Causal Inference

Take-Home Exam

`Due Date: Friday of Week 0, Trinity Term`

`Please use your candidate number`

Exercise 1

Total points: 40

Natural disasters are often used as exogenous shocks to examine the extent to which incumbents are rewarded for providing services to their constituents (Wolfinger and Rosenstone 1980; Mettler and Stonecash 2008; Bechtel and Hainmueller 2011). In this exercise, we aim to evaluate two possible explanations for this phenomenon. On the one hand, voters may reward incumbents after a natural disaster because they are grateful for any help they may have received. On the other hand, voters may perceive politicians as more competent after a natural disaster, especially if they demonstrate skills that helped mitigate the consequences of the disaster.

Context of the Exercise

To examine these two potential mechanisms, we will look at a case where the incumbent did not generally handle the natural disaster well, but did provide significant financial support (disaster relief spending) to the affected areas. Specifically, we will examine the Prestige accident, which occurred in late 2002 in Spain off the coast of Galicia and involved a 26-year-old structurally deficient oil tanker. The accident resulted in a severe oil spill along the northern coast of Spain. Although the government arguably failed to manage the crisis, it generously compensated residents of the affected areas. Here is a brief chronicle of events:

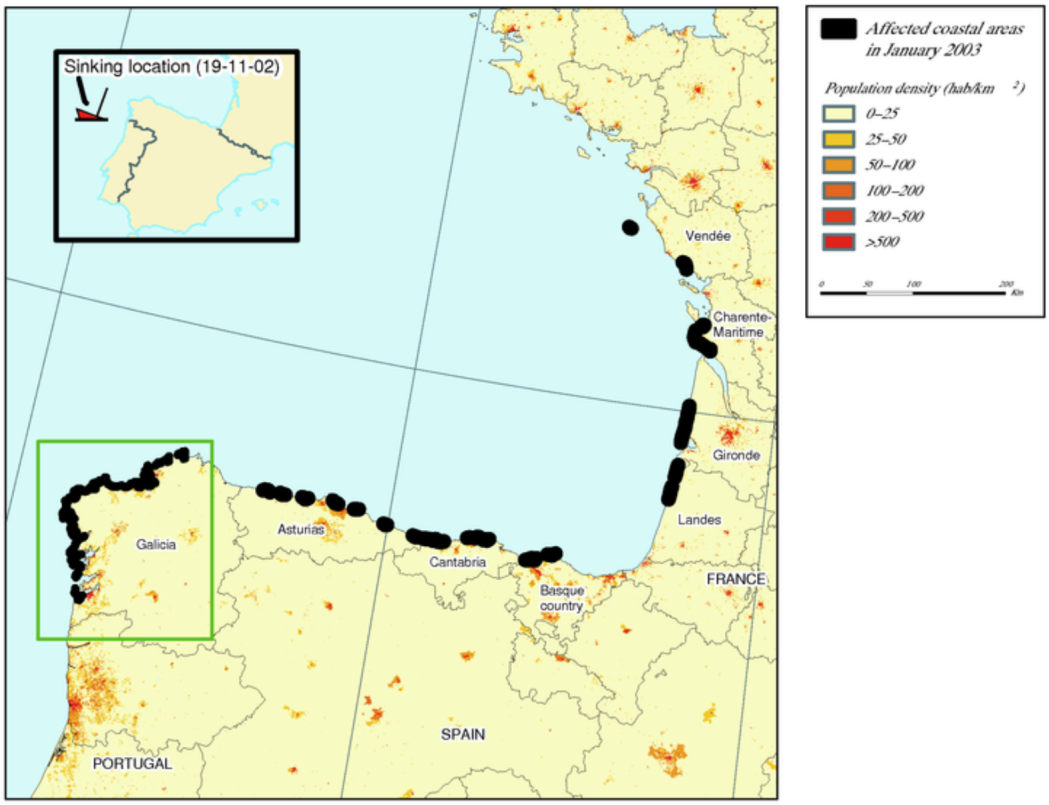
- On 13 November 2002, the Bahamas-registered oil tanker *Prestige*, carrying 77,000 tonnes of heavy fuel oil, broke in two off the coast of Galicia (Spain), spilling a significant amount of its cargo.
- Although the accident occurred within 5 miles of the Galician coast, the Spanish government (then in the hands of the conservative Popular Party [PP]) refused to give the tanker a

safe harbour. Instead, the government sent the tanker away from the Galician coast in a north-westerly direction.

