# Empirical Methods in Economics

Liyousew G. Borga

February 17, 2016

## The Evaluation Problem

# “What if ...”: Answering Counterfactual Questions

Empirical methods in economics have been developed to try to answer “counterfactual” questions.

- What would have happened to this person’s behavior if she had been subjected to an alternative treatment?

## “What if ...”: Answering Counterfactual Questions

Empirical methods in economics have been developed to try to answer “counterfactual” questions.

- What would have happened to this person’s behavior if she had been subjected to an alternative treatment?

We are concerned about identifying the effect of some action (e.g. a policy, an investment) on one or more outcomes of interest

# “What if ...”: Answering Counterfactual Questions

Empirical methods in economics have been developed to try to answer “counterfactual” questions.

- What would have happened to this person’s behavior if she had been subjected to an alternative treatment?

We are concerned about identifying the effect of some action (e.g. a policy, an investment) on one or more outcomes of interest

The goal of the analysis is to “rule out” other possibilities /explanations for the observed effects (internal validity)

- The effect of counseling job search program for the unemployed youth
- The effect of education on wages
- The effect of migration influx on local labor market
- The effect of competition between schools on schooling quality

# The Evaluation Problem

There are two complementary economic approaches for quantitative analysis

- "structuralist": theoretical model of agents' behavior. Could be used for "ex ante" evaluation, to anticipate the result of a new policy, or a change in parameters (e.g. new rules for retirement system)
- "empirical": testing the impact of a policy, without formally modeling the agent's reactions. "reduced-form/experimentalist" approach

The methods presented in this class are related to this latter trend

**Take the "con" out of econometrics**

*"Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously" (Leamer, 1983)*

**Take the "con" out of econometrics**

*"Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously" (Leamer, 1983)*

**Taking the "Econ" out of Econometrics too?**

*"People think about the question less than the method . . . so you get weird papers, like sanitation facilities in Native American reservations" (Chetty, 2007)*

*"In some quarters of our profession, the level of discussion has sunk to the level of a New Yorker article" (Heckman, 2007)*

Suppose we wish to measure the impact of treatment on an outcome

We have a population of units; for each unit we observe a variable $D$ and a variable $Y$

- We observe that $D$ and $Y$ are correlated. Does correlation imply causation?

We would like to understand in which sense and under which hypotheses one can conclude from the evidence that $D$ causes $Y$

# The Evaluation Problem

Suppose we wish to measure the impact of treatment on an outcome

**Treatment**

$D_i$: Indicator of treatment intake for individual $i$

$$D_i = \begin{cases} 1 & \text{if individual } i \text{ received the treatment} \\ 0 & \text{otherwise} \end{cases}$$

**Outcome**

$Y_i$: Observed outcome variable of interest for individual $i$

**Potential Outcomes**

$Y_{0i}$ and $Y_{1i}$: Potential outcomes for individual $i$

$Y_{1i}$: Potential outcome for individual $i$ with treatment

$Y_{0i}$: Potential outcome for individual $i$ without treatment

The treatment effect or the causal effect of treatment is $\Delta_i = Y_{1i} - Y_{0i}$

# The Evaluation Problem: Potential Outcomes

The observed outcome $Y_i$ can be written in terms of potential outcomes as:

**The "Rubin Causal Model"**

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$
$$Y_i = \alpha + \beta_i D_i + \mu_i$$

A fundamental problem is that we cannot observe both $Y_{0i}$ and $Y_{1i}$ for each individual. We can therefore not directly observe:

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

We need to estimate the average effect of treatment by comparing average outcomes of those who were and those who were not treated.

# The Evaluation Problem: Selection Bias

Comparisons of outcome for the treated and the untreated do not usually give the right answer:

**Selection Bias**

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{observed difference in outcome}}=\underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{average treatment effect on the treated}}+\underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{selection bias}}$$

# The Evaluation Problem: Selection Bias

Comparisons of outcome for the treated and the untreated do not usually give the right answer:

**Selection Bias**

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{observed difference in outcome}}=\underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{average treatment effect on the treated}}+\underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{selection bias}}$$

- The selection bias term is not likely to be zero for most public policy applications
- It tells us that, beside the effect of the treatment, there may be systematic differences between the treated and the non-treated group

# The Evaluation Problem: Selection Bias

Comparisons of outcome for the treated and the untreated do not usually give the right answer:

