

UNIVERSITY OF BAYREUTH

Causal Inference Spring 22

Take-Home Exam

Due on 22 August at Noon

Word limit: 4000 words

Exercise 1: Rewards or Competence?

A voluminous literature uses natural disasters or other events to examine the extent to which incumbents are rewarded for their performance in delivering services to their constituents (Wolfinger and Rosenstone 1980; Mettler and Stonecash 2008; Bechtel and Hainmueller 2011). What remains unclear in these cases is whether voters reward incumbents' post-disaster efforts because they are grateful (similar to how voters might act when directly bribed by politicians) or because the politician's disaster response reveals to them that the politician has higher competence than they thought. To address this question, one could look at an instance of a natural disaster in which the incumbent did not perform well, but did financially reimburse the affected areas.

One such example is the *Prestige* accident that took place in late 2002 in the Galician coast, in Spain. The accident led to the severe pollution of the Spanish north coast. Although the government failed to manage the crisis, it generously reimbursed the residents of the affected areas. Here comes a quick chronicle of the events:

On 13 November 2002, the Bahamas-registered oil tanker *Prestige*, laden with 77,000 tonnes of heavy fuel oil, broke in two off the Galician coast (Spain) spilling an unknown but substantial quantity of its cargo. Although the accident occurred within 5 miles of the Galiza, the Spanish national government (in the hands of the Conservative Popular Party [PP]) refused it safe harbour, instead sending it away from shore in a north-western direction. This was the first in a series of decisions that later provoked on November 19 the sinking of the ship 150 miles off the coast and the biggest oil spill, in terms of affected area, since the Exxon-Valdez accident (note that the entire Spanish northern Atlantic coast

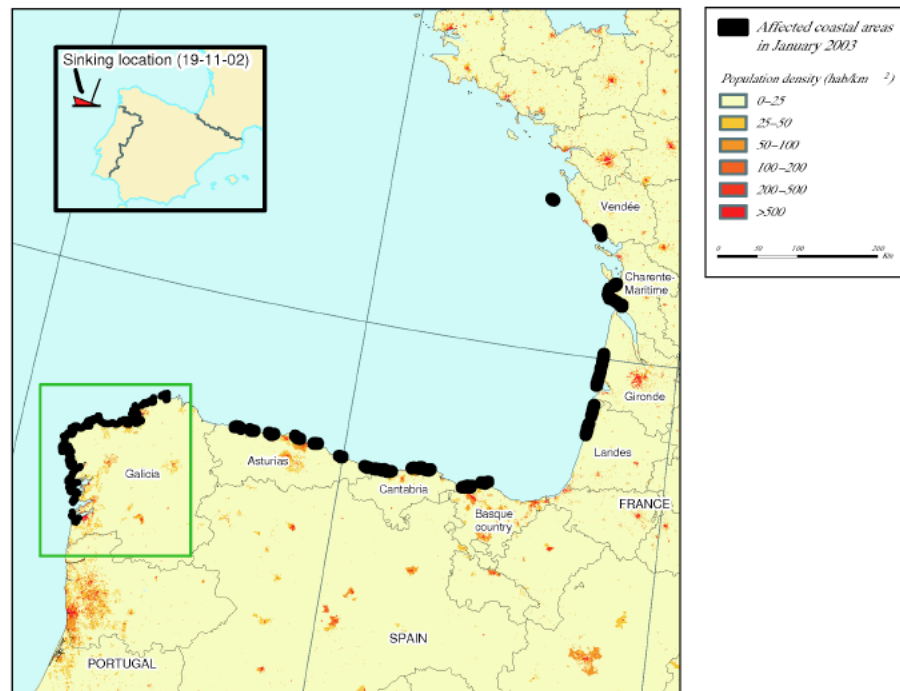
was affected as well as some Portuguese and French coastline). Pristine ecosystems were harmed, and all fishing and seafood collection - which are important industries in Galicia - were forbidden for months along the 1,390 km Galician coast. Figure 1 presents the geographical area that was affected by the accident.