- A longer series of miscalculations followed until the Prestige sank 150 miles off the coast on 19 November. This was the largest oil spill in terms of area since the Exxon Valdez disaster. In fact, the entire north Atlantic coast of Spain was affected, as well as parts of the Portuguese and French coastlines.

- As a result, pristine ecosystems were damaged and all fishing and seafood collection along the Galician coast was banned for months. Figure 1 shows the geographical area affected by the accident.

Figure 1: Affected Coastal Areas



The Prestige oil spill occurred only about six months before Spain's regional and local elections (held in 2003). Some political parties (e.g. the Galician Nationalist Party [BNG], the United Left [IU] and the Socialist Party [PSOE]) blamed the ruling party in Madrid, calling the central government "incompetent" and presenting the accident as evidence of the government's disregard for Galician issues. Some argue that the disaster has provided future Prime Minister Rodriguez

Zapatero and the main opposition party - the Socialist Party (PSOE) - with an important opportunity to win over voters and improve their performance in the upcoming elections.
In addition, the *regional* Galician government launched a major aid programme (the *Plan Galicia*), which included rapid and massive transfers of aid to citizens in the affected areas. The first payments were made two months after the disaster. These funds were considered particularly important given that Galicia's main economic asset is its fishing industry. To sum up, the Prestige oil spill was characterised by two key features: poor management of the spill itself, but an extraordinary response from the government. *How did the voters react? How did they weigh up these two contradictory actions?*

Working Information

The 2003 Spanish municipal elections were held in 8,108 municipalities across the country, including Galicia. In this exercise, we are particularly interested in the municipalities affected by the accident (henceforth dubbed *Prestige*). Since the oil spread all along the coast, it is difficult to make precise estimates of the damage in each municipality. One, but not the only, possible solution to this challenge is to simply consider all Galician coastal municipalities ($n = 56$) as affected (or "treated", in statistical jargon). Take another look at Figure 1, which shows the treated municipalities on the Galician coast. Note, however, that other parts of the Spanish coast were also affected.

Tasks

Your task is to develop a *hypothetical* research design aimed at estimating the causal effect of the *Prestige* accident on the then incumbent's vote share. You should use one of the methods we have covered in the course. Pay particular attention to the following subtasks:

1. Explain what kind of data you would collect. Clearly define your units of analysis as well as your treatment and control group.

2. Explain and justify your identification strategy for estimating the causal effect of interest. If you are estimating a model (e.g. OLS, DiD, etc.), provide the equation and explain all its terms. What are the identification assumptions of your research design and what do they mean in the context you are studying here? Can you test them? If so, how?

3. Can your research design disentangle the effects of the two proposed mechanisms, namely (i) the initial central government response to the emergency and (ii) the subsequent transfer of aid? If yes, explain your reasoning in detail. If not, explain how you could modify your originally proposed research design or data collection to separate the two mechanisms.

4. Elaborate on your research design's limitations and threats to causal identification.

Exercise 2

Total points: 60

Context of the Exercise

Following the 2016 US presidential election, Jill Stein, the defeated Green Party candidate, said in an interview that voter turnout would have been lower if she had not run. When asked about the possibility that her candidacy contributed to Hillary Clinton's defeat, Stein claimed that her supporters would have abstained if they had not been able to vote for Stein. This intuition is consistent with some political science research: in countries where political competition revolves around a few candidates (such as the United States, the United Kingdom or Canada), people are more likely to abstain from voting if they have preferences that do not align with any of the major candidates, e.g. because they feel alienated from the system (for a review of the literature, see Blais, 2006). As a result, some scholars argue that elections with (i) more than two competitive candidates and (ii) flexible state regulations that facilitate candidacy can increase voter turnout and reduce inequalities in political representation (Gallego, 2014). Several existing comparative and cross-sectional studies have examined this issue. However, their findings are mixed: some find that the number of candidates reduces turnout (e.g. Jackman, 1987), others that it increases turnout (e.g. Taagepera et al., 2013), while others find null effects (e.g. Fornos et al., 2004).

