Eugene Baldonado
SID: 862070316

ECON 108: Spring 2022
Midterm Exam
Due on Canvas at 9:30AM on Tuesday, May 10th

1. The university is currently trying to figure out how much to value Teaching Evaluation scores. A common concern is that instructors can "buy scores" by having easy classes and giving out lots of A's. To explore this, you gather evaluation information on all classes at UCR in the fall of 2021. Specifically, you have the average evaluation score (Y) of each class i on a scale of 1 to 5 and the GPA (X) in the class. There are 250 classes during the term.

You regress evaluation score on GPA and get the following result:

$$\widehat{Eval_i} = \underset{(0.36)}{\widehat{1.54}} + \underset{(0.45)}{\widehat{0.73}}\, GPA_i\,,\ R^2 = 0.42$$

a. What is the interpretation of $\beta_0 = 1.54$ and $\beta_1 = 0.73$?

b. Assume that $E[u|\ GPA_i] \neq 0$. Discuss the elements of the error term could cause this to be the case. In other words, identify elements of the error term that could potentially be related to GPA_i.

c. Determine whether the predominant bias from $E[u|\ GPA_i] \neq 0$ causes the estimated $\widehat{\beta}_1$ to be biased towards or away from zero. Support your answer with the sign of the bias and connect your answer to a simple graph.

d. Because of the bias that you are concerned with, you decide to use the natural log of GPA instead of the GPA. You get the following result:

$$\widehat{Eval_i} = \underset{(1.01)}{\widehat{2.54}} + \underset{(0.10)}{\widehat{0.23}}\ln GPA_i\,,\ R^2 = 0.42$$

Interpret the new coefficients, β_0 and β_1. Do you think that this adjustment to the regression alleviates your concerns about bias? Support your answer.

a) The average Teaching Evaluation Score of each class (i) is $\hat{\beta}_0(1.54)$ when the GPA for that class is $\emptyset$. (Estimated $\hat{\beta}_0$ s.s. at 5% level)

$\frac{\hat{\beta}_0}{SE(\hat{\beta}_0)}$ $t = \frac{1.54}{.36} > 2$

If the GPA for a class increases by 1, the Teaching Evaluation score will increase by $\hat{\beta}_1 (0.73)$, on average.

$\frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$ $t = \frac{0.73}{0.45} = 1.62$ (This estimated $\hat{\beta}_1$ is not statistically significant at the 5% level)

$1.62 < 1.96$

b) There are a # of elements that affect GPA that could potentially be in the error term; examples are sleeping hours, stress levels, study hours for each class.

Example: True Model: $Eval_i = \beta_0 + \beta_1 GPA + \beta_2 \text{Stress levels} + u_i$

X_2 = Stress levels

c) $\hat{\beta}_1 = \beta_1 + \beta_2 \hat{\delta}_1$, where Stress Levels$_i = \delta_0 + \delta_1 GPA_i + \varepsilon_i$

1. $\hat{\delta}_1$: Direct relationship between S. Levels & GPA $\Rightarrow \hat{\delta} < 0$
2. β_2: Direct relationship between S. levels & Eval $\Rightarrow \beta_2 < 0$
3. Bias (C) $= \beta_2 \cdot \hat{\delta}_1 > 0$ (Positive!)