The Prestige oil spill occurred only about six months prior to the 2003 regional and local elections and was without any doubt the most devastating natural disaster in Spain in over 25 years of democracy. Some opposition political parties (e.g., the Galician Nationalist Party [BNG], United Left [IU], and the Socialist Party [PSOE]) presented the decision of taking the ship far from the Spanish coast as a political one, revealing the “incompetence” of the ruling party in Madrid, and as evidence of the small weight attached by the central government to Galician issues. More broadly, it provided future Prime Minister Rodriguez Zapatero and the main opposition party—the Socialist Party (PSOE)—with a key opportunity to win over voters and improve their performance in the upcoming election. In fact, one Socialist MP in the regional assembly of Madrid had to resign when one radio station made public that the deputy believed that the PSOE was “overrun” of voters after the disaster and he subsequently joked about the convenience of “sinking another boat, if necessary.”¹

Yet, this is only part of the story. In response to the spill, the Galician *regional* government - also in hands of the PP - launched a huge relief program that included the abundant allocation of rapid and massive transfers of aid to citizens in affected areas (the ‘Plan Galicia’). The first payments went out two months later. These funds were seen as particularly suitable given the fact that the major economic benefit of Galicia is its fishing industry. To sum up, two features characterise the Prestige oil spill: a disastrous management of the sinking, but an extraordinary relief action by the government. How did voters react to this? How did they solve the task of weighting up two contradictory inputs? The 2003 Spanish municipal elections took place in all 8,108 municipalities throughout the country. Your interest lies in those municipalities that were affected by the accident (henceforth dubbed *Prestige*). Since the oil spread throughout the coast, it is extremely difficult to have precise estimates about the extent of damage incurred in each municipality. One solution could be to denote as affected all the coastal municipalities of the region ($n = 56$). Other solutions can also be suggested. Figure 1 presents the treated municipalities located in the coast of Galicia. Note that other parts of the Spanish coast were affected, too.

¹ The political controversy generated around the issue did also give birth to a grassroots popular movement (the so-called, *Plataforma Nunca Más*) that sought to fight against the most negative consequences of the accident at the environmental level, to blame the national (and the regional) incumbents for their bad performance during the catastrophe, and to avoid the occurrence of similar situations in the future.

Figure 1: Affected Coastal Areas



Your task is to develop a research design aimed at isolating the effect of *Prestige* on the vote share of the incumbent party. Make sure to include the following information:

1. A clear definition of your unit of analysis as well as treatment and control groups.
2. The identification assumption of your estimation strategy and how you want to examine whether it holds or not.
3. Does your research design allow you to isolate the effect of the two channels (the effect of the incumbent's response to the disaster and the effect of the transfers of aid)? If yes, describe how so.
4. The potential limitations of your research design.

Exercise 2: Variation & Counterfactuals

In your own words, describe what the counterfactual is for the potential outcomes of the control group that allows you to estimate the treatment effect between the treated and control groups in the following research designs. (You are welcome to use hypothetical empirical examples if you like, but you are not required.):

1. Time *and* unit fixed effects in panel data.
2. Exact matching using three covariates in cross sectional data.
3. Difference-in-differences estimator in panel data.
4. Sharp Regression Discontinuity Design LATE estimator.

Exercise 3: The Number of Candidates and Political Representation

In the aftermath of the 2016 US Presidential election, Jill Stein, the defeated candidate of the Green party, said in an interview that turnout would have been lower if she had not competed. When asked about the possibility that her candidacy contributed to the defeat of Hillary Clinton, Stein claimed that her supporters would have abstained if they could not have voted for her. This general intuition is shared by many: in countries in which political competition revolves around a few candidates like in the United States, the United Kingdom, or Canada, people whose preferences do not resonate with any of the candidates are more likely to abstain because they feel *alienated* by the system (for a review of the literature, see Blais, 2006). This is a strong case in favor of elections with more than two candidates, and flexible state regulation facilitating candidacy, as a way to increase turnout and reduce inequalities in political representation (Gallego, 2014).

Several comparative and cross-sectional studies have investigated this question, but with mixed results: some find that the number of candidates decreases turnout (e.g. Jackman, 1987), some that it increases it (e.g. Taagepera et al, 2013), and yet others that it has no effect (e.g., Fornos et al, 2004).

Question a) Discuss why cross-sectional studies across countries **and** across regions within a given country could lead to mixed results. What type of selection bias could be present in the country cross-sectional studies and what type of selection bias could be present in within country regional cross-sectional studies? Write out a model specification you imagine they estimate and discuss sources of omitted variables bias using formal notation.

