

Multiple Linear Regression

Part B: Ordinary Least Squares

THOMAS WIEMANN
University of Chicago

Econometrics
Econ 21020

Updated: May 18, 2022

Summary

In Part A, we introduced $\text{BLP}(Y|X)$ as approximation to $E[Y|X]$.

- ▷ BLP-coefficients are well-defined when $E[XX^\top]^{-1}$ exists;
- ▷ Used the Frisch-Waugh Theorem for subvector analysis;
- ▷ Discussed interpretation using a generalized Yitzhaki's Theorem;

The BLP and its coefficients β are theoretical concepts.

In Part B, we bridge the gap between BLP and real data using statistics.

- ▷ Develop the *ordinary least squares* estimator;
- ▷ Analyze its statistical properties under an iid sample;
- ▷ Use matrix calculus for implementation.

1. Ordinary Least Squares
2. Estimator Properties
 - ▷ Bias
 - ▷ Consistency
 - ▷ Asymptotic Distribution
3. Implementation

1. **Ordinary Least Squares**
2. Estimator Properties
 - ▷ Bias
 - ▷ Consistency
 - ▷ Asymptotic Distribution
3. Implementation

Ordinary Least Squares

Let Y be a random variable and $X = (1, X_1, \dots, X_k)^\top$ be a random vector. Consider a random sample $(Y^1, X^1), \dots, (Y^n, X^n) \stackrel{iid}{\sim} (Y, X)$.

From Lecture 8A, we know that the BLP-coefficients are given by

$$\beta = E[XX^\top]^{-1}E[XY], \quad (1)$$

whenever $E[XX^\top]^{-1}$ exists.

This suggests the sample analogue estimator

$$\hat{\beta}_n = \quad (2)$$

Notation: Superscripts – i.e., X^1, \dots, X^n – are used as sample indices throughout.

Ordinary Least Squares (Contd.)

The estimator $\hat{\beta}_n$ is known as *ordinary least squares* (OLS). This is because it can also be motivated as solutions to the least-squares sample criterion:

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^{k+1}} \frac{1}{n} \sum_{i=1}^n (Y^i - X^{i\top} \beta)^2, \quad (3)$$

whenever $E[\sum_{i=1}^n X^i X^{i\top}]^{-1}$ exists. In particular, we have:

Ordinary Least Squares (Contd.)

For our analysis, it's useful to rewrite $\hat{\beta}_n$ using $\varepsilon^i \equiv Y^i - \text{BLP}(Y^i|X^i)$:

$$\hat{\beta}_n = \tag{4}$$

1. Ordinary Least Squares
2. **Estimator Properties**
 - ▷ **Bias**
 - ▷ Consistency
 - ▷ Asymptotic Distribution
3. Implementation

Our analysis of the OLS estimator begins with its bias.

We assume here that X is continuous to ensure existence of $E[\sum_{i=1}^n X^i X^{i\top}]^{-1}$ (for $n > k + 1$) when $E[XX^\top]^{-1}$ exists.

The bias of $\hat{\beta}_n$ when X is continuous and $E[XX^\top]^{-1}$ exists is given by

$$\text{Bias}(\hat{\beta}_n) = E[\hat{\beta}_n] - \beta =$$

(5)

Hence, if $E[\varepsilon^i|X^i] = 0$, then $\text{Bias}(\hat{\beta}_n) = 0$.

- ▷ Does $E[\varepsilon^i|X^i] = 0$ hold generally? No: $E[\varepsilon^i X^i] = 0 \not\Rightarrow E[\varepsilon^i|X^i] = 0$.
- ▷ When do we know that $E[\varepsilon^i|X^i] = 0$? Special case: Linear $E[Y|X]$.

Many textbooks state that the OLS estimator $\hat{\beta}_n$ is unbiased for β .

- ▷ Importantly: Strong assumption are made along the way!
- ▷ We *only* showed $\text{Bias}(\hat{\beta}_n) = 0$ if $E[Y|X]$ linear *and* X is continuous.

Generally, little reason to believe $\text{Bias}(\hat{\beta}_n) = 0$ in economic applications:

- ▷ Economic theory rarely implies linear $E[Y|X]$ with continuous X .
- ▷ Horrible news? No: Most estimators are biased in practice...

1. Ordinary Least Squares
2. **Estimator Properties**
 - ▷ Bias
 - ▷ **Consistency**
 - ▷ Asymptotic Distribution
3. Implementation

Theorem 1 ensures OLS satisfies the minimum requirement: Consistency.