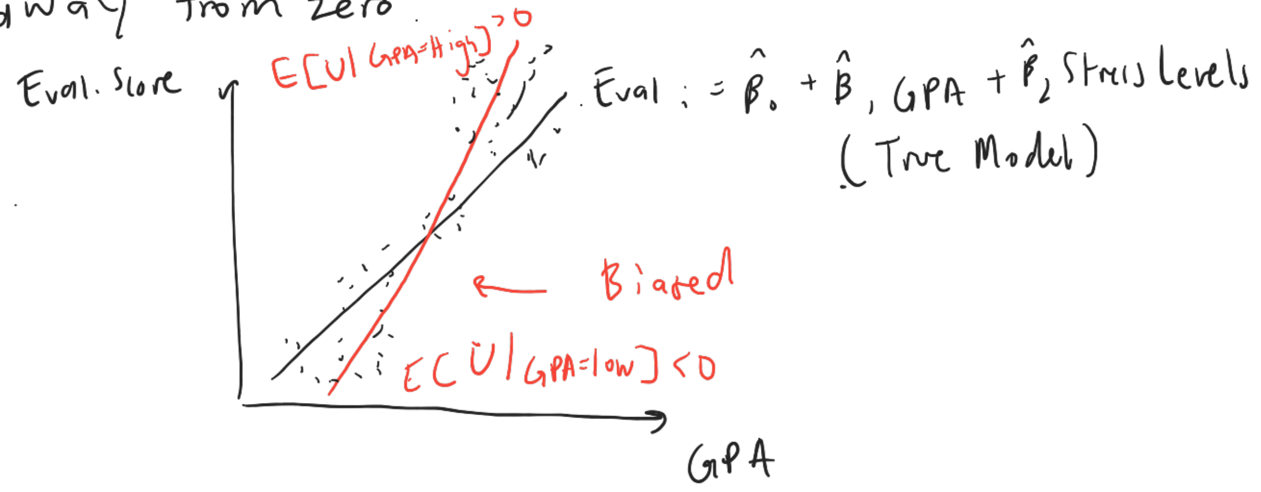
→ Continued

continued

Continue to answer Question 1 on this page.

c) Since $\hat{\beta}_1 > 0$, $c > 0$ $\Rightarrow$ $\beta_1 = \hat{\beta}_1 - \beta_2 \hat{\delta}_1$

Therefore true β_1 is closer to zero and $\hat{\beta}_1$ is biased away from zero.



d) The average Teaching Evaluation Score of each class is $\hat{\beta}_0$ (2.54) when the natural log of the GPA for that class is $\emptyset$.

A 1% change in GPA is associated with a (0.01 × 0.23) 0.0023 change in Teaching Evaluation Score, on average.

(Both are statistically significant ($\frac{2.54}{1.01} > 2$, $\frac{.23}{.10} > 2$, β_0, β_1) at the 5% level)

Although $\hat{\beta}_1$ is now statistically significant at the 5% level, this adjustment to the regression does not alleviate biases, this can be seen as the R^2 stays at 0.42, indicating that log of GPA still is not explaining much in the variation of average teaching evaluation score.

2. With stay-at-home orders being lifted over the past few months you have noticed that traffic is starting increase again and most drivers are on their phone! This leads you to wonder if those cell-phone and driving laws were effective at reducing traffic accidents. It turns out that a number of states enacted "no-texting laws" in 2018.

You gather data on state-level traffic accident rates per 100,000 people in the state and the texting laws for all states in the years 2015 to 2021 and run the following regression:

$$\begin{array}{lcllll} \widehat{AccidentRate}_{it} & = & 41.05 & -\ 5.96 Treatment_i & -\ 3.82 After_t & -\ 8.38 Treatment_i \times After_t \\ & & (6.23) & \quad (4.71) & \quad (2.17) & \quad (1.51) \end{array}$$

where $Treatment$ is equal to 1 if state i is assigned to the treatment group and 0 otherwise (control group). The variable $After$ is equal to 1 if the time period t is 2018 or after and 0 otherwise (2015 through 2017). The interaction variable is equal to 1 if $Treatment \times After$ is equal 1 and 0 otherwise.

a. Interpret the difference-in-difference estimator in the regression.

b. Show the results of the difference-in-difference regression estimates graphically. Specifically note how each regression coefficient is represented on the graph and what the average is for each group.

c. Discuss whether or not there is omitted variable bias in your regression. Given your answer, would you conclude that the texting laws have been effective in reducing the accident rate? Support your claim.

a)

