Lecture 21: Diff-in-Diff

Criminology 250

Prof Maria Cuellar

University of Pennslyvania

Motivation

What if you don't have a randomized experiment? How do you calulate causal effects?

- Natural experiments, or quasi-experiments: Observational study in which an event or a situation that allows for the random or seemingly random assignment of study subjects to different groups is exploited to answer a particular question.
- Less than perfect: Natural experiments are often used to study situations in which controlled experimentation is not possible, such as when an exposure of interest cannot be practically or ethically assigned to research subjects.
- Examples: Situations that may create appropriate circumstances for a natural experiment include policy changes, weather events, and natural disasters. Natural experiments are used most commonly in the fields of epidemiology, political science, psychology, and social science.

Reference: https://www.britannica.com/science/natural-experiment

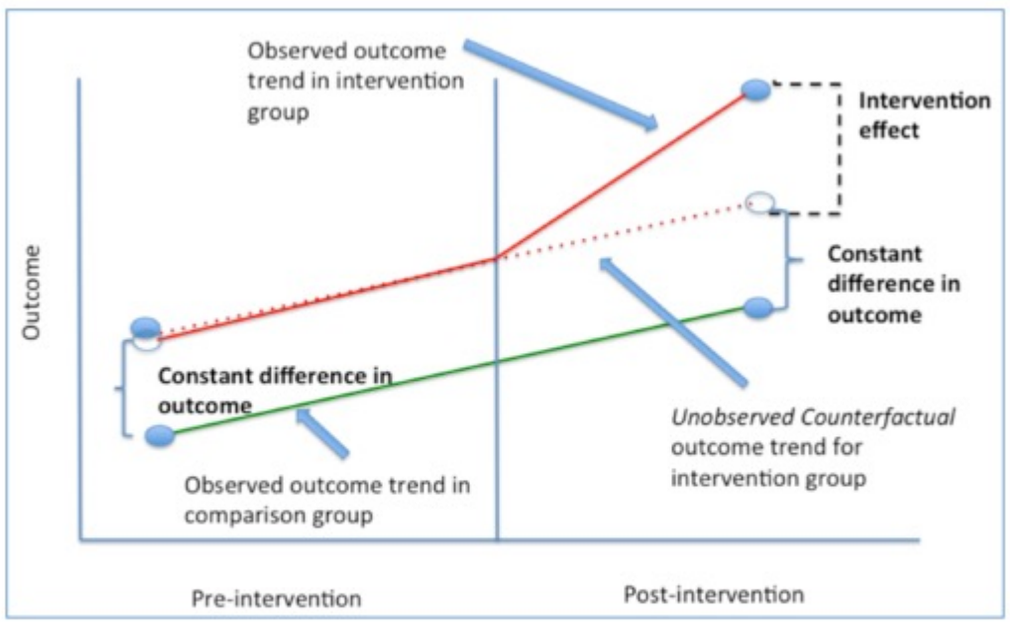
Three methods for estimating causal effects in observational studies

- Difference-in-differences
- Regression discontinuity
- Instrumental variables

Difference-in-differences (Diff-in-diff)

John Snow

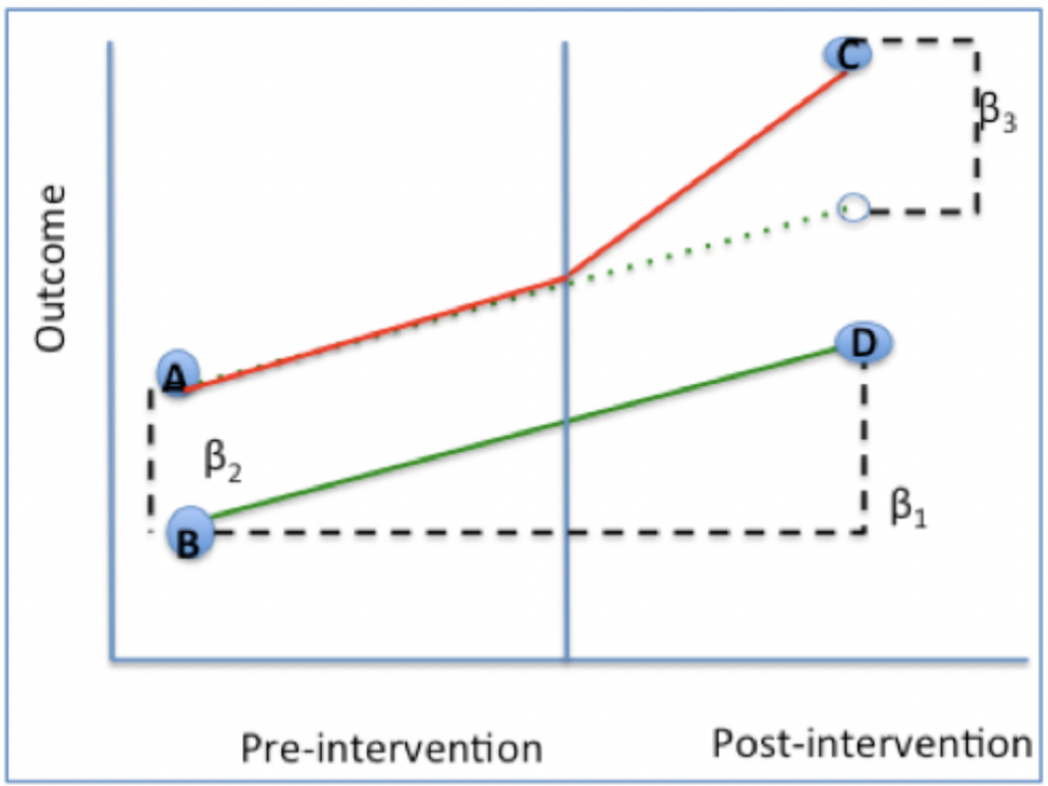
Difference-in-differences (Diff-in-diff)