## Selection Bias

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{observed difference in outcome}}=\underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{average treatment effect on the treated}}+\underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{\text{selection bias}}$$

- The selection bias term is not likely to be zero for most public policy applications
- It tells us that, beside the effect of the treatment, there may be systematic differences between the treated and the non-treated group

Causality is defined by potential outcomes, not by realized (observed) outcomes; observed association is neither necessary nor sufficient for causation

# The Evaluation Problem: Assignment Mechanism

Assignment mechanism is the procedure that determines which individuals are selected for treatment intake

- Random assignment
- Selection on observables / Selection on unobservables

# The Evaluation Problem: Assignment Mechanism

Assignment mechanism is the procedure that determines which individuals are selected for treatment intake

- Random assignment
- Selection on observables / Selection on unobservables

## Random assignment

Random assignment of treatment makes treatment $D_i$ independent of potential outcomes $Y_i$ ($Bias = E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ , but if $D_i$ is independent of $Y_i$, then $E[Y_{0i}|D_1] = E[Y_{0i}|D_0]$)

$$
\begin{aligned}
E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
&= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\
&= E[Y_{1i} - Y_{0i}|D_i = 1] \\
&= E[Y_{1i} - Y_{0i}]
\end{aligned}
$$

The selection bias term therefore vanishes with random assignment

**Treatment:**

A fertilizer program where fertilizers are given for free to some farmers

**Effect**

(Yield for the farmers who got fertilizer) - (Yield at the same point in time for the same farmers in the absence of the program)

**Problem**

We never observe the outcome of the same individual with and without program at the same point in time

**Solution**

Compare before and after?

**Solution**

Compare before and after?

- Other things may have happened over time

**Solution**

Compare before and after?

Simply compare those who get fertilizers with those who did not get?

**Solution**

Compare before and after?

Simply compare those who get fertilizers with those who did not get?

- Some may choose not to participate/ those not offered somehow participate

# Practical Examples: Fertilizer Application

## Solution

Compare before and after?

Simply compare those who get fertilizers with those who did not get?

Find a valid Counterfactual

- Find a good proxy for what would have happened to the outcome in the absence of program
- Compare the farmer with someone who is exactly like her but who was not exposed to the intervention
- Make sure that the **only reason** for different outcomes between treatment and counterfactual is the intervention

**Do hospitals make people healthier?**

Consider a poor elderly population that uses hospital emergency rooms for primary care. What will be the answer?

- YES, because those admitted to the hospital get many valuable services
- NO, exposure to other sick patients by those who are themselves vulnerable might adversely affect their health

What does the data say? Compare the health status of those who have been to the hospital to the health of those who have not

**Do hospitals make people healthier?**

| Group | Observations | Mean health status | Std. dev. |
|---|---|---|---|
| Hospital | 7,774 | 2.79 | 0.014 |
| No Hospital | 90,049 | 2.07 | 0.003 |

Health status ranges between 1-excellent and 5-bad
Difference in the means $= 0.71$
$t - statistic = 58.9$
*Source:* MHE, Ch.2, p10

This result suggests that going to the hospital makes people sicker

# Practical Examples: Effect of Hospitalization

**Do hospitals make people healthier?**

The comparison is unfair:

- People who go to hospital are, on average, less healthy than people who never get hospitalized in the first place
- People go to hospitals because they already have poor health conditions
- This is what we called the "selection bias"
- Random assignment, once again, solves the selection problem

**Do subsidized training programs help people find jobs?**

The comparison is unfair:

- People who need training later have higher unemployment rates
- People who need training programs already have less ability to find jobs than people who don't
- We need to compare individuals with the same observed characteristics
- However, some important characteristics such as innate ability and social skills are usually not observed
- This is what we call the "unobserved heterogeneity" problem

# Evaluation Methods

Constructing the counterfactual in a convincing way is a key requirement of any serious evaluation method

- Social experiments methods (RCTs)
- Natural experiments
- Matching methods
- Instrumental methods
- Discontinuity design methods

All are an attempt to deal with endogenous selection (assignment)

# Randomized Experiments: The "Gold Standard"?

The ideal set-up to evaluate the effect of a policy $X$ on outcome $Y$ is a randomized experiment.

