University of Oxford: MPhil in Politics

Causal Inference: Take-Home Exam

1090063

Contents

1 Problem 1: Natural Disasters and Voting Behaviour

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.

1.1 Data Collection

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

To investigate the causal effect of the *Prestige* oil spill on the incumbent party's vote share, whilst also disentangling the mechanisms of of the government's response to the disaster and the subsequent transfer of aid, a panel data set measured at the municipality unit of analysis is required.

To help disentangle the mechanisms, there will be two treatments: (i) the oil spill treatment and (ii) the payment treatment. The oil spill treatment is a binary variable indicating whether a municipality was affected by the oil spill disaster, while the payment treatment is a binary variable indicating whether a municipality received government aid in response to the disaster. Therefore, the control groups are those municipalities that were not affected by the oil spill and did not receive government aid.

The primary dependent variable is the vote share of the incumbent party in each municipality, measured at three time points: before the oil spill treatment, after the oil spill treatment, and after the payment treatment, with the pre-election time periods being measured as vote intention.

Below, `Table 1` is a mock-up of the panel data set that would be collected:

Table 1: Example Panel Dataset of Data Collected

Municipality ID	**Time Period**	**Controls**	**Oil Spill Treated**	**Payment Treated**	**Vote Share**
1	0	$\mathbf{X}_{1t}$	0	0	0.72
2	2	$\mathbf{X}_{2t}$	0	0	0.35
3	2	$\mathbf{X}_{3t}$	1	1	0.15
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
i	$t = [0\|1\|2]$	$\mathbf{X}_{it}$	$spill = [0\|1]$	$pay = [0\|1]$	$\mathbf{Vote}_{it}$

- `Municipality ID`: A unique identifier for each municipality (unit of analysis), indexed by i.
- `Time Period`: Indicates the time point t at which the observation is recorded. The dataset is structured as a panel, with multiple observations for each municipality over time. Before the oil spill treatment $t = 0$, after the oil spill treatment $t = 1$, and after the payment treatment $t = 2$. The time periods are indexed by t.
- `Controls`: Denotes a vector of time-variant observed covariates for each municipality at time t, represented as $\mathbf{X}_{it}$. This vector of covariates would include socio-economic, geographic, or political variables used to adjust for confounding.
- `Oil Spill Treated`: A binary indicator for whether a municipality was affected by the oil spill natural disaster. Equal to 1 if treated, 0 otherwise. This indicates whether a municipality is in the initial treatment or control group.
- `Payment Treated`: A binary variable indicating whether a municipality received a government payment as aid in response to the disaster. Equal to 1 if aid was received, 0 otherwise. This indicates whether a municipality is in the second treatment or control group.
- `Vote Share`: Vote share is the proportion of voting intentions measured for the incumbent party in each municipality at time t. This variable is used to measure electoral outcomes and public sentiment towards the incumbent after the disaster. It is denoted as $\mathbf{Vote}_{it}$.

1.2 Identification Strategy

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?

To estimate the causal effect of the *Prestige* oil spill disaster on the vote share of the incumbent party, a generalised Difference-in-Differences (DiD) with two-way fixed effects will be used. In particular, along with the panel dataset, this DiD model is used to allow for the estimation of the causal effect of the oil spill treatment and the payment treatment on the vote share of the incumbent party simultaneously in one regression specification, while controlling for time-variant covariates.

The generalised DiD approach works by calculating how much the treated group's outcomes changed compared to the control group's outcomes before and after the treatment. The two-way fixed effects approach accounts for unobserved time-invariant characteristics of the municipalities and common shocks that affect all municipalities at the same time.

The equation for the generalised DiD model is as follows:

$$\mathbf{Vote}_{it} = \alpha_i + \lambda_t + \beta_1(\text{spill}_i \times \text{post\_spill}_t) + \beta_2(\text{pay}_i \times \text{post\_pay}_t) + \gamma \mathbf{X}_{it} + \varepsilon_{it} \quad (1)$$

Where:

Table 2: Explanation of Terms in theGeneralised DiD Regression Model

Term	Description
$\mathbf{Vote}_{it}$	Share of vote intentions for the incumbent party in municipality i at time t.
α_i	Municipality-level fixed effect, capturing all time-invariant characteristics of municipality i.
λ_t	Time fixed effect, capturing all time-specific shocks that affect all municipalities at time t.
β_1	Coefficient for the interaction between the oil spill treatment and the post-treatment period.
spill_i	Binary indicator for whether municipality i was affected by the oil spill.
post_spill_t	Binary indicator for post-treatment period for the oil spill.
β_2	Coefficient for the interaction between the payment treatment and the post-treatment period.
pay_i	Binary indicator for whether municipality i received government aid.
post_pay_t	Binary indicator for post-treatment period for the payment.
γ	Vector of coefficients for the covariates.
$\mathbf{X}_{it}$	Vector of time-variant covariates for municipality i at time t.
ε_{it}	Error term of unobserved factors affecting vote share in municipality i at t.

