$$
\begin{aligned}
\mathbb{E}(b|X) &= \mathbb{E}((X'X)^{-1}X'y|X) \\
&= \mathbb{E}((X'X)^{-1}X'(X\beta + \varepsilon)|X) \\
&= \mathbb{E}((X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon)|X) \\
&= \beta + (X'X)^{-1}X' \underbrace{\mathbb{E}(\varepsilon|X)}_{\underset{(A1.2)}{=}\ 0} = \beta
\end{aligned}
$$

$\square$

Proof of (2)

$$
\begin{aligned}
\mathbb{V}(b|X) &= \mathbb{V}(b-\beta|X) && (\beta \text{ is not random}) \\
&= \mathbb{V}((X'X)^{-1}X'\varepsilon|X)) && (b = \beta + (X'X)^{-1}X'\varepsilon) \\
&= (X'X)^{-1}X'\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}X(X'X)^{-1} && (\text{A1.4: Spherical Error}) \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

$\square$

Proof of (2)

$$
\begin{aligned}
\mathbb{V}(b|X) &= \mathbb{V}(b-\beta|X) && (\beta \text{ is not random}) \\
&= \mathbb{V}((X'X)^{-1}X'\varepsilon|X)) && (b=\beta+(X'X)^{-1}X'\varepsilon) \\
&= (X'X)^{-1}X'\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}X(X'X)^{-1} && (\text{A1.4: Spherical Error}) \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

□

Proof of (2)

$$\begin{aligned}
\mathbb{V}(b|X) &= \mathbb{V}(b-\beta|X) && (\beta \text{ is not random})\\
&= \mathbb{V}((X'X)^{-1}X'\varepsilon|X)) && (b = \beta + (X'X)^{-1}X'\varepsilon)\\
&= (X'X)^{-1}X'\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}X(X'X)^{-1} && (\text{A1.4: Spherical Error})\\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

$\square$

Proof of (2)

$$
\begin{aligned}
\mathbb{V}(b|X) &= \mathbb{V}(b-\beta|X) && (\beta \text{ is not random})\\
&= \mathbb{V}((X'X)^{-1}X'\varepsilon|X)) && (b = \beta + (X'X)^{-1}X'\varepsilon)\\
&= (X'X)^{-1}X' \underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n} X(X'X)^{-1} && (\text{A1.4: Spherical Error})\\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

$\square$

Proof (3)

Assumptions

Let $\widetilde{b}$ be an unbiased estimator and linear in y with

$$\widetilde{b} = Cy$$

where $C \in \mathbb{R}^{K \times n}$, which is a function of X and/or nonrandom components (e.g. $C = (X'X)^{-1}X'$ as in OLS Model).

Proof of (3)
Decomposition

With $D := C - (X'X)^{-1}X'$, $\widetilde{b}$ can be decomposed to:

$$\begin{aligned}\widetilde{b} = Cy &= (C \overbrace{-(X'X)^{-1}X' + (X'X)^{-1}X'}^{=0})y \\ &= Dy + \underbrace{(X'X)^{-1}X'y}_{=b} \\ &= D(X\beta + \varepsilon) + b \\ &= DX\beta + D\varepsilon + b\end{aligned}$$

Proof of (3)

Decomposition

With $D := C - (X'X)^{-1}X'$, $\widetilde{b}$ can be decomposed to:

$$\begin{aligned}\widetilde{b} = Cy &= (C \overbrace{-(X'X)^{-1}X' + (X'X)^{-1}X'}^{=0})y \\ &= Dy + \underbrace{(X'X)^{-1}X'y}_{=b} \\ &= D(X\beta + \varepsilon) + b \\ &= DX\beta + D\varepsilon + b\end{aligned}$$

Proof of (3)
Decomposition

With $D := C - (X'X)^{-1}X'$, $\widetilde{b}$ can be decomposed to:

$$\begin{aligned}\widetilde{b} = Cy &= (C \overbrace{-(X'X)^{-1}X' + (X'X)^{-1}X'}^{=0})y \\ &= Dy + \underbrace{(X'X)^{-1}X'y}_{=b} \\ &= D(X\beta + \varepsilon) + b \\ &= DX\beta + D\varepsilon + b\end{aligned}$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\Longrightarrow DX = 0 \\
\Longrightarrow \widetilde{b} &= D\varepsilon + b \\
\Longleftrightarrow \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$
\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\implies DX = 0 \\
\implies \widetilde{b} &= D\varepsilon + b \\
\iff \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}
$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\Longrightarrow DX = 0
\end{aligned}$$

$$\begin{aligned}
\Longrightarrow \widetilde{b} &= D\varepsilon + b \\
\Longleftrightarrow \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\Longrightarrow DX = 0
\end{aligned}$$

$$\Longrightarrow \widetilde{b} = D\varepsilon + b$$

$$\begin{aligned}
\Longleftrightarrow \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\Longrightarrow DX = 0 \\
\Longrightarrow \widetilde{b} &= D\varepsilon + b \\
\Longleftrightarrow \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\Longrightarrow DX = 0 \\
\Longrightarrow \widetilde{b} &= D\varepsilon + b \\
\Longleftrightarrow \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}$$

