ECON 7010: Applied Microeconomics

Fall 2023

Problem Set 2

Laurel Krovetz, Shubhalakshmi Nag, Annie Kaur

1 Part 2

a) The data used in this table came from multiple sources. House of Representatives voting data came from David Leip's Atlas of Elections and included data from the 2006, 2008, and 2010 House elections. This data was provided to us on Canvas and only required cleaning for use. The demographic data came from multiple sources -

1. American Community Survey, 2009 - Population, median income, percentage of white, African-American, and Hispanic population, and unemployment rate.

2. Census, 2000 - Rural share of population

The data was accessed using an API from the US census website. The data on precipitation was also accessed using an API from the NOAA website. The API was quite faulty and as a result there are some counties for which we couldn't get the precipitation data. Lasty, data on rally participation was accessed from two main sources - 1) the link provided on the problem set, and 2) the New York Times data available online. We computed the average, minimum, and maximum across the two source, but stuck with using the average.

Below is our replication of Table I. Comparing our values with the values from the paper allows us to assess how closely our data appears to match the data used by the authors. The number in our replicated table are not too different from the numbers in the table from the paper. This would suggest that our data is a good match of the data used by the authors but is not an exact match. Therefore, our estimates in the following replications are likely to be slighly off.

	(1) Rain	(2) No Rain	(3) Difference
Precipitation	0.374 (0.010)	0.008 (0.000)	0.366 (0.006)
Election 2008			
Republican vote (% of votes)	50.335 (0.793)	51.361 (0.485)	-1.026 (0.985)
Republican votes (% of pop)	22.275 (0.397)	21.894 (0.228)	0.381 (0.471)
Democratic votes (% of pop)	20.325 (0.328)	19.536 (0.226)	0.790 (0.454)
Total votes (% of pop)	43.386 (0.316)	42.359 (0.225)	1.027 (0.448)
Election 2006			
Republican vote (% of votes)	51.783 (0.693)	51.395 (0.428)	0.388 (0.872)
Republican votes (% of pop)	16.275 (0.306)	15.978 (0.159)	0.297 (0.337)
Democratic votes (% of pop)	13.703 (0.235)	14.821 (0.175)	-1.118 (0.348)
Total votes (% of pop)	30.598 (0.341)	31.458 (0.206)	-0.860 (0.422)
Demographic controls 2009			
Median household income	25,997.9 (230.47)	24,714.14 (93.953)	1,283.754 (213.748)
Population	134,191.6 (10,305.68)	88,886.36 (6,813.472)	45,305.29 (13,762.92)
Rural pop (%)	55.811 (1.218)	60.535 (0.633)	-4.724 (1.341)
White pop (%)	85.023 (0.627)	86.002 (0.331)	-0.979 (0.699)
African American pop (%)	11.26 (0.567)	8.514 (0.293)	2.746 (0 .623)
Immigrant pop (%)	2.684 (0.148)	2.737 (.077)	-0.053 (0.163)
Hispanic pop (%)	4.981 (0.258)	8.466 (0.295)	-3.489 (0.562)

b) In this paper, the challenge in estimating the impact of protest is the existence of unobserved political preferences that are likely to determine both the number of protesters (independent variable) and policy outcomes (dependent variable). Therefore, it is improbable that a simple regression of policy on protest size would represent a causal influence. To address the endogeneity problem, authors use the rainfall during the day of the protest as the instrumental variable. By doing so, the underlying assumption is that rainfall on the rally day affects the policy and voting behavior only through the number of protesters. It is believed that people are less likely to attend a rally if it rains. Hence, the first step of estimation is to ensure the instrumental variable rainfall satisfies the relevance assumption; that is, the negative relationship between the number of protesters and rainfall actually exists.

- To estimate the effect of the protests, authors first investigate whether rainfall decreases attendance by regressing the number of protesters in a county on a dummy variable that indicates whether there was significant rain in the county.

- On the left- hand side, the variable " Protesters " measures rally attendance in the county. On the right-hand side, the variable "Rainy Rally" is equal to one if there was more than 0.1 inch of rain in the county on the day of the rally. Moreover, a set of control variables are added to the equation to ensure a more precise estimation result. If the researcher's assumption about the rainfall and the number of protesters is valid, the estimation coefficient in the equation should be negative indicating that relevance assumption is satisfied.