- A sample of $N$ individuals is selected from the population
- This sample is then divided randomly into two groups: the Treatment group ($N_T$ individuals) and the Control group ($N_C$ individuals); $N_T + N_C = N$
- The Treatment group is then treated by policy $X$ while the Control group is not
- The outcome Y is observed and compared for both Treatment and Control groups
- The effect of policy $X$ is measured in general by the difference in empirical means of $Y$ between Treatments and Controls:

$$\hat{D} = \hat{E}(Y|T) - \hat{E}(Y|C)$$

**Attrition Bias**

**Randomization Bias**

**Hawthorne and John Henry Effects**

**Substitution Bias**

**Supply Side Changes**

**Cost, Ethics, Power, and Generalizability**

**Attrition Bias**

- Attrition rates (i.e. leaving the sample between the baseline and the follow-up surveys) may be different in treatment and control groups
- The estimated treatment effect may therefore be biased

**Randomization Bias**

**Hawthorne and John Henry Effects**

**Substitution Bias**

**Supply Side Changes**

**Cost, Ethics, Power, and Generalizability**

# Randomized Experiments: Limitations

**Attrition Bias**

**Randomization Bias**

- Can occur if treatment effects are heterogeneous
- The experimental sample may be different from the population of interest because of randomization
- People selecting to take part in the randomized trial may have different returns compared to the population average

**Hawthorne and John Henry Effects**

**Substitution Bias**

**Supply Side Changes**

**Cost, Ethics, Power, and Generalizability**

# Randomized Experiments: Limitations

**Attrition Bias**

**Randomization Bias**

**Hawthorne and John Henry Effects**

- People behave differently because they are part of an experiment and cause bias ("Hawthorne" effects)
- If people from the control group behave differently ("John Henry" effects)

**Substitution Bias**

**Supply Side Changes**

**Cost, Ethics, Power, and Generalizability**

**Attrition Bias**

**Randomization Bias**

**Hawthorne and John Henry Effects**

**Substitution Bias**

- Control group members may seek substitutes for treatment. This would bias estimated treatment effects downwards
- Can also occur if the experiment frees up resources that can now be concentrated on the control group

**Supply Side Changes**

**Cost, Ethics, Power, and Generalizability**

**Attrition Bias**

**Randomization Bias**

**Hawthorne and John Henry Effects**

**Substitution Bias**

**Supply Side Changes**

- If programmes are scaled up the supply side implementing the treatment may be different
- In the trial phase the supply side may be more motivated than during the large scale roll-out of a programme

**Cost, Ethics, Power, and Generalizability**

# Randomized Experiments: Limitations

**Attrition Bias**

**Randomization Bias**

**Hawthorne and John Henry Effects**

**Substitution Bias**

**Supply Side Changes**

**Cost, Ethics, Power, and Generalizability**

- Experiments are very costly and difficult to implement properly
- Substantial economic or social outcomes of the Treated
- Samples are often small (e.g. when unit of randomization is a group)
- Difficult to generalize the results of an experiment to the total population

# Examples of RCT: The Tennessee STAR

### The Effect of Class Size on Educational Achievement

What if a student had only 15 classmates instead of 30? Krueger (1999) econometrically re-analyses a randomized experiment of the effect of class size on student achievement

- The project is known as Tennessee Student/Teacher Achievement Ratio (STAR) and was run in the 1980s
- 11,600 students and their teachers were randomly assigned to one of three groups
  - Small classes (13-17 students)
  - Regular classes (22-25 students)
  - Regular classes (22-25 students) with a full time teacher's aide
- Randomization occurred within schools
- In the STAR experiment $D_i$ (being in a small class) is randomly assigned and therefore the selection bias disappears

## The Effect of Class Size on Educational Achievement

Krueger estimates the following econometric model:

$$Y_{ics} = \beta_0 + \beta_1 SMALL_{cs} + \beta_2 Reg/A_{cs} + \beta_3 X_{ics} + \alpha_s + \varepsilon_{is}$$

where the omitted category is regular class, and $\alpha_s$ is School FE because random assignment occurred within schools

- The estimated treatment-control differences for kindergartners, show a small-class effect of about 5 to 6 percentile points
- Attrition problem: attrition is likely to be non-random: especially good students from large classes may have enrolled in private schools
- Krueger's solution: imputing test scores (from their earlier test scores) for all children who leave the sample
- Non-Compliance problem: students changed classes after random Assignment

# Examples of RCT

- Ashraf, Karlan and Yin (2006): Time-Inconsistency (procrastination)
- DellaVigna, List, and Malmendier (2010) - Charitable Giving
- Miguel and Kremer (2004) - Deworming and school outcomes
- Bertrand and Mullainathan (2004): Are Emily and Greg more employable than Lakisha and Jamal?