Question b) You learn that France has an interesting twist to their electoral rules on how many candidates run in legislative and cantonal elections. Legislative and cantonal elections in France are held under a two-round majority system. **The second round is held one week following the first one. The two candidates with the largest number of votes in the first round are automatically qualified to the second round. Yet, if another candidate receives a vote share higher than the qualifying threshold, they also qualify to the second round.** Importantly, the vote share for this threshold is calculated out of the number of registered voters. For legislative elections, the qualifying threshold is 12.5%. In cantonal elections, it is 10%, except in 2011 for which it was 12.5%. You also obtain official results of all French legislative and cantonal elections between 1978 and 2012, including close to 14,000 electoral district races. If a candidate receives a number of votes larger than 50% of the total number of valid votes in the first round, and if this number is higher than 25% of the total registered voters in the district, they are elected and there is no second round. These elections are excluded from this data set. The dataset, `france.dta`, is available on the course website (see the last page for variable descriptions. Do approach me if you have any questions about this).

1. As a first step, you should get to know the data: i) What is the unit of observation (a row) in the dataset? ii) What types of elections are covered in the data? iii) How many candidates are there in the first and second round on average? iv) What is the average turnout and null and blank votes in the second round if there are two candidates vs three candidates? What would you conclude from that? v) How many third ranked candidates in the first round pass the qualifying threshold? How many of those run in the second round? (*Hint: Use the variables `threshold` and `threshold_party_can3`.*)
2. What causal inference method would you use to study this question? Write out the model specification and discuss what the identification assumption is in this case.

3. How would you test the identification assumption? Choose at least one way using the data provided to empirically test that your choice of the research design is internally valid.
4. Estimate the effect of a third candidate by estimating the first and second stage on the main outcomes: voter turnout and null and blank votes. Operationalise the main outcome variables and the running variable as you see appropriate. Interpret the results, both in terms of the size of the estimates in the first and second stages and their statistical significance.

variable name	type	format	label	variable label
year	int	%10.0g		
id_canton	byte	%8.0g		
departement	str29	%29s		
election	str8	%9s		type of election
party_can1	str14	%14s		party of first ranked candidate
voteshare_t1~1	long	%8.0g		total votes in round 1 of first ranked candidate
voteshare_t2~1	long	%10.0g		total votes in round 2 of first ranked candidate
ran_t2_can1	float	%9.0g		whether first ranked candidate ran in round 2
ideology_can1	float	%9.0g		ideology of first ranked candidate
party_can2	str14	%14s		party of second ranked candidate
voteshare_t1~2	long	%8.0g		total votes in round 1 of second ranked candidate
voteshare_t2~2	long	%10.0g		total votes in round 2 of second ranked candidate
ran_t2_can2	float	%9.0g		whether second ranked candidate ran in round 2
ideology_can2	float	%9.0g		ideology of second ranked candidate
party_can3	str14	%14s		party of third ranked candidate
voteshare_t1~3	long	%8.0g		total votes in round 1 of third ranked candidate
voteshare_t2~3	long	%10.0g		total votes in round 2 of third ranked candidate
threshold_par~3	float	%9.0g		qualifying share (votes/registered voters) for third candidate in first round
ran_t2_can3	float	%9.0g		whether third ranked candidate ran in round 2
ideology_can3	float	%9.0g		ideology of third ranked candidate
inscrits_t1	long	%12.0g		Registered voters, 1st round
parties_t1	float	%9.0g		Number of candidates, 1st round
elected_t1	float	%9.0g		Elected in first round
parties_t2	float	%9.0g		Number of candidates, 2nd round
threshold	float	%9.0g		qualifying threshold for third ranked candidate
turnout_t1	float	%9.0g		Turnout, 1st round
cvotes_t1	float	%9.0g		Valid Turnout, 1st round
blancsnull_t1	float	%9.0g		Null/Blank Votes, 1st round
turnout_t2	float	%9.0g		Turnout (2nd)
cvotes_t2	float	%9.0g		Valid Turnout (2nd)
blancsnull_t2	float	%9.0g		Null/Blank Votes (2nd)
nosecond	float	%9.0g		No second round, 3 criteria
margin_t2	float	%9.0g		Votes share difference 1st-2nd cand
margin_t1	float	%9.0g		Votes share difference 1st-2nd cand
threecand	float	%9.0g		Three candidates in second round

Figure 2: Variables in the Data and their Labels