- A key identifying assumption is that rainfall is uncorrelated with other determinants of political outcomes. Evidence shows that rainy and non-rainy counties and districts are quite similar on average and the rainfall dummy is not significantly correlated with any of the pre-rally political outcomes grants credibility to identification strategy.

- The results of the estimation of effect of rainfall on rally attendance in section V show that the instrumental variable satisfies the relevance assumption. Three sets of regression are run depending on three dependent variables to show the existence of negative and statistically significant

relationship between "Rainy Protest" and "Protesters". In the first regression, the dependent variable is scaled by the percent of the population attending. Column 1 of Table 3 implies that rainfall decreased the share of the county population protesting by 0.082 percentage points. In order to avoid measurement error and under-reporting, the highest reported attendance among the three sources was used, and the result is also negative and statistically significant. The dummy independent variable is replaced by the precipitation amount and results are still negative and statistically significant. Negative and statistically significant coefficients consistent with the hypothesis obtained when dependent variable of protesters is measured in thousands of protesters, and dependent variable is scaled by log. Therefore, the instrumental variable chosen by authors satisfies the relevance assumption.

(c) i) Exclusion restriction assumption: We can estimate the impact of protest size using an instrumental variables approach assuming that the absence of rainfall affects policy and voting behavior only through the number of protesters and rally attendance.

- The main challenge in measuring the effectiveness of political protests is that unobserved political beliefs or a culture of activism are likely to be correlated with both the number of protesters and other political behavior such as voting.
- The authors provide their reasoning to this problem in section "IV.B. Exogeneity Check" of the paper. First, as shown Table I (also A.1 and A.2) of the paper, rainy and non-rainy counties are not significantly different on average for the key attributes.
- Second, Table II of the paper presents the regression results of 2008 House and presidential election outcomes on rainy weather on Tax Day 2009. The results suggest no significant correlation between rainy counties and previous election outcomes.
- Moreover, in the appendix section, the authors also present results using the pre-rally donations to Our Country Deserves Better PAC and the ACU's roll-call scores years before the actual rally took place as regressors.
- They conclude that no statistically significant correlation is found between rainy counties and past political outcomes, and hence this provides sufficient evidence that the exclusion restriction is satisfied.

ii) If the instrumental variable is correlated with the dependent variable (through channels other than rally participation) in the model, it could act as a potential source of bias. The weather in various locations shouldn't be correlated with the vote shares for various parties in order to meet the exclusion requirement in this model. We can test this by performing t-tests on voting shares for Republican and Democratic parties on a rainy and non rainy day and if there is no effect of regional weather conditions on the voting shares of the parties, then the claim that there is no serious endogeneity problem is strengthened. In order to have a rough test of the exclusion restriction, we 1) regress the rain instrument on the given controls and attain the residuals, 2) regress the dependent variable (both republic and democratic vote share) on the set of controls and obtain the residuals, and finally conduct a t-test using the residuals from 1) and 2). The t-test results for both the republic and Democratic vote shares were insignificant, suggesting that the exclusion restriction holds. The analysis is provided in the do-file 'exclusion_restriction.do'.

iii) Although rainfall appears to be a valid IV, the concerns regarding the violation of exclusion restriction are still valid. A potential violation is that the precipitation or frequency of rainfall in a

specific geographical region may affect residents' political tendencies through different channels, except for the number of protesters on Tax Day.

- This exclusion restriction seems plausible, but another valid concern is that bad weather may also make a rally less pleasant for actual attendees and the consequent movement less. We would then be measuring the effect of a combination of rally size and rally impact per attendee as determined by the likelihood of new social ties forming.

- If we exclude weather anomalies, the existence of rain patterns is quite normal across the regions, and are contingent upon local geographical features. A vast body of research has established the existence of relationship between weather and people's behavior, mood, and socioeconomic traits. Similar geographic, demographic, and socioeconomic traits exist in some regions of the nation, and these factors influence people's political preferences and actions. For instance, it has been observed that people in the mid-west exhibit solid demographic and socioeconomic characteristics and lean towards Republicans-leaning states are less rainy and Democratic states are on rainy coastlines (but it's important to avoid making direct causation claims).

- A related worry is whether the likelihood that the protests will be covered by the media is influenced directly, as opposed to indirectly, by the number of participants. If there is such a direct effect and media coverage of political protests affects voting behavior and policy-making, the exclusion restriction would again be violated.