For this DiD model to be valid, the core assumption of parallel trends must hold. This assumption states that if there was no treatment, the average outcomes for the treated and control groups would have followed the same trend over time. In this context, it means that the vote share of the incumbent party in municipalities affected by the oil spill would have followed the same trend as those not affected by the oil spill if the disaster had not occurred.

This assumption cannot be directly tested due to the fact that we cannot observe the counterfactual of whether the oil spill did not happen. However, we can conduct a pre-treatment analysis to check whether the trends in vote share for the treated and control groups were similar before the oil spill occurred. This can be done by plotting the average vote share over time for both groups and checking to see if they follow the same trend. This is potentially a problematic assumption due to spillover effects from the oil spill treatment. For example, if the oil spill affected the views and thus vote share of the incumbent party in neighbouring municipalities that were not directly affected by the spill, this could violate the parallel trends assumption.

We also are assuming that we are able to control for all time-variant confounders that may affect the vote share of the incumbent party. This is done by including a vector of time-variant covariates $\mathbf{X}_{it}$ in the regression model. If these are unobserved confounders that are not included and are correlated with both the treatment and the outcome, this could bias our estimates. For example, there can be other political events other than the oil spill that could affect the vote share of the incumbent party, especially in the run up to an election. If these events are not controlled for, they could confound the relationship between the oil spill and the vote share. However, ppsephological research suggests that voter intentions are generally stable over time and it is hard to persuade a change in behaviour in the run up to an election. Therefore, it is likely that the time-variant covariates will be sufficient to control for confounding.

1.3 Disentangling the Mechanisms

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.

As previously mentioned, the design of the research design is to use a generalised DiD approach with two-way fixed effects. Using the panel data spanning multiple time periods, this allows for a generalised model to simultaneously estimate the causal effect of the oil spill treatment and the payment treatment on the vote share of the incumbent party. This is done by including both treatments in the regression model as separate interaction terms with their respective post-treatment indicators.

To estimate the initial central government response, we want to see how the share of vote intentions for the incumbent party changes after the oil spill treatment between those municipalities that were affected by the oil spill and those that were not. This is calculated by the interaction term of $(spill_i \times post\_spill_t)$. By running an OLS regression, the coefficient β_1 estimates the average difference in the change in voting intentions from the pre-spill period ($t = 0$) to the post-spill periods ($t = 1$, $t = 2$) for the spill-affected municipalities, relative to the change experienced by the control municipalities over the same periods.

Then, to estimate the subsequent transfer of aid, we want to see how the share of vote intentions for the incumbent party changes after the payment treatment between those municipalities that received government aid and those that did not. This is calculated by the interaction term of $(pay_i \times post\_pay_t)$. Here, β_2 estimates the average additional difference in the change in voting intentions from the pre-compensation period ($t = 1$) to the post-compensation period ($t = 2$) for the compensated towns, relative to the change experienced by the uncompensated spill-affected towns over that same specific period ($t = 1$ to $t = 2$). The control group here are the municipalities that were affected by the oil spill but did not receive government aid, whilst also controlling for time-based trends through the use of time fixed effects.

By using two-way fixed effects, we can control for the unobserved time-invariant characteristics of the municipalities and common shocks that affect all municipalities at the same time. This allows us to isolate the effects of the two treatments on the vote share of the incumbent party through the specific interaction terms and their coefficients to be comparing only the relevant control and treatment groups for each mechanism.

1.4 Limitations and Threats

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

With this research design comes a few possible limitations and threats to having non-biased and consistent OLS estimates, as well as accurately estimating the causal nature of the effects. As previously mentioned, the parallel trends assumption is a key assumption of the DiD model. If this assumption does not hold, the estimates of the treatment effects may be biased. This could happen if there are other confounding factors that affect the vote share of the incumbent party in the treated and control groups differently over time. For example, if there were other political events or changes in public opinion that affected the vote share of the incumbent party in one group but not the other, this could violate the parallel trends assumption.

Another threat is the extent and scale of the treatments of the oil spill and the compensation received. A core assumption thus far is that the effects of the oil spill and the compensation are homogenous across all municipalities. However, for example, the severity of the oil spill may have varied across municipalities, leading to different levels of impact on businesses and livelihoods, which could affect voting intentions differently across municipalities. Variation in compensation levels should be less of an issue as all municipalities within Galicia received compensation, and as the unit of analysis is the municipality, the average compensation received should be similar across municipalities.

One of the biggest limitations to the research design comes from the quality and accuracy of the panel data set created to analyses these effects. The nature of the panel data requires the same municipalities to be observed over time, which can be difficult to achieve in practice. If there are missing data points, or differences in how the data is collected over time, this can lead to bias in the estimates. In particular, this will be problematic when estimating the share of the voting intentions for the incumbent party. This will require accurate sampling from an identically and independently distributed sample of the municipality each time period is measured. If the sample is not representative of the population, this can lead to bias in the estimates, with only the final election results showing the true vote shares. This means that the estimates of **Vote**_{it} may be biased.