Problem Set 2

Statistics for International Relations and Political Science 2
IHEID
2021-2022

April 14, 2022

Please complete the following problem set and submit the answers on Moodle. For R-related questions, please submit the R script in addition to the results.

1. Please open the replication dataset (cgss06.dta) and import it to R. The dataset contains survey responses of Chinese citizens in 2006 between September 05 and November 29, as indicated in Yang and Jiang (2016).
   (a) First, we would like to estimate the effect of happiness (*happiness*) on the belief that there is no need for democracy in the country (*nodemo*). Write the equation and estimate the relationship.
   (b) Now we would like to treat the data as pooled cross sections. Each time period will be the week that the survey was conducted. Add the week fixed effects and rewrite the equation. Run the model and see if the main result changed. What could be the potential advantage of adding the week-fixed effects and when should we include them?
   (c) We also would like to see if there is any week in which the main effect was substantially larger or smaller than the first week. Rewrite your equation and estimate the effect. What would be your conclusion and why?
   (d) Now we want to see if there was a meaningful change in the dependent variable *nodemo* after the political purge had been announced. Restrict the data to those districts (*local_num*) where some, but not all, respondents were surveyed after the purge, as indicated in the paper. The purge took place in Shanghai, so we first would like to compare the average *nodemo* of those in Shanghai and those in other provinces differed after announcement. Do you find a meaningful difference?
   (e) What could be a potential problem of making a conclusion based on the previous exercise?
   (f) Now write the standard DiD equation without including fixed effects and control variables. Run the model and estimate DiD. How would you interpret the coefficient?

(g) The author included both the region fixed effects and time fixed effects in the model. Do you think that including these fixed effects help our estimation? If so, how? If not, why not?

(h) What about individual-level control variables? What do we gain by including control variables? Or do you think that it does not help?

(i) Include region fixed effects using *province* and date fixed effects using *date*, in addition to age(*age*), education (*college*), employment status (*working*), and being a party member (*party*). Rewrite the DiD equation and estimate the model. Did the main results change?

(j) Can you confirm that the effect of purge was particularly stronger in Shanghai than in other provinces?

(k) Try four different cut-off points for your DiD estimation, which are one week before the purge and one, two, and three weeks after the purge. Compare the DiD estimates. What do you find and what can you conclude?

2. Read Bechtel and Hainmueller 2011 carefully and critically evaluate the DiD estimation. In doing so, please include your answers to the following questions.

   (a) What are the assumptions underlying the DiD estimation in this paper?

   (b) How do the authors address (or fail to address) each point?

   (c) Is there any issue that has not been successfully addressed by the authors?