

# ECON 7201

## Applied Econometrics

### Assignment 3

#### Due Date

Friday November 7, 2025 at 11:59 PM

#### Directions

Answer each question clearly and concisely. Write equations using LaTeX where appropriate and explain all assumptions in your own words. Submit both a PDF and Quarto file to the nexus assignment portal.

#### Git and GitHub

1. (a) Create a new R project in your **econ\_3201** directory called **assignment\_3**.  
(b) Download the assignment PDF and Quarto file the **assignment\_3** folder.  
(c) Commit and push the changes to your **econ\_3201** repository on [GitHub.com](https://GitHub.com). **Ans**

Changes have been committed to <https://github.com/Emmanuel-spec493>

#### Conceptual Foundations

1. (5 pts) Define the *potential outcomes framework*. Clearly explain what is meant by  $Y_i(1)$ ,  $Y_i(0)$ , and  $D_i$ .

**Ans.**

$Y_i(1)$  is the potential outcome if individual i receives treatment

$Y_i(0)$  is the potential outcome if individual i does not receive treatment

$D_i$  is the treatment variable

$D_i = 1$  if an individual receives treatment

$D_i = 0$  if an individual does not receive treatment.

2. (5 pts) State and explain the **Stable Unit Treatment Value Assumption (SUTVA)**. Why is this assumption critical for causal inference?

**Ans.** The Stable Unit Treatment Value Assumption states that, the potential outcomes of each  $i$  depends solely on that's unit's own treatment status, and there are no different or varying versions of the treatment.

$$Y_i(d_1, d_2, \dots, d_n) = Y_i(d_i).$$

the treatment  $d_i$  has a single,well-defined version.

Stable Unit Treatment Value Assumption(SUTVA) has two meaning.

(a).

No hidden versions of the treatment: The treatment and control conditions are always uniform.  
ie. any individual  $i$  treated receives the same version of the treatment.

- (b) No Interference between units: Each individual  $i$ 's treatment depends only on their own treatment.

SUTVA is crucial for causal inference because SUTVA guarantees that causal effects are clearly defined for each individual, remain consistent across the study, and can be meaningfully interpreted — making it a cornerstone of reliable causal inference.

3. (5 pts) What is the **Conditional Independence Assumption (CIA)**? Write it formally and explain its meaning in words.

**Ans.** Conditional Independence Assumption (CIA) means that after controlling for all relevant observed factors, the assignment of individuals to the treatment or control group is unrelated to what their outcomes would be in either case.

It is written formally as  $(Y_{oi}, Y_{1i}) \perp\!\!\!\perp c_i | X_i$

This means that after controlling for  $X$ ,the treatment assignment is as good as random-the treated and untreated groups are comparable in expectation.

4. (5 pts) Describe the **Overlap (Common Support)** condition and its practical importance.

**Ans .** It is formally defined as  $0 < P(D_i = 1 | X_i) < 1$  for all  $X_{-i}$

for every combination of observed characteristics  $X_i$  ,  $D_i = 1$  if an individual receives treatment

$D_i = 0$  if an individual does not receive treatment

5. (5 pts) Explain why the **Average Treatment Effect on the Treated (ATT)** may differ from the **Average Treatment Effect (ATE)**. Ans.

The Average Treatment Effect on the Treated may differ from the Average Treatment effect because of the selection bias problem. If selection is not randomised we get a selection bias leading to the difference in the two.

## Identifying Causal Effects (25 points)

1. (5 pts) Suppose treatment \$ D\_i \$ is assigned completely at random. Show that

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] = ATE$$

Explain why randomization justifies this equality.

**Ans.**

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] + E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]$$

Under Randomisation

$$E[Y_{i0}|D_i = 1] = E[Y_{i0}|D_i = 0].$$

Hence we can replace

$$E[Y_{i0}|D_i = 1] \text{ with } E[Y_{i0}|D_i = 0] \text{ and still get ATE.}$$

$$\text{Therefore } E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0] = E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0] = ATE$$

2. (10 pts) Suppose treatment depends on observed covariates \$ X\_i \$, such that \$ (Y\_i(1), Y\_i(0)) | D\_i | X\_i \$ Derive an expression for ATE using the **law of iterated expectations**, and interpret it.