- Origin: Technique originated in the field of econometrics, but the logic underlying the technique has been used as early as the 1850's by John Snow.
- John Snow: Snow exploited the effective randomization in the design and applied a rudimentary form of Difference-in-Differences (DiD) to compare dirty-versus-clean and before-versus-after.
- Data: Uses longitudinal data (i.e. a measurement before shock and one after schock).
- Assumptions: DID is used in observational settings where exchangeability cannot be assumed between the treatment and control groups. DID relies on a less strict exchangeability assumption, i.e., in absence of treatment, the unobserved differences between treatment and control groups are the same over time.

Diff-in-diff estimation

Coefficient	Calculation	Interpretation
β_0	B	Baseline average
β_1	D-B	Time trend in control group
β_2	A-B	Difference between two groups pre-intervention
β_3	(C-A)-(D-B)	Difference in changes over time



Diff-in-diff parameter of interest vs. estimator

Parameter of interest: Average effect of treatment on the treated (ATT): This compares the potential outcomes with treatment to the potential outcomes with no treatment, in the treated group.

Here, $Y^{a}(t)$ is the potential outcome given treatment a (0 or 1, not treated or treated) at time t (1 or 2, before or after).

$$ATT = E(Y^{1}(2) - Y^{0}(2)|A = 1)$$

Estimator: Once we have made some causal, identification assumptions, we can estimate this by using the Diff-in-diff estimator:

$$\hat{\beta} = (\bar{Y}_{\text{after,treatment group}} - \bar{Y}_{\text{after, control group}}) - (\bar{Y}_{\text{before,treatment group}} - \bar{Y}_{\text{before, control group}})$$

One way to use the estimator

Regression model: Diff-in-diff is usually implemented as an interaction term between time and treatment group dummy variables in a regression model.

$$Y = \beta_0 + \beta_1 * \text{Time} + \beta_2 * \text{Intervention} + \beta_3 * \text{Time} \times \text{Intervention} + \beta_4 * \text{Covariates} + \epsilon.$$

Reference: AP, Mostly Harmless Econometrics: Chapter 5.2 (pg 169-182)

Diff-in-diff assumptions

- Exchangeability, positivity, and SUTVA (more on this in the next slide).
- Intervention unrelated to outcome at baseline (allocation of intervention was not determined by outcome).
- Parallel Trends: Treatment/intervention and control groups have Parallel Trends in outcome

References: https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation

Parallel trend assumption

- The parallel trend assumption is the most critical of the above assumptions to ensure internal validity of Diff-in-diff models and is the hardest to fulfill.
- It requires that in the absence of treatment, the difference between the ‘treatment’ and ‘control’ group is constant over time. Although there is no statistical test for this assumption, visual inspection is useful when you have observations over many time points.
- It has also been proposed that the smaller the time period tested, the more likely the assumption is to hold. Violation of parallel trend assumption will lead to biased estimation of the causal effect.

SUTVA: Stable Unit Treatment Values Assumption

- SUTVA (Stable unit treatment value assumption): the [potential outcome] observation on one unit should be unaffected by the particular assignment of treatments to the other units. This goes beyond the concept of independence. (Cox 1958, §2.4)

It requires:

1. No interference: The potential outcomes for any unit do not vary with the treatments assigned to other units.
2. Consistency: No hidden variations of treatment. For each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes. (Imbens and Rubin 2015, 10; Keele 2015b, 5)

Controversy around consistency (Keele 2015, Keele, L. (2015). The statistics of causal inference: A view from political methodology. Political Analysis, 23(3), 313-335. (SUTVA discussion on page 317.)