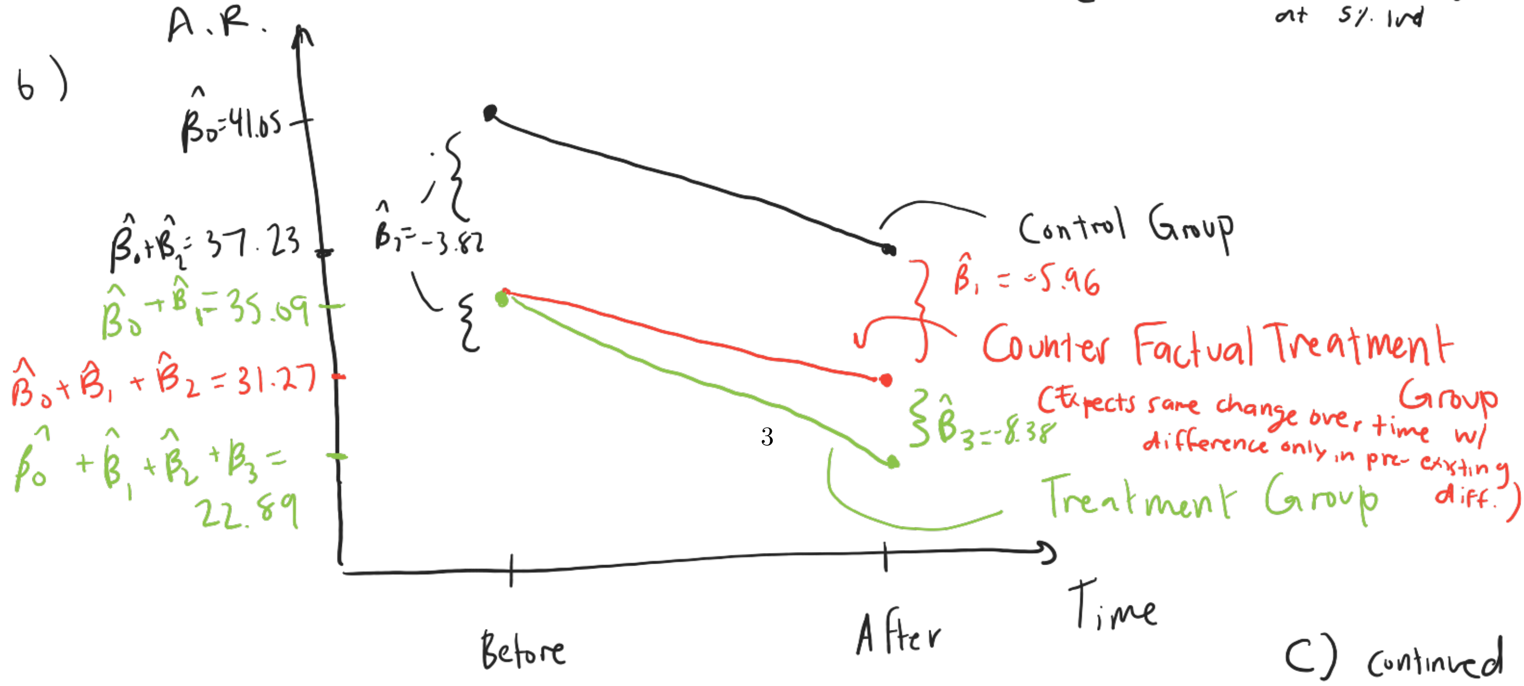
		⓪ Before ($t<2018$)	① After ($t \geq 2018$)
Treatment	⓪ Control Group	$E[A.R \mid T=0, A=0] = \hat{\beta}_0 = 41.05$	$E[A.R. \mid T=0, A=1] = \hat{\beta}_0 + \hat{\beta}_2 = 41.05 - 3.82 = 37.23$
	① Treatment Group	$E[A.R \mid T=1, A=0] = \hat{\beta}_0 + \hat{\beta}_1 = 41.05 - 5.96 = 35.09$	$E[A.R \mid T=1, A=1] = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \beta_3$ $41.05 - 5.96 - 3.82 - 8.38 = 22.89$

A.R = Accident Rate
T = Treatment
A = After

β_0 = Average accident rate for a state before "n.t.l" (no texting laws) (S.S.)
β_1 = Average effect of "n.t.l" on Accident Rate (Treatment)
β_2 = Average effect of time on Accident Rate (S.S)
β_3 = Average effect of interacting terms time + "n.t.l" on Accident Rate (S.S)

(S.S = Statistically significant at 5% lvl)

b)



C) continued

Continue to answer Question 2 on this page.

c) There definitely can be omitted variable bias as there isn't control for per-state differences in accident rates. For example, one state may have narrower roads, different driving culture, on avg. higher accident rates across the years, etc. Although the treatment group states did have lower Accident rates than the control group on average, given the omitted variable biases, the conclusion that texting laws have been effective in reducing the accident rate cannot be confirmed.

3. After your study from question 2, you realize that laws are irrelevant if cell phone coverage is not available. As a result, you decide to run a regression of traffic accident rates in state i in year t on $WifiFraction$ which is the fraction of a state that has high-speed wireless coverage (3G or 4G). Again assume you have data on all states for the years 2015 to 2021 and that you run the following regression:

$$\widehat{AccidentRate_{it}} = \underset{(15.36)}{\widehat{50.60}} - \underset{(0.21)}{\widehat{0.27}} \times WifiFraction_{it}.$$

The dependent variable, $AccidentRate_{it}$, is the accident rate per 100,000 in state i in year t. The independent variable of interest is $WifiFraction_{it}$, which is the fraction of state i with 3G or 4G coverage in year t.

a. Interpret the coefficient of interest ($\beta_1 = -0.27$). Does this estimator suffer from omitted variable bias? Support your answer.

b. You decide to add state and year fixed effects by running an $(n-1)$ and $(T-1)$ binary variable regression:

$$\widehat{AccidentRate_{it}} = \underset{(52.11)}{\widehat{18.26}} + \underset{(0.15)}{\widehat{0.52}} \times WifiFraction_{it} + \gamma_2 D2_i + ... + \gamma_{50} D50_i + \delta_1 B2016 + ...\delta_6 B2021.$$