# Empirical Methods in Economics

Liyousew G. Borga

February 17, 2016

## Controlling for Observables

# Causality

- We often observe that two variables are correlated
  - Parental Income is correlated with child's education
  - Pupil performance is correlated with the performance of peers.
  - Advertising is correlated with firm cash flow
  - Health and Income are correlated

# Causality

- We often observe that two variables are correlated
- However, this does not establish causal relationships. If a variable $Y$ is causally related to $X$, then changing $X$ will LEAD to a change in $Y$

- We often observe that two variables are correlated
- However, this does not establish causal relationships. If a variable $Y$ is causally related to $X$, then changing $X$ will LEAD to a change in $Y$
- In order to explain how $Y$ varies with changes in $X$,
  - we need to specify the functional relationship between $Y$ and $X$
  - we need to be sure that we are capturing the *ceteris paribus* relationship between $Y$ and $X$

# The Regression Model

- The basic tool in Econometrics is the Regression Model. Its simplest form is the two variable regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where the error term ($\varepsilon$) reflects *all other* factors than $X$ that affect $Y$

### An Example: Measuring the returns to education

A very common example in economics: what are the returns to education?

- Public Policy towards education is predicated on the assumption that education has economic value
- Education is supposed to promote earnings growth and as a result overall economic growth

A simple approach is to compare funds advanced for an extra year of education to the stream of earnings (i.e. the causal effect of extra education on earnings)

$$wage = \beta_0 + \beta_1 educ + \varepsilon$$

- The error term ($\varepsilon$) may include: experience, ability, personal characteristics, ...
- Under certain assumptions, we can capture a ceteris paribus relationship between *wage* and *educ*

# Assumptions of the Simple Regression Model

1. Linear in parameters
   - restrictive, but we may still be able to model non linear relationships in variables
2. Zero Conditional Mean: $E(\varepsilon|X) = 0$
   - $E(\varepsilon) = 0$; $Cov(X, \varepsilon) = 0$
3. Observations are drawn from random distributions
   - Too strong in some applications. For example, if we want to study female wages and we observe the salaries of working women

Given the first two assumptions (Linearity and Zero Conditional Mean), it is true that the population regression function (PRF) is a linear function of $X$:

$$E(Y|X) = \alpha + \beta X_i$$

## An Example: Measuring the returns to education

$$wage = \beta_0 + \beta_1 educ + \varepsilon$$

- assume for simplicity that $\varepsilon$ is innate ability
- $E(ability|educ = 0)$ denotes the average ability for the group of people with no education, and $E(ability|educ = 12)$ denotes the average ability among people in the population with 12 years of education
- Then the zero conditional mean assumption implies that $E(ability|educ = 0) = E(ability|educ = 12)$. In fact, the average level of ability is the same for all levels of education
- Since we do not observe innate ability, there is no way of knowing whether or not average ability is the same for all levels of education

# Threats to identification

Potential Violations of Conditional Mean Independence

- Wrongly specified model
- Omitted Variable Bias
- Measurement error
- Simultaneous causality

# Threats to identification

Potential Violations of Conditional Mean Independence

- Wrongly specified model
  - Example: *wages* do not depend linearly on *educ*
- Omitted Variable Bias
- Measurement error
- Simultaneous causality

# Threats to identification

Potential Violations of Conditional Mean Independence

- Wrongly specified model
  - Example: *wages* do not depend linearly on *educ*
- Omitted Variable Bias
  - Explanatory variables are potentially correlated with missing variables. (e.g. We do not control for ability and ability is correlated with *educ*)
- Measurement error
- Simultaneous causality

# Threats to identification

Potential Violations of Conditional Mean Independence

- Wrongly specified model
  - Example: *wages* do not depend linearly on *educ*
- Omitted Variable Bias
  - Explanatory variables are potentially correlated with missing variables. (e.g. We do not control for ability and ability is correlated with *educ*)
- Measurement error
  - Measurement error in explanatory variables (e.g. a proxy to ability (IQ) is included, but measured with error)
- Simultaneous causality

