Natural and Quasi-Experiments

Political Analysis, Week 6 (MT 2015)
Oxford Q-Step Centre
Prof. David Kirk
Outline

- Examples of natural experiments
- Research designs making use of natural experiments:
  - Difference-in-differences
  - Instrumental variables
- Applications related to (1) the media influence on voting, and (2) civil conflict
Random vs. Natural Experiments

- Unlike true randomized experiments, natural experiments arise from naturally occurring phenomena that are often the product of social and political forces.

- Manipulation of the treatment is not under the control of the analyst; rather the “natural” intervention approximates the characteristics of a randomized experiment.
  - Some kind of external situation creates a counterfactual.

- Then why do we still use the word “experiment?”
  - Because we can credibly make the case that subjects were assigned to treatment or control “as if” they have been randomly assigned.
  - Need to make the case that subjects do not self-select into the treatment or control group.
London Cholera Epidemic, 1853-54

- Pandemics of cholera affected several regions of the world throughout the 19th century, including London.
- A common hypothesis was that transmission was air-born.
- Dr John Snow – a physician and later to be known as a one of the fathers of epidemiology – hypothesized that cholera was actually carried through the water.
- To test this hypothesis, he capitalized on a natural experiment.

Dr John Snow
Snow’s Natural Experiment

- Some parts of London were served by two different water companies: Lambeth (red) and Southwark & Vauxhall (green).

- In the overlapping areas, pipes from both water companies went down the same streets, and the particular company supplying water to a home was unrelated to the characteristics of the home or occupants (i.e., “as-if” random).

- Snow (1855: 74-75) stated: “Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different companies”
• The Lambeth Company moved its water intake source upstream in 1852, thereby obtaining a water supply free from the sewage of London

• Arguably there were otherwise identical homes in the same parts of London, where a “treatment” group of homes was supplied by Lambeth water from an uncontaminated source and the “control” group of homes was supplied by Southwark and Vauxhall from a contaminated source

<table>
<thead>
<tr>
<th>Snow’s Table IX</th>
<th>Number of Houses</th>
<th>Deaths from Cholera</th>
<th>Deaths Per 10,000 Houses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southwark and Vauxhall</td>
<td>40,046</td>
<td>1,263</td>
<td>315</td>
</tr>
<tr>
<td>Lambeth</td>
<td>26,107</td>
<td>98</td>
<td>37</td>
</tr>
</tbody>
</table>
Examples of Other Natural Experiments

➢ Explore the effect of policy shifts on households on one side of a political/jurisdictional boundary versus households on the other side of the boundary not subject to the policy change (e.g., Card and Krueger 1994; Miguel 2004; Krasno and Green 2008).

➢ Assume that households/people on opposite sides of the border are roughly similar

➢ Miguel (2004) examined the effect of ethnic policies on funding for public goods (e.g., education, infrastructure) in western Kenya and western Tanzania. He argued that these regions have similar geographies and colonial institutional legacies, so that the main difference was ethnic policies by the government

➢ Sudden surge in immigrants (Card 1990)

➢ Card examined the sudden surge of immigrants into Miami, Florida in 1980 (called the Mariel boatlift) on the wages and employment rates among existing low-skilled workers in Miami
More Natural Experiments

- Exogenous health shocks:
  - Almond (2006) used the 1918 flu epidemic to examine the effect of in utero conditions on long-term health
  - Several scholars have used the Dutch famine of 1994-45 to examine the long-term physical and mental health effects of starvation and malnutrition (Lumey and van Poppel 2013)
  - Rich et al (2015) used the sudden and temporary improvement in air quality in Beijing during the 2008 Olympics to examine the health effects of air pollution

- Weather as a natural experiment: the effect of economic and housing shocks induced by heavy rainfall or hurricanes (e.g., Miguel et al. 2004; Kirk 2009)
Additional Natural Experiments in Politics and IR

- Lottery winnings as an exogenous source of economic affluence, to predict the effect of affluence on political attitudes (Doherty, Gerber, and Green 2006)

- Ansolabehere et al (2000) used political redistricting as a natural experiment to study the voting advantages of incumbents

- The influence of Western media on regime change (Kern and Hainmueller 2009)
  - Because of East Germany’s topography, West German television broadcasts could be received in some but not all of East Germany. The authors exploit this naturally occurring variation to estimate the effect of West German media exposure on support for the East German regime.
Ladd and Lenz (2009), from Pol Soc core paper


- Examined the effect of the media on voting patterns
- Some scholars argue that the media simply reinforces pre-existing views and attitudes, and that individuals seek out media outlets that adhere to their political views
  - Hence, media influence may actually be limited
Ladd and Lenz’s Research Design

- Media outlets tend to support the same political parties from election to election (Ansolabehere et al. 2006) – e.g., the *New York Times* endorses Democrats and has done so for decades.