The first component of SUTVA is often referred to as the consistency assumption in the epidemiological literature, and under this assumption we assume that for units exposed to a treatment we observe the potential outcomes for that treatment. The consistency assumption is somewhat controversial. Hernán and VanderWeele (2011) argue that the consistency assumption must be evaluated by analysts since it links observed data to the counterfactual outcomes. They argue that in the absence of consistency, one would not know which counterfactual contrast is being estimated by the data, which makes it difficult to base decision-making on a causal analysis. For example, if the treatment were "15 min of exercise," there are many different forms of exercise. They contend that it will be difficult to justify any decision-making based on effect estimates since we may not know which form of exercise actually made the treatment effective. In contrast, van der Laan, Haight, and Tager (2005) say that consistency is an axiom which can be taken for granted, while Pearl (2010) maintains that consistency immediately follows so long as the causal model is correct. In some sense, there are elements of truth to both sides. If potential outcomes are independent of the exercise treatment, we can rule out the presence of other causes. However, generating policy recommendations about this treatment may be difficult given the fact that the treatment may contain a large number of components.

The second part of the SUTVA assumption tends to be a more serious problem in many social science settings. The problem is that if we treat a unit and that unit can then spread some of that treatment to a control unit or units, the comparison is no longer between treated and control, but between a treated unit and partially treated unit. If one specifies a model of contagion for how the treatment spreads, one can make some progress toward identification, but if we have no knowledge of treatment spillovers, causal parameters will not be identified. Taking interference into account is currently a very active area of research in both the social sciences and statistics (Sinclair, McConnell, and Green, 2012; Tchetgen and VanderWeele 2012; Bowers, Fredrickson, and Panagopoulos 2013). Treatments that vary over time can also lead to SUTVA violations, as well as other complications. There has been considerable focus on treatments that vary over time in the biostatistics literature, and I suspect time-varying treatments will eventually be a topic of interest in political science (Robins 1997, 1999).

Strengths and Limitations

Strengths

- Intuitive interpretation
- Can obtain causal effect using observational data if assumptions are met
- Can use either individual or group level data
- Comparison groups can start at different levels of the outcome. (Diff-in-diff focuses on change rather than absolute levels)
- Accounts for change due to factors other than intervention

Limitations

- Requires baseline data & a non-intervention group
- Cannot use if intervention allocation determined by baseline outcome
- Cannot use if composition of groups pre/post change are not stable
- Cannot use if comparison groups have different outcome trend

Example in R

```
# Getting sample data.
library(foreign)
mydata = read.dta("http://dss.princeton.edu/training/Panel101.dta")
head(mydata)
```

```
##   country year           y y_bin        x1         x2          x3   opinion
## 1       A 1990  1342787840     1 0.2779036 -1.1079559  0.28255358 Str agree
## 2       A 1991 -1899660544     0 0.3206847 -0.9487200  0.49253848     Disag
## 3       A 1992   -11234363     0 0.3634657 -0.7894840  0.70252335     Disag
## 4       A 1993  2645775360     1 0.2461440 -0.8855330 -0.09439092     Disag
## 5       A 1994  3008334848     1 0.4246230 -0.7297683  0.94613063     Disag
## 6       A 1995  3229574144     1 0.4772141 -0.7232460  1.02968037 Str agree
##   op
## 1  1
## 2  0
## 3  0
## 4  0
## 5  0
## 6  1
```

Example in R

```r
# Create a dummy variable to indicate the time when the treatment started. Let's assume that treatment start
mydata$time = ifelse(mydata$year >= 1994, 1, 0)


# Create a dummy variable to identify the group exposed to the treatment. In this example lets assume that c
mydata$treated = ifelse(mydata$country == "E" |
                        mydata$country == "F" |
                          mydata$country == "G", 1, 0)


# Create an interaction between time and treated. We will call this interaction 'did'.
mydata$did = mydata$time * mydata$treated


# Estimating the DID estimator
didreg = lm(y ~ treated + time + did, data = mydata)
```

Example in R

```
##
## Call:
## lm(formula = y ~ treated + time + did, data = mydata)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -9.768e+09 -1.623e+09  1.167e+08  1.393e+09  6.807e+09
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.581e+08  7.382e+08   0.485   0.6292
## treated      1.776e+09  1.128e+09   1.575   0.1200
## time         2.289e+09  9.530e+08   2.402   0.0191 *
## did         -2.520e+09  1.456e+09  -1.731   0.0882 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.953e+09 on 66 degrees of freedom
## Multiple R-squared:  0.08273,    Adjusted R-squared:  0.04104
## F-statistic: 1.984 on 3 and 66 DF,  p-value: 0.1249
```

Example in R

The coefficient for ‘did’ is the differences-in-differences estimator. The effect is significant at 10% with the treatment having a negative effect.

Example in R

```r
# Can also estimate the Diff-in-diff effect (multiplying the two variables in the formula, no need to genera
didreg1 = lm(y ~ treated*time, data = mydata)
summary(didreg1) # Check, but it's the same as before
```

```
##
## Call:
## lm(formula = y ~ treated * time, data = mydata)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -9.768e+09 -1.623e+09  1.167e+08  1.393e+09  6.807e+09
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.581e+08  7.382e+08   0.485   0.6292
## treated       1.776e+09  1.128e+09   1.575   0.1200
## time          2.289e+09  9.530e+08   2.402   0.0191 *
## treated:time -2.520e+09  1.456e+09  -1.731   0.0882 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```