POLS0012 Causal Analysis: Tutorial 7 Exercises

Leah Stokes (2016)[1] examines whether governments are punished electorally for building wind farms, a policy that mitigates climate change but may impose costs on the communities where turbines are sited. She looks at Ontario in Canada, where from 2009 the provincial government removed local communities' rights to make planning decisions on the building of wind turbines. Instead, decision-making was centralised and turbines were imposed by the government. It chose to build turbines in places where their construction was most feasible and they would generate the most electricity. In particular, they were more likely to be sited in places with higher prevailing wind speeds. Whilst certain broad areas are in general better suited for turbines (more rural and more elevated places, and areas closer to the windy great lakes), she argues that within these broad areas wind speed varies at random at the local level. This means that local communities could not select out of (or into) receiving a wind farm based on their levels of support for the policy or for the government. This is therefore a natural experiment where wind speed is an instrument that randomly encouraged the government to site turbines in particular places.

Her outcome of interest is change in support for the incumbent government from 2007 (before the wind farm policy) to 2011 (after it began) at a highly localised level known as "precincts" in Canada, which typically contain around 300 voters. Using GIS software, she geo-located all wind turbines that were built or proposed in the period and matched them to precincts, where she collected voting data, localised prevailing wind speeds, and background covariates. The dataset for this question is contained in the file Stokes.Rda and contains the following variables:

- *chng_lib* - outcome: pp change in support for the incumbent government, 2007-11
- *prop_3km* - treatment: =1 if a wind turbine was built or proposed within 3km, 0 otherwise
- *avg_pwr_log* - instrument: prevailing wind speed in the precinct, logged
- *longitude* - of the precinct
- *longitude* - of the precinct
- *ed_id* - the broader district within which the precinct is located
- *mindistlake* - distance to the great lakes in km
- *mindistlake_sq* - distance to the great lakes in km, squared

a) Assess whether wind speed can be considered to be as-if randomly assigned geographically, by regressing *avg_pwr_log* on all of the geographical covariates. What do you conclude?
Code Hint: Remember to use `factor()` for the *ed_id* variable

[1] Leah Stokes (2016). "Electoral Backlash against Climate Policy: A Natural Experiment on Retrospective Voting and Local Resistance to Public Policy." *American Journal of Political Science* 60 (4): 958-974

b) Estimate the first-stage relationship between *prop_3km* and *avg_pwr_log* using a regression with no added covariates. Interpret the result precisely.

c) Stokes actually estimates the first and second stages with a full set of geographic controls included. Why do you think she does this?

d) Estimate the first-stage relationship between *prop_3km* and *avg_pwr_log* using a regression, this time with a full set of geographic controls. Interpret the result, and explain why it does or does not differ from part (b)

e) Conduct an F test for the strength of the *avg_pwr_log* instrument. Using this test and your answer to (d), do you think that the instrument can be considered to be relevant?
Code Hints: Use the function `waldtest()` in the `lmtest` library. Your code should take the form `waldtest(model1,model2)`, where `model1` and `model2` are the names of estimated regression models with and without the instrument

f) Estimate the Local Average Treatment Effect of *prop_3km* on *chng_lib* using two-stage least squares with *avg_pwr_log* as the instrument and the full set of geographic controls. Interpret the coefficient on *prop_3km* and its statistical significance precisely.
Code Hints: Remember to use `ivreg()` in the `AER` library. Your code should take the form:

```
ivreg(outcome ~ treatment + covariates | instrument + covariates)
```

g) Now find the same Local Average Treatment Effect of *prop_3km* on *chng_lib*, but this time using two-stage least squares manually in two separate stages with covariates. Does your result differ from part (f)?
Code Hints: Extract fitted values from the first stage using `fitted.values()` and add them to the dataset as a variable. In the second stage, include them as an explanatory variable

h) Outline one way in which the randomisation assumption could be violated, here. How serious do you think this violation is likely to be for the internal validity of the results?

i) What does the 'exclusion restriction' mean in this study? Do you think it is likely to be violated?

j) Briefly, assess the external validity of this study. To what extent do you think the results can be generalised to other settings, times, treatments, etc.?