# Threats to identification

Potential Violations of Conditional Mean Independence

- Wrongly specified model
  - Example: *wages* do not depend linearly on *educ*
- Omitted Variable Bias
  - Explanatory variables are potentially correlated with missing variables. (e.g. We do not control for ability and ability is correlated with *educ*)
- Measurement error
  - Measurement error in explanatory variables (e.g. a proxy to ability (IQ) is included, but measured with error)
- Simultaneous causality
  - Explanatory variables simultaneously determined with $Y$ (e.g. The level of education chosen depends on the expected returns to education)

# Matching on observables

- The main purpose of matching is to reproduce the treatment group among the non-treated, this way re-establishing the experimental conditions in a non-experimental setting
- Under certain assumptions, the matching method constructs the correct sample counterpart for the missing information on the treated outcomes had they not been treated by pairing each participant with members of non-treated group
- The matching assumptions ensure that the only remaining difference between the two groups is programme participation

## Basic Idea of Matching

- For each person who is enrolled in the program, match them with someone who is as similar as possible and not enrolled
- Compute the difference in outcomes for each match
- The treatment effect is the weighted average of these differences

$$ATT : \hat{D}_{ATT} = \frac{1}{N_T} \sum_{i=1}^{N_T} (Y_{i1}^{T} - Y_{j0}^{C})$$

where $N_T$ is the number of treated individuals, $Y_{i1}^{T}$ is a treated observation, and $Y_{j0}^{C}$ is the untreated observation that is matched with observation $i$

## The CIA: Conditional Independence Assumption

- The CIA (sometimes referred to as "unconfoundedness" or "selection on observables") requires that the common variables that affect treatment assignment and treatment-specific outcomes be observable
- Conditional on the set of observables $X$, the non-treated outcomes are independent of the participation status,

$$Y_{0i} \perp T_i | X_i$$

- This means that, conditional on $X$, treated and non-treated individuals are comparable with respect to the outcome $Y$ in the non-treatment case
- Thus, there is no remaining selection on the unobservable term. Only observable factors influence participation and outcome variable simultaneously
- The choice of the appropriate matching variables, $X$, is a delicate issue

# How to: Matching Methods

- In an ideal setup, the treatment effects are calculated by comparing individuals for whom the values of $X$ are identical (i.e. exact matching on observables)

| **Treated Group** | | | | **Control Group** | | | |
|---|---|---|---|---|---|---|---|
| **Age** | **Gender** | **Unemp** | **Univ.** | **Age** | **Gender** | **Unemp** | **Uni. Grad** |
| 19 | 1 | 3 | 0 | 24 | 1 | 8 | 1 |
| 35 | 1 | 12 | 1 | 38 | 0 | 2 | 0 |
| 41 | 0 | 17 | 1 | 55 | 0 | 10 | 1 |
| 23 | 1 | 6 | 0 | 23 | 0 | 2 | 1 |
| 55 | 0 | 21 | 1 | 34 | 1 | 20 | 0 |
| 27 | 0 | 4 | 1 | 41 | 0 | 17 | 0 |
| 24 | 1 | 8 | 1 | 19 | 1 | 3 | 0 |

- Typically in applied work, it is either impractical or impossible to divide up the data into $(X, D)$ specific cells because there are usually many $X$ variables and/or some or all of these may be continuous variables (i.e. there are typically no non-treated individuals in the data that have exactly the same $X$ values as a given treated individual.

# How to: Matching Methods

- Typically in applied work, it is either impractical or impossible to divide up the data into $(X, D)$ specific cells because there are usually many $X$ variables and/or some or all of these may be continuous variables (i.e. there are typically no non-treated individuals in the data that have exactly the same $X$ values as a given treated individual.
- This makes it more difficult to estimate the counterfactuals

# How to: Matching Methods

In the absence of an exact match, we instead revert to using the distance between the $X$. The main alternatives of controlling for observable variables in practice are:

- Nearest Neighbor Matching
- Kernel Matching
- Propensity Score Matching

## Nearest Neighbor Matching

- match with only the closest untreated individual ("nearest neighbor")
- assigns a weight 1 to the closest non-treated observation and 0 to all others

| $X_1$ | $X_0$ |
|---|---|
| 9.9 | 5.4 |
| 13.4 | 16 |
| 9 | 8.7 |
| 12.5 | 8.2 |
| 12.9 | 7 |
| 7.7 | 9.4 |
| 12.9 | 8 |
| 12.3 | 7.7 |
| 10.2 | 9.2 |
| 6.9 | 6.5 |