- Restricting our research to the swing states which are dispersed throughout the US and have distinct characteristics could be one approach to get around this association. In that instance, the correlation between protests, rainfall, and political outcome will be more accurately determined because a protest in a Texas county won't be able to swing the vote from republican to democratic.

(d) Below are our results for the replication of Table III, split into two tables. The first table is the replication when the dependent variables is protesters as a percentage of the population. The second table is when the dependent variable is number of protesters (in thousands).

Note that we only collected rally attendance data from two of the three sources the authors used in their analysis, and so our analyses are not entirely comparable. Finding the correct data was a challenge associated with the problem set overall. Additionally, note that while we used most of the controls used by the authors in these regressions, we do not include some of them, such as population density, and change in unemployment between 2005 and 2009. We include most of the electoral and demographic controls though.

Protesters, % of pop	**(1)**	**(2)**	**(3)**	**(4)**
Rainy protest	0.0013 (0.0089)	-0.0024 (0.0146)	0.0053 (0.0184)	-0.0702 (0.0426)
Observations	2,906	2,906	2,906	469
Protesters variable	Mean	Max	Mean	Mean
Rain variable	Dummy	Dummy	Continuous	Dummy
Sample counties	All	All	All	Protesters > 0
Election controls	Y	Y	Y	Y
Demo controls	Y	Y	Y	Y

Protesters, ,000s	**(1)**	**(2)**	**(3)**	**(4)**
Rainy protest	-0.0599 (0.0189)	-0.1216 (0.0320)	-0.1045 (0.0396)	-0.227 (0.1019)
Observations	2,906	2,906	2,906	469
Protesters variable	Mean	Max	Mean	Mean
Rain variable	Dummy	Dummy	Continuous	Dummy
Sample counties	All	All	All	Protesters > 0
Election controls	Y	Y	Y	Y
Demo controls	Y	Y	Y	Y

(e) To the best of your ability, replicate MSVY Table VI.

Table 3: 2SLS estimates

	(1)	(2)	(3)	(4)	(5)	(6)
	Republic vote (%)	Republic vote (%) - 2SLS	Dem. vote (%)	Republic vote (%) - 2SLS	Republic vote share	Republic vote share - 2SLS
Rainy Protest	-0.693**		0.160		-0.0190*	
	(0.256)		(0.409)		(0.00771)	
% Protesting		83.30		-19.20		2.215
		(86.40)		(59.01)		(2.517)
Observations	2905	2905	2905	2905	2885	2885
R - squared	0.7303	-	0.7534	-	0.7585	-
Protesters Variable -	Mean	-	Mean	-	Mean	
Election Controls Y	Y	Y	Y	Y	Y	
Demographic Controls Y	Y	Y	Y	Y	Y	
Dep. var mean 19.39	19.39	11.54	11.54	0.61	0.61	

Standard errors in parentheses

* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

(f) Challenges, thoughts, concerns regarding the paper

- Although the data collection part was challenging, the learning process was quite beneficial, especially for the API part, which was entirely new to all of us.
- The procedure of collecting and cleaning protester data was also difficult. In the data set provided online, several cities were not mapped to their corresponding state and county. Dealing with missing values, inconsistencies, and merging data from different sources posed challenges.
- The replication process required a lot of trial and error, which was not easy, because we had to obtain access to every data set used and because portions of the paper were not very precise about the sources of data utilized. Although the authors provided some guidance in the form of a data appendix, this was not very helpful in collecting some of the data.
- There could be spillover effects that were not captured within the cross-sectional framework. The estimates using rainfall as an instrument capture the Local Average Treatment Effects (LATE), which may differ from the average effect of protests on political outcomes, or the effect under identical weather conditions. As the exclusion restriction concerns still remain, we believe that the paper's results cannot be fully relied upon. It is challenging to generalize the paper's results to other settings and the external validity is questionable. Nevertheless, the use of rain as an IV was highly innovative and the results survive variety of robustness checks, so, the the finding that number of protesters expressing their beliefs in favor of a policy change is a sufficient statistic to describe the distribution of beliefs when distribution of policy preferences in society is unobservable and protesting is costly seemed to pique people's interest as seen by the numerous citations it received.

- All things considered, it was a worthwhile learning experience that provided insights into the difficulties of data collecting, cleaning, and analysis as well as a critical evaluation of the findings' robustness.