Theorem 1

Let Y be a random variable and $X = (1, X_1, \dots, X_k)^\top$ be a random vector such that $E[XX^\top]^{-1}$ exists, and let β denote the $BLP(Y|X)$ -coefficient. If $\hat{\beta}_n$ are the OLS estimators constructed using $(Y^1, X^1), \dots, (Y^n, X^n) \stackrel{iid}{\sim} (Y, X)$, then

$$\hat{\beta}_n \xrightarrow{P} \beta. \quad (6)$$

Since the OLS estimators are continuous functions of moments of (Y, X) , we can prove this straightforwardly using the WLLN and CMT.

Consistency (Contd.)

Proof.

1. Ordinary Least Squares
2. **Estimator Properties**
 - ▷ Bias
 - ▷ Consistency
 - ▷ **Asymptotic Distribution**
3. Implementation

Theorem 2 shows that OLS is asymptotically normal.

Theorem 2

Let Y be a random variable and $X = (1, X_1, \dots, X_k)^\top$ be a random vector such that $E[XX^\top]^{-1}$ exists, and let β denote the $BLP(Y|X)$ -coefficient. If $\hat{\beta}_n$ are the OLS estimators constructed using $(Y^1, X^1), \dots, (Y^n, X^n) \stackrel{iid}{\sim} (Y, X)$, then

$$\sqrt{n} \left(\hat{\beta}_n - \beta \right) \xrightarrow{d} N(0, \Sigma), \quad (7)$$

where

$$\Sigma = E[XX^\top]^{-1} E[XX^\top \varepsilon^2] E[XX^\top]^{-1}, \quad (8)$$

with $\varepsilon \equiv Y - BLP(Y|X)$.

Asymptotic Distribution (Contd.)

Proof.

OLS Covariance Estimation

Theorem 2 is of no practical use unless we can replace the expression for the asymptotic variance by a consistent estimator. Fortunately, we can.

Theorem 3

Let Y be a random variable and $X = (1, X_1, \dots, X_k)^\top$ be a random vector such that $E[XX^\top]^{-1}$ exists, and let β denote the $BLP(Y|X)$ -coefficient. If $\hat{\beta}_n$ is the OLS estimator constructed using $(Y^1, X^1), \dots, (Y^n, X^n) \stackrel{iid}{\sim} (Y, X)$, then

$$\sqrt{n} \widehat{\Sigma}_n^{-\frac{1}{2}} \left(\hat{\beta}_n - \beta \right) \xrightarrow{d} N(0, \mathbf{I}_{k+1}), \quad (9)$$

where

$$\widehat{\Sigma}_n = \left(\frac{1}{n} \sum_{i=1}^n X^i X^{i\top} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X^i X^{i\top} \hat{\varepsilon}^i{}^2 \right) \left(\frac{1}{n} \sum_{i=1}^n X^i X^{i\top} \right)^{-1} \quad (10)$$

and $\hat{\varepsilon}^i = Y^i - X^{i\top} \hat{\beta}_n$.

OLS Covariance Estimation (Contd.)

Proof.

OLS Covariance Estimation (Contd.)

Theorem 2 and 3 give inference for the vector $\hat{\beta}_n$.

- ▷ Often interested only in a subvector;
- ▷ E.g., the estimator $\hat{\beta}_{jn}$ of β_j .

Corollary 1 and 2 give inference for individual components of $\hat{\beta}_n$.

- ▷ Corollary 1 combines Theorem 2 + Slutsky's Theorem;
- ▷ Corollary 3 gives the standard error formula.

Subvector Asymptotic Distribution

Corollary 1

Let Y be a random variable and $X = (1, X_1, \dots, X_k)^\top$ be a random vector such that $E[XX^\top]^{-1}$ exists, and let $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ denote the $BLP(Y|X)$ -coefficient. If $\hat{\beta}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n}, \dots, \hat{\beta}_{kn})$ is the OLS estimator constructed using $(Y^1, X^1), \dots, (Y^n, X^n) \stackrel{iid}{\sim} (Y, X)$, then

$$\sqrt{n} \left(\hat{\beta}_{jn} - \beta_j \right) \xrightarrow{d} N(0, e_j^\top \Sigma e_j), \quad \forall j = 0, 1, \dots, k, \quad (11)$$

where Σ is defined by Equation (8) and e_j is the j th unit vector.

Proof.



Note: $e_j^\top \Sigma e_j$ simply selects the j th diagonal entry of Σ

Corollary 2

Let Y be a random variable and $X = (1, X_1, \dots, X_k)^\top$ be a random vector such that $E[XX^\top]^{-1}$ exists, and let $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ denote the $BLP(Y|X)$ -coefficient. If $\hat{\beta}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n}, \dots, \hat{\beta}_{kn})$ is the OLS estimator constructed using $(Y^1, X^1), \dots, (Y^n, X^n) \stackrel{iid}{\sim} (Y, X)$, then

$$\frac{\hat{\beta}_{jn} - \beta_j}{\text{se}(\hat{\beta}_{jn})} \xrightarrow{d} N(0, 1), \quad \forall j = 0, 1, \dots, k, \quad (12)$$

where

$$\text{se}(\hat{\beta}_{jn}) = \frac{1}{\sqrt{n}} \sqrt{e_j^\top \hat{\Sigma}_n e_j} \quad (13)$$

with $\hat{\Sigma}_n$ is defined by Equation (3) and e_j is the j th unit vector.

