Simple Linear Regression
Part B: Ordinary Least Squares

Thomas Wiemann
University of Chicago

Econometrics
Econ 21020

Updated: April 27, 2022

Summary

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

- ▷ Showed that the BLP-coefficients are well-defined when $Var(X) > 0$;
- ▷ Discussed interpretation using Yitzhaki's Theorem;

The BLP and its coefficients α and 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;
- ▷ Apply it to study the California STAR data.

Outline

1. Ordinary Least Squares
2. Estimator Properties
   - Bias
   - Consistency
   - Asymptotic Distribution
3. Case Study: The California STAR Data

Outline

1. **Ordinary Least Squares**

2. Estimator Properties
    - ▷ Bias
    - ▷ Consistency
    - ▷ Asymptotic Distribution

3. Case Study: The California STAR Data

Ordinary Least Squares

Let Y and X be two random variables. Throughout, we consider a random sample $(Y_1, X_1), \ldots, (Y_n, X_n) \overset{iid}{\sim} (Y, X)$.

From Lecture 6A, we know that the BLP-coefficients are given by

$$\beta = \frac{Cov(Y, X)}{Var(X)}, \quad \text{and} \quad \alpha = E[Y] - E[X]\beta, \tag{1}$$

whenever $Var(X) > 0$.

This suggests sample analogue estimators

$$\hat{\beta}_n = \frac{\frac{1}{n}\sum Y_i X_i - (\frac{1}{n}\sum Y_i)(\frac{1}{n}\sum X_i)}{\frac{1}{n}\sum X_i^2 - (\frac{1}{n}\sum X_i)^2} \tag{2}$$

and

$$\hat{\alpha}_n = \frac{1}{n}\sum Y_i - (\frac{1}{n}\sum X_i)\hat{\beta}_n \tag{3}$$

Ordinary Least Squares (Contd.)

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

$$(\hat{\alpha}_n, \hat{\beta}_n) = \underset{\alpha,\beta\in\mathbb{R}}{\arg\min}\ \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - (\alpha + X_i\beta)\right)^2, \tag{4}$$

whenever $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 > 0$. In particular, we have:

$$R_n(\alpha,\beta) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i-(\alpha+X_i\beta)\right)^2 = \frac{1}{n}\sum Y_i^2 - 2\frac{1}{n}\sum Y_i(\alpha+X_i\beta) + \frac{1}{n}\sum(\alpha+X_i\beta)^2$$

$$\frac{\partial}{\partial\alpha}R_n(\alpha,\beta) = -2\frac{1}{n}\sum Y_i + 2\frac{1}{n}\sum(\alpha+X_i\beta) = 0$$

$$\iff \alpha = \frac{1}{n}\sum Y_i - \left(\frac{1}{n}\sum X_i\right)\beta$$

$$\frac{\partial}{\partial\beta}R_n(\alpha,\beta) = -2\frac{1}{n}\sum Y_iX_i + \frac{2}{n}\sum(\alpha+X_i\beta)X_i = -2\frac{1}{n}\sum Y_iX_i + 2\alpha\frac{1}{n}\sum X_i + 2\left(\frac{1}{n}\sum X_i^2\right)\beta$$

$$= -2\frac{1}{n}\sum Y_iX_i + 2\left(\frac{1}{n}\sum Y_i\right)\left(\frac{1}{n}\sum X_i\right) - 2\left(\frac{1}{n}\sum X_i\right)^2\beta + 2\left(\frac{1}{n}\sum X_i^2\right)\beta = 0$$

$$\iff \beta = \frac{\frac{1}{n}\sum Y_iX_i - \left(\frac{1}{n}\sum Y_i\right)\left(\frac{1}{n}\sum X_i\right)}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}$$

Ordinary Least Squares (Contd.)

