

PPHA 34600: Program Evaluation
Spring 2020
Problem Set 3

Due: Thursday, May 21, 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:

The state of California is increasingly worried about wildfire season. In particular, sparking electricity transmission lines have been identified as a cause of large fires in 2018 and 2019. In response to this, the state has required that its electric utilities shut off power during times of high fire risk, as part of a program of public safety power shutoffs, which they are calling. However, as part of a cost-benefit analysis to inform future policy, the state needs an estimate of the impact of blackouts on household adoption of solar PV (since, with the right inverter, a solar panel can continue to power a house even when grid power is off). The state have hired an analysis team, called the CALifornia Blackouts and Electricity Analysis Research Service (CALBEARS). However, these analysts are in a bit over their heads, so they've called you in to help (go Maroons!).

1. CALBEARS are interested in answering the following question: *What is the effect of hours of electricity outages experienced by a household on the kW of solar PV they install?* To make sure everybody is on the same page, explain to them what the ideal experiment would be for answering this question. Describe the dataset that you'd like to have to carry out this ideal experiment, and 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. CALBEARS are on board with your explanation, but, as they've discussed with you, they won't be able to implement your preferred solution. They don't think that a selection-on-observables approach will work (they're very sophisticated). They're also limited by state privacy laws: they will only be able to give you one wave of data (no repeated observations). Given these limitations, describe the type of research design you would try to use to answer their question of



interest. Be explicit about the assumptions required for this design to work, describing them in both math and words.

3. CALBEARS are interested in this research design. It sounds promising. They'd like you to propose a specific approach. Please describe a plausible instrumental variable you could use to evaluate the effect of power outage hours on kW of solar PV installed. Why is your proposed instrument a good one? Do you have any concerns about your ability to estimate the treatment effect using your instrument? If yes, why? If no, why not?
4. CALBEARS is intrigued by your approach. After an internal discussion, they've come back to you with great news! It turns out that one of the California utilities ran a small pilot program where they randomly cut power for different lengths of time to different households as part of an equipment test. With this new information, please describe to CALBEARS how you would estimate the impacts of power outages on solar PV adoption? Use both words and math.
5. CALBEARS agree that your approach is a good one. So good, in fact, that they'd like to see it in action! They are willing to share some data with you, in the form of *ps3_data.csv*. Please report the results of an analysis of the impacts of power outage hours on kW of PV adoption, using *utility_outage_hours* as your treatment variable and *installed_pv_contractors* as your outcome variable.
6. CALBEARS like your analysis, but they're a bit worried about the quality of their data on PV adoption. The way they normally collect these data is by collecting details on solar installations from contractors. However, they went and did some back-checks in a subsample of data that they gave you, and noticed that the contractor reports seem to be off. They would like you to make a graph showing the relationship between their back-checks (*installed_pv_backchecks*) and the contractors' estimates (*installed_pv_contractors*). Describe to them what you find. Is this likely to be a problem for your analysis? Why or why not? Next, estimate the impacts of power outages on PV adoption using the backcheck data and using the contractor estimates. Report what you find. Do your estimates differ? If no, explain why not. If yes, explain why.
7. The challenge with back-checks is that they're very expensive to do. Fortunately, CALBEARS realized that there's another field on the contractors' reports that seems to match the back-checks much better. They'd like you to make a graph showing the relationship between their back-checks and this new measurement (*installed_pv_contractors_v2*). Describe to them what you find. Is this likely to be a problem for your analysis? Why or why not? Next, estimate the impacts of power outages on PV adoption using the backcheck data and using the (new) contractor estimates. Report what you find. Do your estimates differ? If no, explain why not. If yes, explain why.
8. CALBEARS comes back to you again with yet another data problem. This time, they're worried that the utilities aren't measuring the hours of power outages very well. CALBEARS explain to you that, in one utility (labeled *iou == 1* in the data, because #privacy), something was going wrong with the power grid meters they were using. Houses that only had an hour or two of power outages appear to be reported accurately, but the higher the hours of outage, the more inflated utility 1's measurements are. In the other utility (labeled *iou == 2*), there are still imperfect



measurements, but CALBEARS is convinced that the measurement problems are random. Explain the implications of these data issues in each utility to CALBEARS. Are these measurement issues going to be a problem for your analysis? Use words and math to explain why or why not. Despite any misgivings you might have, run your analysis anyway, separately for each utility this time (using your preferred PV variable from the three described above), and report your findings.

9. CALBEARS conducted a survey of households to understand their power outage experiences, and asked households to report the number of hours of outages they experienced (*survey_outage_hours*). Describe how you could use these data to correct any issues you reported in (8). What conditions need to be satisfied in order for this to work? Carry out your proposed analysis. Report your results. What do you find? How do your results compare to your estimates in (8)? Which estimates would you send to CALBEARS as your final results?