### Kernel Matching

- Kernel matching defines a neighborhood for each treated observation and constructs the counterfactual using all control observations within the neighborhood, not only the closest one

$$\hat{Y}_{i0} = \frac{1}{N_0} \sum_{T_j=0} w_j \cdot Y_j$$

- It assigns a positive weight to all observations within the neighbour while the weight is zero otherwise

### The Curse of Dimensionality

To make CIA plausible, we should use lots of observable characteristics; the more dimensions, the less likely that we will find an exact match

## The Curse of Dimensionality

To make CIA plausible, we should use lots of observable characteristics; the more dimensions, the less likely that we will find an exact match

- Classic bias vs. efficiency trade-off in choosing how many untreated observations $M$ to consider per treated individual $i$
- The more observations considered, the smaller the variance (and the bigger the bias)
- The optimal M balances the two effects

## Propensity Score Matching

To avoid the curse of dimensionality, we can try and reduce the problem to one dimension - the propensity score

- Use observable characteristics to compute the probability that an individual will enroll in treatment

$$p(X) = Pr(T = 1|X) \in [0, 1]$$

- $p(X)$ summarizes all the observed characteristics that influence the likelihood of being treated
- The propensity score reduces the multi-dimensional vector $X$ to a single-dimensional measure
- The propensity score is a balancing score. The conditional distribution of $X$ given $p(X)$ is independent of assignment to treatment. In other words, when looking at a subgroup of individuals with similar $p(X)$, the distribution of $X$ should be the same in group $T$ and $C$

## Propensity Score Matching: Few technical details

- To employ PSM, first regress the treatment dummy $T$ on the set of available controls $X$ (Probit or Logit)

  Record the predicted probability of treatment (i.e. calculate the propensity score - the fitted values $\hat{T}$)
- Restrict the sample to observations for which there is common support in the propensity score distribution (i.e. If for some values of X there are only treated (or only untreated) individuals, we can't match them with anyone)

$$0 < Pr(T = 1|X) < 1$$

- Match treated individuals with untreated individuals with similar propensity scores
- Check to see if $X$'s are balanced after matching, and calculate ATT/ATE

### Propensity Score Matching: Choice of Covariates

To make conditional independence credible

- Get as many characteristics as possible which may predict treatment
- Careful when dealing with one-off cross sections: Don't use variables which have been affected by treatment
- If we have before-after data, can use differences rather than absolute values
- Data collection challenges: Often studies have different data for treated and untreated individuals
- NOTE: Matching does not allow for heterogeneous treatment effects

## Testing the Identifying Assumption

- Placebo test:
  - If earlier cross-sectional data is available, apply the same matching procedure, but before treatment kicked in
  - Check that you don't get a significant difference ("treatment effect") between $T$ and $C$
- In PSM, count how many controls have a propensity score lower than the minimum or higher than the maximum of the propensity scores of the treated
  - Ideally we would like that the range of variation of propensity scores is the same in the two groups
- Generate histograms of the estimated propensity scores for the treated and the controls with bins corresponding to the strata constructed for the estimation of propensity scores

Example

### Testing the Identifying Assumption

The Lalonde (1986) Study

- very influential study on the validity of matching
- takes an existing RCT study on the effect of an employment programme on income
- finds a non-experimental control group and tries to replicate results using matching techniques
- gets very different results

# Example: Matching Methods

*Angrist (1998): the effect of voluntary military service on earnings later in life*

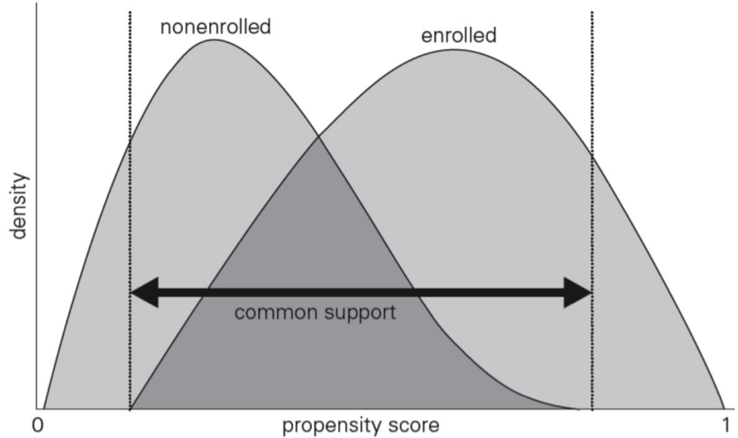
- This research asks whether men who volunteered for service in the US Armed Forces were economically better off in the long run
- Since voluntary military service is not randomly assigned, Angrist used matching and regression techniques to control for observed differences between veterans and non-veterans who applied to get into the all-volunteer forces between 1979 and 1982
- The motivation for a control strategy in this case is the fact that the military screens soldier-applicants primarily on the basis of observable covariates like age, schooling, and test scores
- The CIA is that after conditioning on all these observed characteristics veterans and nonveterans are comparable
- conditional on $X_i$, variation in veteran status comes solely from the fact that some qualified applicants fail to enlist at the last minute