For our analysis, it's useful to rewrite $\hat{\beta}_n$ using $\varepsilon_i \equiv Y_i - \text{BLP}(Y_i | X_i)$:

$$\hat{\beta}_n = \frac{\frac{1}{n}\sum Y_i X_i - \left(\frac{1}{n}\sum Y_i\right)\left(\frac{1}{n}\sum X_i\right)}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2} \qquad \Rightarrow Y_i = \alpha + X_i\beta + \varepsilon_i \tag{5}$$

$$= \frac{\frac{1}{n}\sum(\alpha + X_i\beta + \varepsilon_i)X_i - \left(\frac{1}{n}\sum(\alpha + X_i\beta + \varepsilon_i)\right)\left(\frac{1}{n}\sum X_i\right)}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}$$

$$= \frac{\alpha\frac{1}{n}\sum X_i + \left(\frac{1}{n}\sum X_i^2\right)\beta + \frac{1}{n}\sum \varepsilon_i X_i - \alpha\frac{1}{n}\sum X_i - \left(\frac{1}{n}\sum X_i\right)^2\beta - \left(\frac{1}{n}\sum \varepsilon_i\right)\left(\frac{1}{n}\sum X_i\right)}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}$$

$$= \left(\frac{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}\right)\beta + \frac{\frac{1}{n}\sum \varepsilon_i X_i - \left(\frac{1}{n}\sum \varepsilon_i\right)\left(\frac{1}{n}\sum X_i\right)}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}$$

$$= \beta + \frac{\frac{1}{n}\sum \varepsilon_i X_i - \left(\frac{1}{n}\sum \varepsilon_i\right)\left(\frac{1}{n}\sum X_i\right)}{\frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2}$$

Outline

1. Ordinary Least Squares

2. **Estimator Properties**
   - ▷ **Bias**
   - ▷ Consistency
   - ▷ Asymptotic Distribution

3. Case Study: The California STAR Data

Bias

Our analysis of the OLS estimator begins with its bias.

We assume here that X is continuous to avoid division by zero with positive probability (for $n \geq 2$).

- ▷ Flashbacks? Recall Problem 4d) of Problem Set 2 and the additional complications non-continuous X would introduce.

The bias of $\hat{\beta}_n$ when X is continuous is given by

$$\text{Bias}(\hat{\beta}_n) = E[\hat{\beta}_n] - \beta = E\left[\frac{\frac{1}{n}\sum X_i \varepsilon_i - \left(\frac{1}{n}\sum \varepsilon_i\right)\left(\frac{1}{n}\sum X_i\right)}{Var_n(X)}\right], \quad Var_n(X) = \frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2$$

$$= E\left[E\left[\frac{\frac{1}{n}\sum X_i \varepsilon_i - \left(\frac{1}{n}\sum \varepsilon_i\right)\left(\frac{1}{n}\sum X_i\right)}{Var_n(X)} \,\middle|\, (X_i)_{i=1}^n\right]\right] \tag{6}$$

$$= E\left[\frac{1}{Var_n(X)}\left(\frac{1}{n}\sum X_i E[\varepsilon_i \mid (X_i)_{i=1}^n] - \left(\frac{1}{n}\sum E[\varepsilon_i \mid (X_i)_{i=1}^n]\right)\left(\frac{1}{n}\sum X_i\right)\right)\right]$$

$$= E\left[\frac{1}{Var_n(X)}\left(\frac{1}{n}\sum X_i E[\varepsilon_i \mid X_i] - \left(\frac{1}{n}\sum E[\varepsilon_i \mid X_i]\right)\left(\frac{1}{n}\sum X_i\right)\right)\right]$$

Bias

Hence, for continuous X, 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 $E[\varepsilon_i|X_i] = 0$? When $E[Y|X]$ is linear.
- ▷ When do we know that $E[Y|X]$ is linear? When X is binary.
- ▷ Binary X is not continuous, so need to worry about division by zero...

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

- ▷ Importantly: Strong assumption are made along the way!
- ▷ We showed $\text{Bias}(\hat{\beta}_n) = 0$ if $E[Y|X]$ linear *and* X is continuous.
- ▷ With binary X, showed *conditional* unbiasedness in Problem Set 2.