- Following the 1992 UK election as well as a recession, several long-time media supporters of the Conservative Party (e.g., *Sun*) suddenly switched endorsements to Labour.

- The authors argued that this sudden shift created a natural experiment for examining the effect of the media on voting.
<table>
<thead>
<tr>
<th>Table 1</th>
<th>British Newspapers’ Party Orientations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>What Paper Did You Read Most Often? (%)</td>
</tr>
<tr>
<td>Doesn’t read newspaper</td>
<td>30.4</td>
</tr>
<tr>
<td>Consistent Conservative</td>
<td></td>
</tr>
<tr>
<td>Daily Express (Scottish)</td>
<td>5.6</td>
</tr>
<tr>
<td>Daily Mail</td>
<td>9.9</td>
</tr>
<tr>
<td>Daily Telegraph</td>
<td>5.5</td>
</tr>
<tr>
<td>Consistent Labour</td>
<td></td>
</tr>
<tr>
<td>Daily Mirror/Record</td>
<td>18.1</td>
</tr>
<tr>
<td>Guardian</td>
<td>2.3</td>
</tr>
<tr>
<td>Switched to Labour</td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>9.7</td>
</tr>
<tr>
<td>Daily Star</td>
<td>1.4</td>
</tr>
<tr>
<td>Independent</td>
<td>1.6</td>
</tr>
<tr>
<td>Financial Times</td>
<td>0.7</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Times</td>
<td>2.5</td>
</tr>
<tr>
<td>Other/Not answered</td>
<td>12.3</td>
</tr>
<tr>
<td>n =</td>
<td></td>
</tr>
</tbody>
</table>

Coding of newspaper slant is based on Seymour-Ure (1997), Norris (1998), and Scammell (1997). Readership percentages are from the 1996 wave of the BEPS. Circulation data are from Seymour-Ure (1997).
Research Design and Results

- **Treatment Group**: readers of newspapers that switched endorsements
- **Untreated/Control Group**: non-newspaper readers or readers of newspapers with stable endorsements
- Use a difference-in-differences design to mimic random assignment; use time trend of untreated group as the “counterfactual”—i.e., to establish what would have happened with the treatment group if they had not actually received the treatment
Parallel Trends Assumption:

- In the absence of any kind of change in media endorsements, the temporal change in % voting for Labour would be the same for Treated and Untreated groups.

- The red line is the counterfactual: what would have happened with the treatment group if their newspapers had not switched endorsements (slope of red line is the same as Untreated line)

This is the treatment effect: 15.2 – 6.6 = 8.6
DID Formula: \( (Y^T_1 - Y^T_0) - (Y^C_1 - Y^C_0) \)

\( Y^T_0 = 38.9\% = \% \text{ voting for Labour among Treated group at first time point (time = 0)} \)

\( Y^T_1 = 58.3\% = \% \text{ voting for Labour among Treated group at second time point (time = 1)} \)

\( Y^T_1 - Y^T_0 = 58.3\% - 38.9\% = 19.4\% = \text{ increase in vote for Labour in Untreated group} \)

\( Y^C_0 = 32.2\% = \% \text{ voting for Labour among Untreated group at first time point (time = 0)} \)

\( Y^C_1 = 43\% = \% \text{ voting for Labour among Untreated group at second time point (time = 1)} \)

\( Y^C_1 - Y^C_0 = 43\% - 32.2\% = 10.8\% = \text{ increase in vote for Labour in Untreated group} \)

\( (Y^T_1 - Y^T_0) - (Y^C_1 - Y^C_0) = 19.4\% - 10.8\% = 8.6\% = \text{ treatment effect} \)
Benefit of the Natural Experiment and DID Design:

- If we just compared the Treated and Untreated groups on the % voting for Labour in 1997, we would incorrectly overestimate the size of the effect (15.2 instead of 8.6).

