THE UNIVERSITY OF CHICAGO
Harris Public Policy

PPHA 34600: Program Evaluation
Spring 2020
Problem Set 4
Due: Thursday, May 28, at 9PM to Canvas

Instructions:

This problem set consists of two files: (1) this document with instructions and questions; and (2) a dataset which you will use to answer the questions below.

You can work in groups of up to three. Please identify your group members. Groups can share code, but each group member must turn in their own problem set, and must have separate written answers to the questions. You should submit both written answers -- which should be parsimonious -- and a file which contains your code and results for the data analysis. You must use R. I recommend that you use RMarkdown or knitr, which will allow you to intersperse your code and written answers. Note that you are primarily being graded on your written answers. Problem sets must be submitted in PDF format. Problem sets must be turned in via Canvas; no late submissions will be considered.

Questions:
Local air pollution is among the greatest threats to human health today. In recent years, China has embarked on a "war on pollution," using a variety of approaches to try to reduce pollution exposure. A new blue-ribbon commission, the Pollution Regulation Organization for Greater Regional Air Monitoring, Evaluation, Valuation, And Life (PROGRAMEVAL), has come to the Harris School to find a team of experts to help them understand the effectiveness of China's pollution regulations on air quality, but all the faculty are busy -- so you've been asked to step in.

1. PROGRAMEVAL are interested in answering the following question: *What is the impact of provincial air quality regulations on local particulate matter (PM 2.5)?* In order to get started, they'd like you to present your ideal experiment. Explain what you would do to answer this question in a completely unconstrained world, and describe the dataset that you'd like to have to perform the analysis. Use math, words, and the potential outcomes framework to explain what you would estimate and how you would do so. Make sure to be clear about the unit of analysis (ie, what is "i" here?)

2. PROGRAMEVAL, being a commission and not a Chinese regulator, can't impose pollution regulations themselves. But they do have some data that they'd be willing to let you work with. They have a single temporal snapshot of air quality across many municipalities. They'd like you to look at average differences in air quality between municipalities with and without air quality regulations to get a sense of what these regulations do to air quality. You have a sense that this is not a great idea. Describe three concrete examples of why this comparison might not provide the answer that PROGRAMEVAL want.

3. Explain to the PROGRAMEVAL what the benefit would be of being able to observe municipalities at multiple points in time. Their database goes from 2001 to 2019, but they only want to share what's absolutely necessary, for confidentiality reasons. First, describe, in words

and math, what you would do with data on many municipalities, all of which imposed air quality regulations in 2004. Be sure to discuss the identifying assumption that would be required for this approach to recover the causal effect of air quality regulations on particulate matter. Provide three concrete examples of concerns with this approach.

4. Next, explain why it would be even better to have data on multiple municipalities, divided into two groups: municipalities that never imposed air quality regulations, and municipalities that imposed air quality regulations in 2004. Explain, in words and math, the estimator that you would use with this dataset. You should include an estimating equation in the form of a regression. Describe how this larger dataset would allow you to resolve the concerns you had above. Be sure to discuss the identifying assumption that would be required for this approach to recover the causal effect of air quality regulations on particulate matter. Provide two examples of remaining concerns, even in this larger dataset.

5. PROGRAMEVAL, given your even-handed discussion of various approaches, is willing to put their faith in you. They will give you data on the universe of their consumers from 2003 to 2007. This includes municipalities that imposed air quality regulations across several different years. Describe, in words and math, how you would estimate the effect of air quality regulations on particulate matter using this dataset. You should include an estimating equation in the form of a regression.

6. Use the included *ps4_data.csv* dataset to implement a simple comparison of average particulate matter between municipalities with and without air quality regulations. Describe what you find. Use regression to perform a time-series analysis of the effect of air quality regulations on particulate matter, using only municipalities who introduced regulations in 2004. Describe what you find. How does this differ from what you estimated using the initial estimator. Plot particulate matter against time for municipalities that imposed air quality restrictions in 2004. What do you see? (It may also be helpful to plot *average* consumption across municipalities). Does this figure affect how you interpret your estimates?

7. Plot (average) particulate matter against time for municipalities who never imposed air quality regulations. Assess the viability of using these municipalities as a control group for the 2004 regulators. Plot (average) particulate matter against time for municipalities who passed air quality regulation in 2006. Assess the viability of using the non-regulating municipalities as a control group for the 2006 regulators.

8. Using just the non-regulators and the 2006 regulators, estimate the causal impact of imposing air quality regulation on particulate matter. To do this, begin with a simple difference in means (rather than regression). Next, use a simple regression (no fixed effects). Finally, use fixed effects to control for common time shocks and time-invariant municipality characteristics (you can do this either via dummy variables or de-meaning). Report what you find. Be sure to adjust your standard errors appropriately in the regression-based estimates. Describe how this compares to what you estimated in (6) and (7).

9. Plot average particulate matter over time, separately (but on the same graph) for municipalities that imposed air quality regulations in each of the years from 2003 to 2007. Drop the municipalities that regulated air quality in the year that looks different from the rest of the years, and explain why you can't estimate a credible causal effect for these municipalities. Using the remaining municipalities to estimate a panel fixed effects regression to identify the causal effect of air quality regulation on particulate matter. Describe what you find. How does this compare to what you estimated in (9)? Use an event study regression to estimate how this treatment effect varies over time. Note that you will have to omit one of the event study treatment dummies (otherwise everything will be collinear). Standard practice is to leave out the T-1 dummy. Plot the resulting event study point estimates and 95 percent confidence intervals. Describe how the treatment effect varies over time, if at all.

10. PROGRAMEVAL would like a summary of your findings. Explain which of the results you've come up with is your preferred estimate, and why. Make sure to describe at least one remaining potential shortcoming with these results. Finally, interpret the magnitude of your estimated effects: do your results suggest that the PROGRAMEVAL should be strongly promoting air quality regulations?