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...

Outline

1. Ordinary Least Squares
2. **Estimator Properties**
   - Bias
   - **Consistency**
   - Asymptotic Distribution
3. Case Study: The California STAR Data

Consistency

Theorem 1 ensures OLS satisfies the minimum requirement: Consistency.

Theorem 1

Let Y and X be random variables such that $Var(X) > 0$, and let (α, β) denote the $BLP(Y|X)$-coefficients. If $(\hat{\alpha}_n, \hat{\beta}_n)$ are the OLS estimators constructed using $(Y_1, X_1), \ldots, (Y_n, X_n) \overset{iid}{\sim} (Y, X)$, then

$$\begin{bmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{7}$$

Since the OLS estimators are continuous functions of moments of (Y, X), we can prove this straightforwardly:

- Show $\hat{\beta}_n \xrightarrow{p} \beta$ using WLLN + CMT;
- Show $\hat{\alpha}_n \xrightarrow{p} \alpha$ using WLLN + CMT;
- Then use that marginal convergence in probability implies joint convergence in probability.

Consistency (Contd.)

Proof. $$\hat{\beta}_n = \frac{\frac{1}{n}\sum Y_i X_i - (\frac{1}{n}\sum Y_i)(\frac{1}{n}\sum X_i)}{\frac{1}{n}\sum X_i^2 - (\frac{1}{n}\sum X_i)^2}; \quad WTS: \hat{\beta}_n \xrightarrow{p} \beta$$

i) Four steps:

1. $A_n \equiv \frac{1}{n}\sum Y_i X_i$, $B_n \equiv \frac{1}{n}\sum Y_i$, $C_n \equiv \frac{1}{n}\sum X_i$, $D_n \equiv \frac{1}{n}\sum X_i^2$
2. $g(a,b,c,d) = \frac{a - b\cdot c}{d - c^2}$
3. By WLLN, $A_n \xrightarrow{p} E[YX]$, $C_n \xrightarrow{p} E[X]$, $B_n \xrightarrow{p} E[Y]$, $D_n \xrightarrow{p} E[X^2]$
4. By CMT, $g(A_n, B_n, C_n, D_n) \xrightarrow{p} \frac{E[YX] - E[Y]E[X]}{E[X^2] - E[X]^2} = \frac{Cov(Y,X)}{Var(X)} = \beta$

   $\forall\ Var(X) > 0.$

ii) $\hat{\alpha}_n = \frac{1}{n}\sum Y_i - (\frac{1}{n}\sum X_i)\hat{\beta}_n$

By part i) + CMT, $\hat{\alpha}_n \xrightarrow{p} \alpha$. □

Outline

1. Ordinary Least Squares
2. **Estimator Properties**
   - Bias
   - Consistency
   - **Asymptotic Distribution**
3. Case Study: The California STAR Data

Asymptotic Distribution

OLS is asymptotically normal, which is useful for approximating its sampling variance in practice. Theorem 2 focuses on $\hat{\beta}_n$ for conciseness.

Theorem 2

Let Y and X be random variables such that $Var(X) > 0$, and let (α, β) denote the $BLP(Y|X)$-coefficients. If $\hat{\beta}_n$ is OLS estimator for β constructed using $(Y_1, X_1), \ldots, (Y_n, X_n) \overset{iid}{\sim} (Y, X)$, then

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \xrightarrow{d} N\left(0, \frac{E\left[\varepsilon^2 \left(X - E[X]\right)^2\right]}{Var(X)^2}\right), \tag{8}$$

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

***Note**: Classical treatments of OLS would refer to the variance expression in Theorem 2 as the asymptotic variance under heteroskedasticity. By this, they mean the general case that $Var(\epsilon|X)$ is allowed to vary with X. As ε is simply the BLP-residual, there is no obvious reason to think otherwise. We thus don't highlight this as a special case.*

Asymptotic Distribution (Contd.)

