## Preliminary Schedule

|  |  |  |
| --- | --- | --- |
| **Date** | | **Tasks** |
| 11/3/2020 | 11/8/2020 | Find 2-3 articles on energy efficiency/ energy burden; add articles to the google doc and try to answer the project questions on canvas for each article (questions are copied below) |
| 11/8/2020 |  | Zoom meeting to discuss articles and select one<https://georgetown.zoom.us/j/2645661712> |
| 11/9/2020 | 11/15/2020 | Complete a first draft of question 1 |
| 11/15/2020 |  | Zoom meeting to go over the first draft; merge results into one document; and discuss next steps<https://georgetown.zoom.us/j/2645661712> |
| 11/16/2020 | 11/22/2020 | Complete a first draft of question 2 |
| 11/22/2020 |  | Zoom meeting to go over the second draft; merge results into one document; and make revisions<https://georgetown.zoom.us/j/2645661712> |
| 11/22/2020 | 11/29/2020 | Review question 1 and 2; make final changes/revisions |
| 11/29/2020 |  | Zoom meeting to discuss final revisions<https://georgetown.zoom.us/j/2645661712> |

## GitHub

Link to project’s github repository:

## Project Questions

### Part 1

Discuss existing research

1) Find a research paper on a topic of interest that uses OLS or probit/logit/LPM.

a. Provide a complete citation to this work. Please include a link (or send the paper) if

possible. Discuss the key finding of the paper

2) What is the analytical question? (If there are multiple analyses/questions addressed in the

paper, focus on a single important question).

a. Why is it important to answer this question?

b. Write out a statistical model and pay special attention to your subscripts, as these

indicate your unit-of-analysis (e.g., individuals at a given point in time, or national

data over time or perhaps individual data over time).

i. What is dependent variable?

ii. What is the key independent variable (or, in some cases, variables)?

iii. What are the null and alternative hypotheses?

c. Does the model use any specific functional form (e.g., probit, log)? If yes, explain

why.

3) Endogeneity

a. Which sources of endogeneity are the authors able to control for with additional

control variables?

b. Which sources of endogeneity are the authors unable to control for? Explain the

implications for your analysis (only work through one such (important) example in

detail). Use the conditions for omitted variable bias and our discussion of the

implications of omitted variable bias to discuss whether the particular omitted factor

will bias results upward or downward. If the paper is a RCT, explain a source of bias

that the RCT is designed to avoid.

c. Could measurement error cause endogeneity? Explain why or why not.

4) Post-treatment: are any of the included variables possibly “post-treatment” variables?

5) What factors affect the precision of the results?

a. Discuss, as appropriate autocorrelation, heteroscedasticity and multicollinearity and

what the authors do (or should do) about them.

b. Is the test statistically powerful? What are the implications of your answer to this

question for how you will interpret your results?

6) Have the authors considered heterogeneous treatment effects? (That is, have they considered

whether the effect of the key independent variable(s) differs across sub-groups?) Describe

how they do or could address this possibility.

7) Are the results generalizable?

### Part 2

Discuss a new research design: Develop a hypothesis to test on a question related to the

above article. This hypothesis must be different than the hypothesis tested in the paper discussed

In part 1. (For example, if the paper covers the effect of income on high school graduation in

Mexico, you cannot simply propose research on the effect of income on high school graduation

in the U.S.) If you are in doubt, be sure to see me.

1) What is the analytical question?

a. Why is it important to answer this question?

a. Write out a statistical model and pay special attention to your subscripts, as these

indicate your unit-of-analysis (e.g., individuals at a given point in time, or national

data over time or perhaps individual data over time).

i. What is dependent variable?

ii. What is the key independent variable (or, in some cases, variables)? [Please

focus on no more than two independent variables.]

iii. What are the null and alternative hypotheses?

b. Does the model use any specific functional form (e.g., probit, log)? If yes, explain

why.

2) Endogeneity

a. Which sources of endogeneity would you expect to be able to control for with control

variables?

b. Which sources of endogeneity would you expect to be unable to control for? Explain

the implications for your analysis. Use the conditions for omitted variable bias and

our discussion of the implications of omitted variable bias to discuss whether the

particular omitted factor will bias results upward or downward. If you propose a RCT,

explain a source of bias that the RCT will avoid. (That is, for a RCT paper, explain a

source of endogeneity that would occur if a RCT were not used.)

c. Could measurement error cause endogeneity? Explain why or why not.

3) Post-treatment: are any of the included variables possibly “post-treatment” variables?

4) What factors affect the precision of the results?

a. Discuss, as appropriate autocorrelation, heteroscedasticity and multicollinearity and

what you would do about them.

b. Is the test statistically powerful? What are the implications of your answer to this

question for how you will interpret your results?

5) Do you expect the treatment effects to be heterogeneous? (That is, will the effect of the key

independent variable differ across sub-groups?) Describe how to address this possibility.

6) Are the results generalizable?

## Articles

### 

### Gloria

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### 

### Merykokeb