What are the state fixed effects capturing? What are the time fixed effects capturing? Provide an example for each fixed effect.

c. Interpret γ_2, the coefficient on the dummy variable $D2$. Interpret δ_4, the coefficient on the dummy variable $B2019$.

d. Interpret the coefficient of interest ($\beta_1 = 0.52$). Explain how the interpretation is different than the regression without fixed effects and how the coefficient could have changed so much.

a) β_1 = The average effect of a 1-unit increase in the fraction of wireless coverage on the accident rate AND the effects of that individual states characteristics and time characteristics

Yes, the estimator suffers from omitted variable bias, as certain states may have more 4G antennas or are more urban than rural, etc.

b) State Fixed Effects captures the per state differences eliminating the risk of bias from omitted factors that vary per state; an example of this can be if the state is more urban or rural.

Time Fixed effects capture the omitted variables that vary over time that may cause biases, an example of this would be all increases in connection as # of sattelites delivering 4G increases. over the years.

c → continued

c) γ_2 captures the state-fixed effect from the State D_2 (if it is the state this Dummy Variable = 1) on the Accident Rate on average. δ_y is the time-fixed effect from the year 2019 on the accident rate on average.

d) 0.52 is the estimate on how much the Accident Rate changes when a specific state has 1 more fraction of the state covered with wireless connection, on average. (controlling for State-fixed effects)

The interpretation of the coefficient is different as now the controls for time fixed effects & state fixed effects are added. Instead of the OLS regression coefficient containing all the characteristics, now the coefficient estimates for a specific state. The coefficient also becomes stronger and statistically significant at the 5% level

$\left(\frac{0.52}{0.15} > 2\right.$ in controlling for State & Time fixed effects, v.s.

$\left.\frac{0.27}{0.21} \not> 2\right)$.

4. In this question, you will use a dataset that I provide on Canvas (teachingevals.dta), upload the data to Stata and run regressions that allow you to answer the following questions. You should create a table using *outreg*2 and submit the table and do-file, along with your test answers.

The main question you are exploring is whether there is discrimination in teaching evaluations.

In the panel data, you are given the following variables:

courseid: a unique ID for a course (for example, ECON 108 may be coded as 1368). A course can be taught multiple times.

term: a specific quarter (ranging from 1 to 18)

enrollment: the enrollment in the course for the specific quarter

*A*19: the average evaluation score for question 19 - ”The course overall as a learning experience was excellent” - and rated on a scale of 1 (lowest) to 5 (highest).

response: the evaluation response rate in the course.

Dummy variables for *female*, *othergender*, *white*

The colleges in the university are also in the dataset.

a. What are the average scores for males, females and those that identify with a gender other than male or female? Propose a regression, explain your specification, run the regression and interpret the coefficient(s) of interest. (You can use as many or as few variables as you deem appropriate.)

b. Do evaluation scores differ by race? Again, propose a regression, explain your specification, run the regression and interpret the coefficient(s) of interest.

c. Do you have enough evidence to conclude that evaluations differ by race and/or gender? Feel free to propose and report additional regressions to support your argument. Do you believe that your estimated coefficients are biased? Explain why or why not.

d. Are there any other variables you would have liked to include in your regressions? If not, explain why adding variables would not matter. If so, explain which variables you would like to have?

a.) Avg. Scores for Male = 4.258727
" " female = 4.350833
" " other gendel = 4.328433

The regression I proposed included female, othergender COURSEENROLLMENT (mean of all quarters) & response. All of these variables show statistical significance on A19 (Average evaluation score) and I believed affected the dependent variable.

b) After controlling for Race, there was statistical significance in evaluation scores differing, producing a coefficient of 0.0447, meaning white professors on avg. received a 0.0447 higher score on average.

→ Continued.

Continue to answer Question 4 on this page

however, their may be omitted variable bias in controlling for the typeof course (major) that can prevent this from being a conclusive report.

c) To make a more conclusive report, there also myl- be controls for the groupings of the course, which yielded a higher R-squared after including business, education, engineering, social sciences, hard sciences, & public policy. Additionally, after controlling for these groups, coefficients decreased with female (0.0836 ⇒ 0.0259), other gender no longer being statistically significant (0.0278 ⇒ -0.00528) & race still making a statistically significant effect on AI9 (0.0447 ⇒ 0.0417). Our estimated coefficients control for field they are in, but still may have differences varying on the term.

d) Adding the last variable, term, we can control for time-fixed effects, but still may have omitted variable bias based on professors teaching the course, this may increase our R-squared, but currently it is hard to say conclusively (despite being statistically significant) that the independent variables describe the variation in the dependent variable.