Proof.

$$\hat{\beta}_n - \beta = \frac{\frac{1}{n}\sum \varepsilon_i X_i - (\frac{1}{n}\sum \varepsilon_i)(\frac{1}{n}\sum X_i)}{Var_n(X)} = \frac{1}{Var_n(X)} \begin{bmatrix} 1 \\ -\frac{1}{n}\sum X_i \end{bmatrix}^T \begin{bmatrix} \frac{1}{n}\sum \varepsilon_i X_i \\ \frac{1}{n}\sum \varepsilon_i \end{bmatrix}$$

i) $$\sqrt{n}\begin{bmatrix} \frac{1}{n}\sum \varepsilon_i X_i \\ \frac{1}{n}\sum \varepsilon_i \end{bmatrix} = \sqrt{n}\left(\begin{bmatrix} \frac{1}{n}\sum \varepsilon_i X_i \\ \frac{1}{n}\sum \varepsilon_i \end{bmatrix} - \begin{bmatrix} E[\varepsilon X] \\ E[\varepsilon] \end{bmatrix}\right) \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \underbrace{Var(\varepsilon X)}_{=E[\varepsilon^2 X^2]} & \underbrace{Cov(\varepsilon X, \varepsilon)}_{= E[\varepsilon^2 X] - \underbrace{E[\varepsilon X]}_{=0}\underbrace{E[\varepsilon]}_{=0}} \\ Cov(\varepsilon X, \varepsilon) & \underbrace{Var(\varepsilon)}_{=E[\varepsilon^2]} \end{bmatrix}\right)$$

ii) $$\frac{1}{\frac{1}{n}\sum X_i^2 - (\frac{1}{n}\sum X_i)^2} \begin{bmatrix} 1 \\ -\frac{1}{n}\sum X_i \end{bmatrix}^T \xrightarrow{p} \frac{1}{Var(X)} \begin{bmatrix} 1 \\ -E[X] \end{bmatrix}^T, \quad \forall\, Var(X) > 0, \text{ by WLLN + CMT.}$$

Slutsky's, $$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \frac{1}{Var(X)} \begin{bmatrix} 1 \\ -E[X] \end{bmatrix}^T \mathcal{N}\left(0, \begin{bmatrix} E[\varepsilon^2 X^2] & E[\varepsilon^2 X] \\ E[\varepsilon^2 X] & E[\varepsilon^2] \end{bmatrix}\right)$$

Asymptotic Distribution (Contd.)

Variance: $\frac{1}{Var(X)} \begin{bmatrix} 1 \\ -E[X] \end{bmatrix}^T \begin{bmatrix} E[\varepsilon^2 X^2] & E[\varepsilon^2 X] \\ E[\varepsilon^2 X] & E[\varepsilon^2] \end{bmatrix} \begin{pmatrix} 1 \\ -E[X] \end{pmatrix} \frac{1}{Var(X)}$

$$= \frac{1}{Var(X)^2} \left(E[\varepsilon^2 X^2] + E[\varepsilon^2] E[X]^2 - 2 E[\varepsilon^2 X] E[X] \right) = \circledast$$

$$\begin{aligned} E[\varepsilon^2 X^2] &= E[\varepsilon^2 (X - E[X] + E[X])^2] \\ &= E[\varepsilon^2 (X - E[X])^2] + 2 E[\varepsilon^2 (X - E[X]) E[X]] + E[\varepsilon^2 E[X]^2] \\ &= E[\varepsilon^2 (X - E[X])^2] + 2 E[\varepsilon^2 X] E[X] - 2 E[\varepsilon^2] E[X]^2 + E[\varepsilon^2] E[X]^2 \end{aligned}$$