**Ans**

$$ATE = E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0]$$

Applying the law of Iterated Expectations.

$$E[Y] = E[E|Y|X]$$

$$E[Y_{i1}] = E[E|Y_{i1}|X_I], E[Y_{i0}] = E[E|Y_{i0}|X_I]$$

$$ATE = E[E|Y_{i1}|X_i] - E[E|Y_{i0}|X_i]$$

Under Conditional Independence Assumption (CIA)

$$E[Y_{i1}|X_i] = E[Y_1|D_i = 1|X_i] =$$

$$E[Y_{i0}|X_i] = E[Y_1|D_i = 0|X_i]$$

$$\text{ATE} = E[E[Y_1|D_i = 1|X_i] - E[Y_1|D_i = 0|X_i]]$$

The ATE can be computed as a weighted average of conditional differences in outcomes across subgroups defined by covariates  $X_i$

3. (10 pts) Explain how *selection bias* arises if the CIA fails. Use the decomposition:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \text{ATE} + \text{Selection Bias.}$$

### Ans

When CIA fails, treatment assignment is correlated with unobserved determinants of potential outcomes. When that happens  $E[Y_{o1}|D_i = 1|X_i]$  is not equal to  $E[Y_{o1}|D_i = 0|X_i]$

$$E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0] = \text{ATE} + \text{Selection bias}$$

$$E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0] = E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] + E[Y_{o1}|D_i = 1|X_i] - E[Y_{o1}|D_i = 0|X_i]$$

$$\text{Selection bias} = E[Y_{o1}|D_i = 1|X_i] - E[Y_{o1}|D_i = 0|X_i]$$

Once these two terms are unequal because CIA has failed,  $E[Y_{o1}|D_i = 1|X_i]$  is not equal to  $E[Y_{o1}|D_i = 0|X_i]$  and selection bias exist.

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \text{ATE} + \text{Selection Bias.}$$

## Applied Exercise in R (30 points)

Use R to simulate data, estimate treatment effects, and interpret your results. Include all R code and output.

```
set.seed(123)

n <- 1000
X <- rnorm(n)
D <- ifelse(X + rnorm(n) > 0, 1, 0)
Y0 <- 2 + 0.5 * X + rnorm(n)
Y1 <- Y0 + 3 + 0.5 * X
Y <- ifelse(D == 1, Y1, Y0)
data <- data.frame(Y, D, X)
```

1. (5 pts) Compute the difference in average outcomes between treated and untreated. Interpret the result in terms of potential bias.

## 2. (5 pts)

- (a) Estimate the treatment effect controlling for (X):

$$Y_i = \alpha + \tau D_i + \beta X_i + \varepsilon_i$$

- (b) Compare the estimate of ( ) to the true treatment effect of 3.

## 3. (5 pts) Propensity Score Matching

- (a) Estimate treatment effects using matching. Search `?matchIt` in the console  
 (b) Compare the matched estimate to OLS and the naïve difference.

## 4. (5 pts) Simulate a violation of the Conditional Independence Assumption by introducing an unobserved confounder ( U\_i ):

```

U <- rnorm(n)
D <- ifelse(X + U + rnorm(n) > 0, 1, 0)
Y0 <- 2 + 0.5 * X + 0.5 * U + rnorm(n)
Y1 <- Y0 + 3 + 0.5 * X
Y <- ifelse(D == 1, Y1, Y0)
bias_model <- lm(Y ~ D + X, data = data.frame(Y, D, X))
summary(bias_model)

```

Visualize and discuss how this unobserved confounder biases the estimated treatment effect.

```

plot(X, Y, col = ifelse(D==1, "red", "blue"), pch = 16,
      main = "Outcome vs Covariate with Confounding",
      xlab = "X", ylab = "Y")
legend("topleft", legend = c("Treated", "Untreated"),
      col = c("red", "blue"), pch = 16)

```