Proof (3)

$\widetilde{b}$ unbiased

Due to $\widetilde{b}$ being unbiased, it follows that $\mathbb{E}(\widetilde{b}|X) = \beta$. Therefore,

$$\begin{aligned}
\mathbb{E}(\widetilde{b}|X) &= \mathbb{E}(DX\beta|X) + \mathbb{E}(D\varepsilon|X) + \mathbb{E}(b|X) \\
&= DX\beta + 0 + \beta = \beta \\
&\Longrightarrow DX = 0 \\
\Longrightarrow \widetilde{b} &= D\varepsilon + b \\
\Longleftrightarrow \widetilde{b} - \beta &= D\varepsilon + (b - \beta) \\
&= D\varepsilon + (X'X)^{-1}X'\varepsilon \\
&= (D + (X'X)^{-1}X')\varepsilon \qquad (*)
\end{aligned}$$

Proof (3)
Rearrangement

By using $(*)$, $\mathbb{V}(\widetilde{b}|X)$ can now be rearranged such that:

$$\begin{aligned}
\mathbb{V}(\widetilde{b}|X) &= \mathbb{V}(\widetilde{b} - \beta|X) \\
&\overset{(*)}{=} \mathbb{V}((D + (X'X)^{-1}X')\varepsilon|X) \\
&= (D + (X'X)^{-1}X')\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}(D' + X(X'X)^{-1}) \\
&\overset{(\mathrm{DX=0})}{=} \sigma^2(D'D + (X'X)^{-1}) \\
&\geq \sigma^2((X'X)^{-1}) = \mathbb{V}(b|X),
\end{aligned}$$

since $D'D$ is positive semidefinite. □

Proof (3)

Rearrangement

By using $(*)$, $\mathbb{V}(\widetilde{b}|X)$ can now be rearranged such that:

$$
\begin{aligned}
\mathbb{V}(\widetilde{b}|X) &= \mathbb{V}(\widetilde{b} - \beta|X) \\
&\overset{(*)}{=} \mathbb{V}((D + (X'X)^{-1}X')\varepsilon|X) \\
&= (D + (X'X)^{-1}X')\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}(D' + X(X'X)^{-1}) \\
&\overset{(\mathrm{DX}=0)}{=} \sigma^2(D'D + (X'X)^{-1}) \\
&\geq \sigma^2((X'X)^{-1}) = \mathbb{V}(b|X),
\end{aligned}
$$

since $D'D$ is positive semidefinite. $\square$

Proof (3)
Rearrangement

By using $(*)$, $\mathbb{V}(\widetilde{b}|X)$ can now be rearranged such that:

$$\begin{aligned}
\mathbb{V}(\widetilde{b}|X) &= \mathbb{V}(\widetilde{b}-\beta|X) \\
&\overset{(*)}{=} \mathbb{V}((D+(X'X)^{-1}X')\varepsilon|X) \\
&= (D+(X'X)^{-1}X')\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}(D'+X(X'X)^{-1}) \\
&\overset{(DX=0)}{=} \sigma^2(D'D+(X'X)^{-1}) \\
&\geq \sigma^2((X'X)^{-1}) = \mathbb{V}(b|X),
\end{aligned}$$

since $D'D$ is positive semidefinite. $\square$

Proof (3)
Rearrangement

By using $(*)$, $\mathbb{V}(\widetilde{b}|X)$ can now be rearranged such that:

$$\begin{aligned}
\mathbb{V}(\widetilde{b}|X) &= \mathbb{V}(\widetilde{b} - \beta|X) \\
&\overset{(*)}{=} \mathbb{V}((D + (X'X)^{-1}X')\varepsilon|X) \\
&= (D + (X'X)^{-1}X')\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}(D' + X(X'X)^{-1}) \\
&\overset{(\text{DX}=0)}{=} \sigma^2(D'D + (X'X)^{-1}) \\
&\geq \sigma^2((X'X)^{-1}) = \mathbb{V}(b|X),
\end{aligned}$$

since $D'D$ is positive semidefinite. □

Proof (3)
Rearrangement

By using $(*)$, $\mathbb{V}(\widetilde{b}|X)$ can now be rearranged such that:

$$\begin{aligned}
\mathbb{V}(\widetilde{b}|X) &= \mathbb{V}(\widetilde{b} - \beta|X) \\
&\overset{(*)}{=} \mathbb{V}((D + (X'X)^{-1}X')\varepsilon|X) \\
&= (D + (X'X)^{-1}X')\underbrace{\mathbb{V}(\varepsilon|X)}_{=\sigma^2 I_n}(D' + X(X'X)^{-1}) \\
&\overset{(\mathrm{DX=0})}{=} \sigma^2(D'D + (X'X)^{-1}) \\
&\geq \sigma^2((X'X)^{-1}) = \mathbb{V}(b|X),
\end{aligned}$$

since $D'D$ is positive semidefinite. $\square$

Conclusions

- Under stated assumptions OLS estimator has very nice properties
- Assumptions in the real life examples are very difficult to satisfy. Example: model with lagged dependent variable
- Team work: communication is more important than you think.