$$\Rightarrow \circledast = \frac{1}{Var(X)^2} E[\varepsilon^2 (X - E[X])^2]$$

□

Standard Error

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 and X be random variables such that $Var(X) > 0$, and let (α, β) denote the BLP$(Y|X)$-coefficients. If $\hat{\beta}_n$ is OLS estimator for β constructed using $(Y_1, X_1), \ldots, (Y_n, X_n) \overset{iid}{\sim} (Y, X)$, then

$$\frac{\left(\hat{\beta}_n - \beta\right)}{se(\hat{\beta}_n)} \xrightarrow{d} N(0, 1), \tag{9}$$

where

$$se(\hat{\beta}_n) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2 \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right)^2}}{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right)}, \tag{10}$$

and $\hat{\varepsilon}_i = Y_i - X_i\hat{\beta}_n$.

Standard Error (Contd.)

We prove Theorem 3 for X such that $E[X] = 0$. The more general result holds, of course, but its proof requires even more painful algebra.

Proof.

WTS: $\frac{1}{\left(\frac{1}{n}\sum X_i^2 - (\frac{1}{n}\sum X_i)^2\right)^2} \frac{1}{n}\sum \hat{\varepsilon}_i^2 X_i^2 \xrightarrow{p} \frac{E[\varepsilon^2 X^2]}{Var(X)^2}$

$$\frac{1}{n}\sum \hat{\varepsilon}_i^2 X_i^2 = \frac{1}{n}\sum (\hat{\varepsilon}_i - \varepsilon_i + \varepsilon_i)^2 X_i = \underbrace{\frac{1}{n}\sum(\hat{\varepsilon}_i - \varepsilon_i)^2 X_i^2}_{\equiv A_n} + \underbrace{2\frac{1}{n}\sum(\hat{\varepsilon}_i - \varepsilon_i)\varepsilon_i X_i^2}_{\equiv B_n} + \underbrace{\frac{1}{n}\sum \varepsilon_i^2 X_i^2}_{\xrightarrow{p} E[\varepsilon^2 X^2] \text{ by WLLN}}$$

Note: $\hat{\varepsilon}_i - \varepsilon_i = \hat{\alpha}_n - \alpha + X_i(\hat{\beta}_n - \beta)$

Then,
$$A_n = \frac{1}{n}\sum(\hat{\alpha}_n - \alpha)^2 X_i^2 + 2\frac{1}{n}\sum(\hat{\alpha}_n - \alpha)(\hat{\beta}_n - \beta)X_i^3 + \frac{1}{n}\sum(\hat{\beta}_n - \beta)^2 X_i^4$$
$$= (\hat{\alpha}_n - \alpha)^2 \frac{1}{n}\sum X_i^2 + 2(\hat{\alpha}_n - \alpha)(\hat{\beta}_n - \beta)\frac{1}{n}\sum X_i^3 + (\hat{\beta}_n - \beta)^2 \frac{1}{n}\sum X_i^4$$
$$\xrightarrow{p} 0^2 \cdot E[X^2] + 2 \cdot 0 \cdot 0 \cdot E[X^3] + 0^2 \cdot E[X^4] = 0 \quad \text{by WLLN + Thrm 1 + CMT}$$

Similarly,
$$B_n = (\hat{\alpha}_n - \alpha)\frac{2}{n}\sum \varepsilon_i X_i^2 + (\hat{\beta}_n - \beta)\frac{2}{n}\sum \varepsilon_i X_i^3 \xrightarrow{p} 0 \cdot E[\varepsilon X^2] + 0 \cdot E[\varepsilon X^3] = 0$$

by WLLN + Thrm 1 + CMT. □

OLS Estimation in R

This completes the theoretical analysis: Now we need to implement it!