# Example: Matching Methods

Jalan, Jyotsna and Martin Ravallion (2003): "Does Piped Water Reduce Diarrhea for Children in Rural India", *Journal of Econometrics*

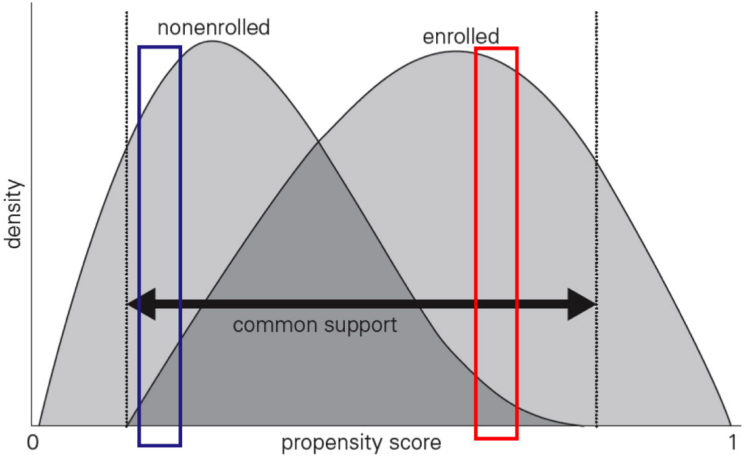
- The impacts of public investments that directly improve children's health are theoretically ambiguous given that the outcomes also depend on parentally provided inputs. Using propensity score matching methods, we find that the prevalence and duration of diarrhea among children under five in rural India are significantly lower on average for families with piped water than for observationally identical households without it. However, our results indicate that the health gains largely by-pass children in poor families, particularly when the mother is poorly educated. Our findings point to the importance of combining water infrastructure investments with effective public action to promote health knowledge and income poverty reduction.

# The Common Support



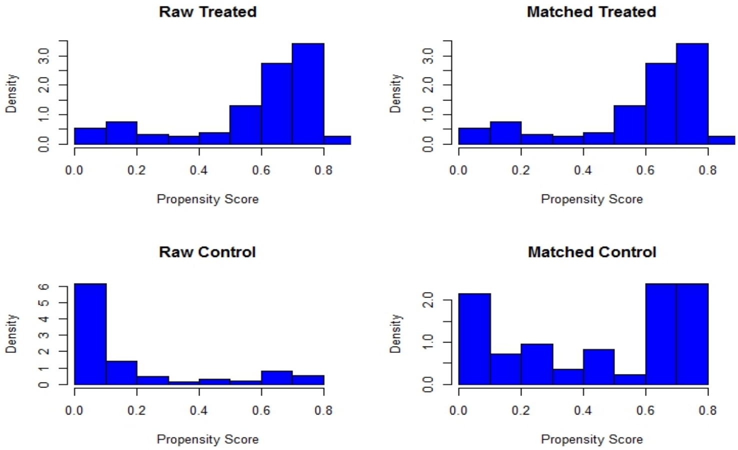

Back

# The Common Support




Back

# The Common Support




Back

# Empirical Methods in Economics

Liyousew G. Borga

February 17, 2016

## Selection on Unobservables

## John Snow's data journalism: The cholera map

- In the world of the 1850s, cholera was believed to be spread by miasma in the air,
- The sudden and serious outbreak of cholera in London's Soho was a mystery
- John Snow (1813-1858), a medical doctor, exploited a natural experiment to provide evidence contrasting the popular belief

