EXERCISE SHEET 10: FUJIWARA (2015): VOTING TECHNOLOGY

Main learning goals: 1. Using difference and lag operators

2. Some basics on bootstrapping
3. Supporting graphs and robustness checks

Hand-in: As usual: do-file and the corresponding log-file (saved in .log-format) containing the following steps (numbered in the do-file).

1. Move from the municipal level dataset that we have seen last session to the state level dataset `state.dta` and make yourself familiar with the dataset, i.e. the data structure and the contained variables.
2. Estimate equation (6) and (7) for the outcome variable *Valid Votes/Turnout*. Compare your results to table 4. Do you get the same coefficients and standard errors? What are the statistical challenges for the estimation at the state level, in particular for the p-values?
3. Implement a bootstrapping procedure, using `cgmwildboot`, to replicate the bootstrapped p-values for the previous regressions, reported in table 4. Use 1000 replications and make sure that you set the seed. Briefly describe the basic idea that we implement here.
4. Generate residual scatter plots and linear fits in which you plot the change of the valid votes against the share of voters above the threshold. Use the residuals for which you controlled for regional dummies. Plot the graphs both, for 1994-1998 and for 1998-2002. Briefly discuss what additional support these graphs give to the results of table 4.
5. Now go on with equation (8). Think about the identification behind this equation and implement it. Why is the estimate equal to $(\theta^{98} - \theta^{02})/2$, with some approximation error? Estimate this regression. Compare your results to the values in table 4. What do you think about this identification strategy?
6. Generate a new variable that will help you to identify $(\theta^{98} + \theta^{02})/2$ in an estimation. To do so you have to implement the strategy described in the text as well as footnote 45. Estimate the corresponding regression to estimate the coefficient for *Valid Votes/Turnout* in column and (4).
7. Do a robustness check in which you investigate the effect of outliers. Write a loop in which you estimate equation (8), but leave out one of the 27 states at a time. Store the estimates and set up a scatter plot with the 27 values and their confidence bounds for the 95% confidence interval. For the sake of computational time, use the standard confidence intervals and not the bootstrapped ones.
8. As a last check we move away from the electoral outcomes to the fiscal outcomes, in particular the share of spending in health care. Open the file `yearly.dta` to access annual outcomes. Then, regress the share of spending in health care on the share above cut-off interacted with the corresponding years. Add regional dummies and constrain your regression to the period 1995-2006. Save your coefficients and plot the coefficient estimates.

Discussion Questions:
- What do you think about the analysis/article/story? Are you convinced?
- Why did the paper end up in Econometrica?

Further comments: Submit 3 files (.do, .log, .pdf(with graphs)) to Ilias. Usual procedures apply.