OLS Estimation in R

```
# Compute OLS estimates
beta <- cov(y, x) / var(x)
alpha <- mean(y) - mean(x) * beta

# Compute BLP estimates
blp_yx <- alpha + x * beta

# Compute standard error for beta
epsilon <- y - blp_yx
se_numer <- sqrt(mean(epsilon^2 * (x - mean(x))^2))
se <- (se_numer / var(w)) / sqrt(n)
```

Note: *There exists an OLS implementation in R – the* `lm`*-command. But importantly: Base-R does not implement the standard error of Theorem 3! So have some faith in your abilities and implement OLS yourself. See Problem 7 of Problem Set 3.*

Outline

1. Ordinary Least Squares
2. Estimator Properties
   - ▷ Bias
   - ▷ Consistency
   - ▷ Asymptotic Distribution
3. **Case Study: The California STAR Data**

The California STAR Data

The California Standardized Testing and Reporting (STAR) data:

- ▷ All 420 K-6 and K-8 districts in California for 1998/99;
- ▷ Average reading and math scores on 5th grade standardized test;
- ▷ Average school characteristics: E.g., student-teacher ratios, income.

Suppose we are interested in assessing the relationship between students' math scores and student-teacher ratios. For this purpose, let (Y, W, U) be random variables, where $Y = g(W, U)$ and

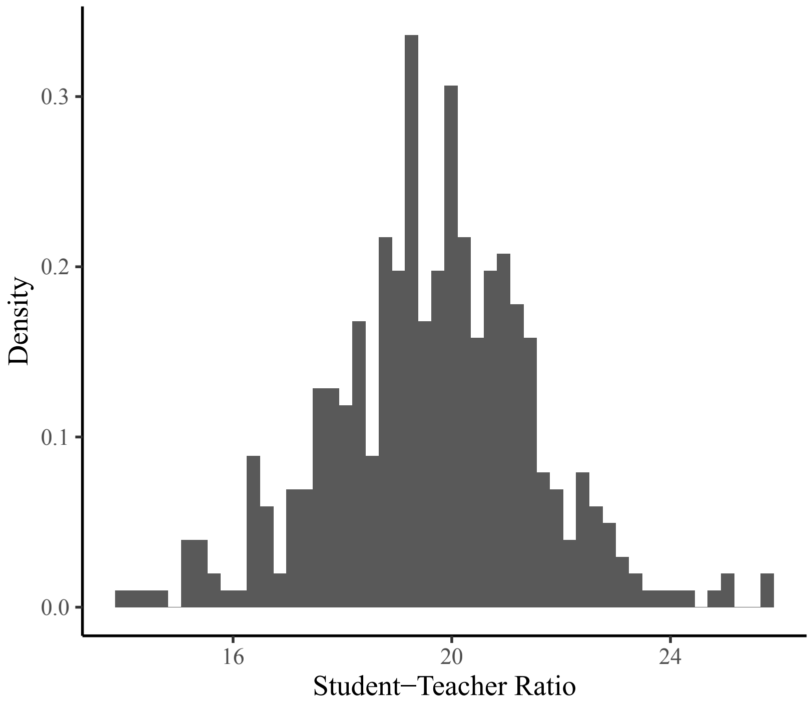
- ▷ $Y \equiv$ students' average math scores in a California district;
- ▷ $W \equiv$ the average student-teacher ratio of the district;
- ▷ $U \equiv$ all determinants of Y other than W.

We assume that the STAR data is the realization of the sample $(Y_1, W_1), \ldots, (Y_{420}, W_{420}) \overset{iid}{\sim} (Y, W)$.

Note: *The data is used as an example in Stock and Watson (2015) and readily available on their website. You can find the dataset on Canvas:* `star98.csv`.

The California STAR Data (Contd.)

Figure 1: Histogram of the Student-Teacher Ratios in the STAR Data



Notes. You can find the code generating the figure on GitHub: `example_star98.R`.

The California STAR Data (Contd.)

Since the student-teacher ratio is (nearly) continuous, need BLP:

- Conditional expectations can't be estimated directly;
- Choose to focus on best linear approximation instead for feasibility.

In particular, we consider estimating the BLP$(Y|W)$-coefficient β.

- β captures the approximate expected change in students' math score associated with an additional student per teacher.