Note: $e_j^\top \hat{\Sigma}_n e_j$ simply selects the j th diagonal entry of $\hat{\Sigma}_n$

Standard Error (Contd.)

Proof.

1. Ordinary Least Squares
2. **Estimator Properties**
 - ▷ Bias
 - ▷ Consistency
 - ▷ Asymptotic Distribution
3. **Implementation**

OLS Implementation

Implementing OLS by brute force (e.g., $\sum_{i=1}^n X^i X^{i\top}$) is difficult.

▷ Instead: Use matrix operations for straightforward computation.

Define the stacked sample matrices \mathbb{X}_n and \mathbb{Y}_n :

$$\mathbb{X}_n \equiv \begin{bmatrix} X^{1\top} \\ X^{2\top} \\ \vdots \\ X^{n\top} \end{bmatrix}, \quad \mathbb{Y}_n \equiv \begin{bmatrix} Y^1 \\ Y^2 \\ \vdots \\ Y^n \end{bmatrix}. \quad (14)$$

Then, matrix calculus shows that we have

$$\mathbb{X}_n^\top \mathbb{X}_n = \sum_{i=1}^n X^i X^{i\top}, \quad \mathbb{X}_n^\top \mathbb{Y}_n = \sum_{i=1}^n X^i Y^i. \quad (15)$$

The OLS estimator can then equivalently be stated as

$$\hat{\beta}_n = (\mathbb{X}_n^\top \mathbb{X}_n)^{-1} (\mathbb{X}_n^\top \mathbb{Y}_n). \quad (16)$$

OLS Implementation (Contd.)

For the OLS covariance estimator $\hat{\Sigma}_n$, we define stacked residual vector:

$$\epsilon_n \equiv \mathbb{Y}_n - \mathbb{X}_n \hat{\beta}_n = \begin{bmatrix} Y^1 \\ Y^2 \\ \vdots \\ Y^n \end{bmatrix} - \begin{bmatrix} X^{1\top} \hat{\beta}_n \\ X^{2\top} \hat{\beta}_n \\ \vdots \\ X^{n\top} \hat{\beta}_n \end{bmatrix} = \begin{bmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_2 \\ \vdots \\ \hat{\epsilon}_n \end{bmatrix}. \quad (17)$$

By the same matrix calculus as before, we have

$$(\mathbb{X}_n \odot \epsilon_n)^\top (\mathbb{X}_n \odot \epsilon_n) = \sum_{i=1}^n X^i X^{i\top} \hat{\epsilon}_i^2, \quad (18)$$

where \odot denotes element-wise multiplication (*Hadamard product*). Then

$$\hat{\Sigma}_n = \frac{1}{n} (\mathbb{X}_n^\top \mathbb{X}_n)^{-1} \left[(\mathbb{X}_n \odot \epsilon_n)^\top (\mathbb{X}_n \odot \epsilon_n) \right] (\mathbb{X}_n^\top \mathbb{X}_n)^{-1}. \quad (19)$$

Notation: Strictly speaking, \odot is defined only for matrices of equal dimension. We abuse the notation here to denote multiplication between each row of the matrix \mathbb{X}_n with the corresponding component of the vector ϵ_n .

OLS Estimation in R

```
# Compute OLS estimates
XX_inv <- solve(t(X) %*% X)
XY <- t(X) %*% Y
beta <- XX_inv %*% XY

# Compute BLP estimates
blp_yx <- X %*% beta

# Compute standard error for beta
epsilon <- c(Y - blp_yx)
XX_eps2 <- t(X * epsilon) %*% (X * epsilon)
Sigma <- XX_inv %*% XX_eps2 %*% XX_inv / n
se <- sqrt(diag(Sigma)) / sqrt(n)
```

Note: There exists an OLS implementation in R – the `lm`-command. But importantly: Base-R does not implement the standard error of Corollary 2! So have some faith in your abilities and implement OLS yourself. See Problem 7 of Problem Set 4.

Today, we introduced OLS as an estimator for the $BLP(Y|X)$.

- ▷ Showed that it is consistent and asymptotically normal;
- ▷ Derived standard errors for subvector inference.

We're now well-equipped for causal analysis under selection on observables & common support:

- ▷ Defined interesting causal parameters using the all causes model;
- ▷ Showed identification of the CATE, ATT, ATU, and ATE;
- ▷ Concluded that if (W, X) is discrete, may use the binning estimator;
- ▷ If (W, X) is continuous/mixed, we can leverage OLS to obtain approximate results.