- Comparing the trends from 1992 to 1997 instead of just the % at 1997 allows us to adjust for the fact that the % voting for Labour in 1992 in each group was different (i.e., pre-treatment difference between the groups that led to some selection bias).
Benefit of the Natural Experiment and DID Design:

- If we just compared the vote % for Labour in 1997 for the Treated group (58.3%) with the vote % in 1992 (38.9%), we would overestimate the size of the effect (19.4 instead of 8.6).

- Comparing the trends from 1992 to 1997 for the Treated and Untreated allows us to adjust for the fact that there were factors besides just the switch in newspaper endorsements that affected the vote for Labour.


- Examined the impact of economic conditions on the likelihood of civil conflict
- Hypothesis: economic growth is negatively related to civil conflict
- Challenges: whereas economic conditions and civil conflict are likely correlated, each may be explained by the same antecedent factor, such as government institutional quality
- Also, the causal ordering may be reversed:
  - civil conflict → economic conditions
Selection Bias from Omitted Variables

Estimating the causal impact of economic conditions ($T$) on the likelihood of civil conflict ($Y$) is complicated by the issue of omitted variable bias (i.e., if $T$ is correlated with the omitted variable $W$, then the estimated effect of $T$ on $Y$ will be biased).
Technical Details

Full model:
\[ Y_i = B_0 + B_1 T_i + B_2 W_i + \varepsilon_i \]

But if \( W \) is omitted, the coefficient for the treatment effect ultimately yields:
\[ B_{1*} = B_1 + B_2 \gamma \]

The difference between \( B_1 \) & \( B_{1*} \) is the omitted variable bias (i.e., \( B_2 \gamma \)). The coefficient for the treatment variable \( T \) absorbs the effect of the omitted variable \( W \).
Technical Details

Full model:
\[ Y_i = B_0 + B_1 T_i + B_2 W_i + \varepsilon_i \]

But if \( W \) is omitted, the coefficient for the treatment effect ultimately yields:
\[ B_{1*} = B_1 + B_2 \gamma \]

There are two scenarios where \( B_1 = B_{1*} \) (meaning there is no omitted variable bias (i.e., \( B_2 \gamma = 0 \)).

1) \( B_2 = 0 \) (e.g., no association between the dependent variable (civil conflict, \( Y \)) and the omitted variable (government institutional quality, \( W \))

2) \( \gamma = 0 \) (e.g., no association between the treatment variable (economic growth, \( T \)) and the omitted variable (government institutional quality, \( W \))
Selection Bias from Omitted Variables

• An instrumental variable \( Z \) for \( T \) is one solution to the problem of omitted variable bias.

• A valid instrument is a variable that is (1) correlated with the treatment variable \( T \), yet is (2) uncorrelated with the dependent variable \( Y \) except through the treatment \( T \).

• Miguel et al.’s strategy: use rainfall as an instrumental variable \( Z \) to predict economic growth \( T \)

• “Weather shocks are plausible instruments for growth in gross domestic product in economies that largely rely on rain-fed agriculture, that is, neither have extensive irrigation systems nor are heavily industrialized.” (P.726)
Instrumental Variables & a Natural Experiment

Two-stage estimation:

1) Predict the treatment ($T_i$) – economic growth – with the instruments (rainfall, $Z_i$) and a vector of control variables ($X_i$).

2) Estimate the causal effect of the economic growth (i.e., the predicted $T_i$ from the 1st equation) on the incidence of civil conflict ($Y_i$).
Results, Miguel et al. (2004)

A five-percentage-point decline in annual economic growth increases the likelihood of a civil conflict in the following year by over 12 percentage points.
Challenges with Instrumental Variables

• Recall that a valid instrument is a variable that is (1) correlated with the treatment variable $T$, yet is (2) uncorrelated with the dependent variable $Y$ except through the treatment $T$

• It is often very challenging to find such a variable

• In the case of Miguel et al., they made a credible argument that the weather (rainfall levels) were only indirectly related to civil conflict, through the effect on economic conditions.
Summary

- **Randomized experiment**
  - Treated and Untreated groups are randomly similar to each other, even on omitted variables

- **Covariate adjustment (statistical control)**
  - Attempt to control for all factors that determine treatment and that are related to the dependent variable
  - But...we have no idea how many Ws there are, and/or cannot measure some likely Ws

- **Difference-in-Differences**
  - Multiple time points of data allows us to control for characteristics of our units of analysis (e.g., people, countries) that do not vary over time.

- **Instrumental Variables**
  - Instrument induces quasi-experimental variation in treatment status