Our OLS estimate and the corresponding standard error are

$$\hat{\beta}_n \approx -1.936 \quad , \text{ and } se\left(\hat{\beta}_n\right) \approx 0.519$$

Hence, a 95% confidence interval is given by

$$c_n \approx [-2.956, -0.921]$$

Notes. You can replicate these estimates with the code on GitHub: `example_star98.R`.

The California STAR Data (Contd.)

W/o additional assumptions, a correct interpretation of $\hat{\beta}_n$ is:

- ▷ An additional student per teacher is associated with an approximate expected change of students' math score by -1.939 points.

It's important to qualify your interpretation:

- ▷ "approximate" to highlight that $\text{BLP}(Y|W) \neq E[Y|W]$;
- ▷ "associated" to emphasize that β is a descriptive parameter.

The California STAR Data (Contd.)

Suppose that a peer who has not yet taken Econ 21020 sees your estimate $\hat{\beta}_n$ and concludes:

- "Just as I expected. Low student-teacher ratios are causing disastrous learning outcomes. We should hire more math teachers! "

What's wrong with this interpretation?

- It's is a *causal* interpretation for a *descriptive* parameter: Not good!

From Lecture 6A, we know that under random assignment (RA), a causal interpretation of β would be appropriate. What does RA imply here?

- $W \perp\!\!\!\perp U$, i.e., that the student-teacher ratio is independent of all other determinants of students' average math scores.

Do you think random assignment is a plausible assumption here?

- Probably not. E.g., wealthier districts may hire both 1) better teachers, and 2) more teachers per student.

The California STAR Data (Contd.)

To convince your peer that RA is likely to be implausible in the context of the California STAR data, you consider conducting a balance test.

Let the random variable X denote districts' average parental income.

- RA implies $W \perp\!\!\!\perp U \Rightarrow W \perp\!\!\!\perp X$;
- Hence, $W \perp\!\!\!\perp X \Rightarrow Cov(W, X) = 0 \Rightarrow \beta_X = 0$.

Here: β_X is the BLP$(X|W)$-coefficient.

A balance test may thus consider testing

$$H_0 : \beta_X = 0 \quad \text{versus} \quad H_1 : \beta_X \neq 0.$$

Rejecting H_0 would provide evidence that $W \not\perp\!\!\!\perp X$.

- Potential cause for worry (but of course: Type I errors exist!).

Failure to reject H_0 would *not* provide evidence that $W \perp\!\!\!\perp X$.

- May be because of $W \perp\!\!\!\perp X$ or low power of the test!

The California STAR Data (Contd.)

Our OLS estimate of β_X and the corresponding standard error are

$$\hat{\beta}_{X,n} \approx -0.887 \quad , \text{ and } se\left(\hat{\beta}_{X,n}\right) \approx 0.223$$

The test statistic for a two-sided hypothesis test is

$$T_n \approx \left|\frac{-0.887}{0.223}\right| \approx 3.977$$

The corresponding p-value is

$$p\text{-value} \approx 2\left(1-\Phi\left(3.977\right)\right) \approx 0.00007$$

Hence, on a 5% significance level, we reject H_0.

- On a 5% significance level, there is sufficient evidence to reject that the approximate expected change in average parental income associated with a unit-change in the student-teacher ratio is zero.

As a consequence, we may deem RA implausible for the STAR context.

Summary

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

- Showed that it is consistent and asymptotically normal;
- Considered the California STAR data as an example.

We're now well-equipped for causal analysis under random assignment:

- Defined interesting causal parameters using the all causes model;
- Showed identification of the ATE;
- Concluded that if W is discrete, may use the binning estimator;
- If W is continuous, can leverage OLS to obtain approximate results.

But as we saw, random assignment is not always plausible or desired:

- Rarely applicable in non-experimental settings;
- Does not allow to study selection into treatment;

Next, we study a new identifying assumption: *Selection on Observables*.

References

Stock, J. H. and Watson, M. W. (2015). Introduction to econometrics.