The University of Texas at Austin
ECO 348K (Advanced Econometrics)
Prof. Jason Abrevaya
Fall 2016

PROBLEM SET #4 (due Thursday, October 20th, 8am)

This problem set is based upon the Stata dataset **lee-moretti-butler.dta** available on the Canvas site. These data represent a subset of the data used in the 2004 paper "Do Voters Affect or Elect Policies? Evidence from the U.S. House" by David Lee, Enrico Moretti, and Matthew Butler in the *Quarterly Journal of Economics*. The data provide the results from U.S. Congressional (House of Representatives) elections between 1948 and 1990. Only elections with incumbents running are included within the data; uncontested elections (where vote share is 100%) have been dropped. The full sample size is 9,788.

The variables of interest are as follows:

- **democrat**: 1 if the Democratic candidate wins the *current* election, 0 otherwise (Republican wins)
- **demvoteshare**: fraction of votes for the Democratic candidate in the *current* election
  - Note that **democrat**=1 if and only if **demvoteshare**>0.5
- **lagdemocrat**: 1 if the Democratic candidate won the *previous* election, 0 otherwise (Republican won)
- **lagdemvoteshare**: fraction of votes for the Democratic candidate in the *previous* election
  - Note that **lagdemocrat**=1 if and only if **lagdemvoteshare**>0.5
- **year**: year of the current election
- **pcturban**: fraction of urban population in Congressional district
- **pctblack**: fraction of black population in Congressional district
- **pcthighschl**: fraction of HS graduates in Congressional district

1. Do a scatter plot of **demvoteshare** versus **lagdemvoteshare**. Do you see a discontinuity at **lagdemvoteshare** $= 0.5$? Imagining a line going through the left part of the scatter (for **lagdemvoteshare** < 0.5) and another line going through the right part of the scatter (for **lagdemvoteshare** > 0.5), what is your "eyeball" estimate of the magnitude of the jump?

2. What is the raw difference in **demvoteshare** between elections with a Democratic incumbent (**lagdemocrat** = 1) and those with a Republican incumbent (**lagdemocrat** = 0)?

3. Repeat Question 2, but now looking at smaller "window" widths. Specifically, you should look at the raw difference in **demvoteshare** but only on the subsample for which $|lagdemvoteshare - 0.5| < h$, where h is the window width. Do this for $h = 0.20$, $h = 0.10$, $h = 0.05$, and $h = 0.01$. How do the sample sizes change? How do the estimated differences change?

4. How would you estimate the differences in Question 3 using a regression? Do it for both $h = 0.05$ and $h = 0.01$, using robust standard errors. What are 95% confidence intervals for the incumbency effect from both of these regressions? (Hint: Your dependent variable should be **demvoteshare**.)

5. For this question, return to using the full sample. Run the following Sharp RD regressions for the outcome **demvoteshare**, all with a jump discontinuity at **lagdemvoteshare** = 0.5, and report the estimated incumbency effect from each of the models: (Recall that interpretation of your regression results will be much easier if you center the running variable at its mean.)

   (a) Linear on both sides of the RD cutoff, slopes identical on left and on right

   (b) Linear on both sides of the RD cutoff, slopes differing

   (c) Quadratic on both sides of the RD cutoff, same function on left and on right

   (d) Quadratic on both sides of the RD cutoff, different function on left and on right

   (e) Same as (d), but also include **year**, **pcturban**, **pctblack**, and **pcthighschl** as explanatory variables

6. For the regression in Question 5(d), generate the fitted values for **demvoteshare**. To get a visual representation of your estimated model, do a scatter plot of the fitted values versus the running variable.

7. Repeat Question 5(d) and Question 6 for the binary outcome **democrat** (that is, a linear probability model for **democrat**). What is the incumbency effect here (the effect on the probability of a Democratic victory)? Do the fitted values on your scatter plot remain between zero and one?

8. Using the simpler specification in Question 5(a), this question will assess whether 0.50 is the "right" cutoff value. Re-run the regression in 5(a), but using a different value for the cutoff R^*. Specifically, try doing $R^* = 0.44$, $R^* = 0.47$, $R^* = 0.50$ (original model), $R^* = 0.53$, and $R^* = 0.56$. How do the R-squared values for each of these regressions compare? What do you conclude?