Working Information

We will use data from France in this exercise (the dataset, *france.dta*, is available on Canvas). France has an interesting twist to its electoral rules in terms of the number of candidates standing in legislative and cantonal elections. More specifically, legislative and cantonal elections in France are held under a two-round majority system. The second round takes place one week after the first round. The two candidates with the most votes in the first round automatically qualify for the second round. However, other candidates also qualify if their share of the vote exceeds a certain qualifying threshold. Importantly, the vote share for this threshold is calculated on the basis of the number of **registered voters**. For legislative elections, the qualifying threshold is 12.5%. For cantonal elections it is 10%, except in 2011 when it was also 12.5%. Note that our

dataset includes official election results for all French legislative and cantonal elections between 1978 and 2012 - around 14,000 electoral district races. Note also that a candidate is immediately elected in the first round if his vote share exceeds (i) 50% of the total number of valid votes in the first round, and (ii) 25% of the total number of registered voters in his constituency. See Table 1 for details of the variables in this dataset.

Table 1: Variable Overview

Variable	Type	Format	Variable Label
year	numeric	%10.0g	
id_canton	numeric	%8.0g	
departement	character	%29s	
election	character	%9s	type of election
party_can1	character	%14s	party of first ranked candidate
voteshare_t1_can1	numeric	%8.0g	total votes in round 1 of first ranked candidate
voteshare_t2_can1	numeric	%10.0g	total votes in round 2 of first ranked candidate
ran_t2_can1	numeric	%9.0g	whether first ranked candidate ran in round 2
ideology_can1	numeric	%9.0g	ideology of first ranked candidated
party_can2	character	%14s	party of second ranked candidate
voteshare_t1_can2	numeric	%8.0g	total votes in round 1 of second ranked candidate
voteshare_t2_can2	numeric	%10.0g	total votes in round 2 of second ranked candidate
ran_t2_can2	numeric	%9.0g	whether second ranked candidate ran in round 2
ideology_can2	numeric	%9.0g	ideology of second ranked candidated
party_can3	character	%14s	party of third ranked candidate
voteshare_t1_can3	numeric	%8.0g	total votes in round 1 of third ranked candidate
voteshare_t2_can3	numeric	%10.0g	total votes in round 2 of third ranked candidate
threshold_party_can3	numeric	%9.0g	qualifying share (votes/registered voters) for third candidate in first round
ran_t2_can3	numeric	%9.0g	whether third ranked candidate ran in round 2
ideology_can3	numeric	%9.0g	ideology of third ranked candidated
inscrits_t1	numeric	%12.0g	Registered voters, 1st round
parties_t1	numeric	%9.0g	Number of candidates, 1st round
elected_t1	numeric	%9.0g	Elected in first round
parties_t2	numeric	%9.0g	Number of candidates, 2nd round
threshold	numeric	%9.0g	qualifying threshold for third ranked candidate
turnout_t1	numeric	%9.0g	Turnout, 1st round
cvotes_t1	numeric	%9.0g	Valid Turnout, 1st round
blancsnull_t1	numeric	%9.0g	Null/Blank Votes, 1st round
turnout_t2	numeric	%9.0g	Turnout (2nd)
cvotes_t2	numeric	%9.0g	Valid Turnout (2nd)
blancsnull_t2	numeric	%9.0g	Null/Blank Votes (2nd)
nosecond	numeric	%9.0g	No second round, 3 criteria
margin_t2	numeric	%9.0g	Voteshare difference 1st-2nd cand
margin_t1	numeric	%9.0g	Voteshare difference 1st-2nd cand
threecand	numeric	%9.0g	Three candidates in second round

Tasks

1. Familiarise yourself with the data by answering the following questions: i) What are the units of observation (rows) in the dataset? ii) What types of elections does the dataset cover? iii) How many candidates stood in the first and second rounds in each election? iv) What is the average turnout and null & blank votes in the second round of elections when two and three candidates stand in the second round? v) How many candidates who get the third highest share of votes in the first round qualify for the second round by exceeding the qualifying threshold? How many of them actually run in the second round? (Hint: Use the variables *threshold* and *threshold_party_can3*.)

2. Think about a causal inference method covered in the course that allows you to investigate, using the data provided, whether increasing the number of candidates standing in an election reduces the abstention rate. Explain why this design is a good choice in this setting. Discuss its identification assumption(s) in the context you are studying here.

3. Conduct at least two empirical tests to show that your choice of research design is internally valid (falsification checks).

4. Finally, estimate the causal effect of a third candidate standing in the second round, using the research design of your choice. Consider two main outcome variables: (1) turnout and (2) null and blank votes. Select or construct appropriate variables to operationalise the main outcome variables and the running variable. Interpret the results and discuss the effect size of your estimates in the first/second stage and their statistical significance.