## John Snow's data journalism: The cholera map

- In the world of the 1850s, cholera was believed to be spread by miasma in the air,
- The sudden and serious outbreak of cholera in London's Soho was a mystery
- John Snow (1813-1858), a medical doctor, exploited a natural experiment to provide evidence contrasting the popular belief
- During the 1849 and 1854 epidemics, some parts of London were simultaneously supplied by two water companies
- The Southwark Company pumped water from a dirty part of the Thames during both the 1849 and 1854 cholera epidemics while The Lambeth water company sourced its water from a dirty part of the Thames during 1849, and had moved to a cleaner part by 1854
- Areas supplied by the Southwark company had similar numbers of deaths during 1849 and 1854; Areas supplied by Southwark and Lambeth had similar number of deaths in 1849
- Deaths in Lambeth areas dropped sharply between 1849 and 1854, relative to the (small) change in Southwark deaths

# Natural or Quasi-Experiments

A popular method in empirical studies is exploiting naturally occurring exogenous variation to mimic a randomized experiment

- As random experiments are very rare, we rely on actual policy changes to identify the effects of policies on outcomes
- These are called "natural experiments" because we take advantage of changes that were not made explicitly to measure the effects of policies
- The key issue when analyzing a natural experiment is to divide the data into a control and treatment group
- The most obvious way to do that is to do a simple difference method using data before ($t = 0$) and after the change ($t = 1$); but it is difficult to distinguish the policy effect from a secular change
- Is the outcome before treatment a good counterfactual of the potential outcome without treatment in the treatment period?

# Difference-in-Differences Estimation

- A way to improve on the simple difference method is to compare outcomes before and after a policy change for a group affected by the change (Treatment Group) to a group not affected by the change (Control Group)
- Alternatively: instead of comparing before and after, it is possible to compare a region where a policy is implemented to a region with no such policy

| | **Pre** | **Post** | **ATE** |
|---|---|---|---|
| **Treatment** | $Y_1$ | $Y_2$ | $(Y_2 - Y_1)$-$(Y_4 - Y_3)$ |
| **Control** | $Y_3$ | $Y_4$ | |

- The idea is to correct the simple difference before and after for the treatment group by subtracting the simple difference for the control group

$$DiD = [\hat{E}(Y_1|T) - \hat{E}(Y_0|T)] - [\hat{E}(Y_1|C) - \hat{E}(Y_0|C)]$$

# Dif-in-Dif: An Example

## Card & Krueger (1994): the effect of a minimum wage increase

- Card & Krueger (1994) analyze the effect of a minimum wage increase in New Jersey using a differences-in-differences methodology
- In February 1992 NJ increased the state minimum wage from \$4.25 to \$5.05. Pennsylvania's minimum wage stayed at \$4.25
- One would expect a raise in the minimum wage to result in a decrease in employment
- They surveyed about 400 fast food stores both in NJ and in PA both before and after the minimum wage increase in NJ
- The differences-in-differences strategy amounts to comparing the change in employment in NJ to the change in employment in PA

# Dif-in-Dif: The Setup

To see this more formally:

$Y_{1ist}$ : employment at restaurant $i$ , state $s$, time $t$ with a high $w^{min}$

$Y_{0ist}$ : employment at restaurant $i$ , state $s$, time $t$ with a low $w^{min}$

- In the absence of a minimum wage change, employment is determined by the sum of a time-invariant state effect $\gamma_s$ and a year effect $\lambda_t$ that is common across states:

$$E(Y_{0ist}|s,t) = \gamma_s + \lambda_t$$

- Let $D_{st}$ be a dummy for high-minimum wage states and periods
- Assuming $E(Y_{1ist} - Y_{0ist}|s,t) = \delta$, the treatment effect, observed employment can be written:

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

# Dif-in-Dif: The Setup

- The typical regression model that we estimate is:

$$Outcome_{it} = \beta_0 + \beta_1 \cdot post_t + \beta_2 \cdot treat_i + \tau \cdot (post * treat)_{it} + \varepsilon$$

  where *post* is the treatment dummy and captures time effects, *treat* is a dummy if the observation is in the treatment group and captures constant differences between the two groups

- In the Card & Krueger case the equivalent regression model would be:

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$

  *NJ* is equal to 1 if the observation is from NJ, *d* is equal to 1 if the observation is from November

# Dif-in-Dif: The Setup

This equation takes the following values:

- PA Pre: $\alpha$
- PA Post: $\alpha + \lambda$
- NJ Pre: $\alpha + \gamma$
- NJ Post: $\alpha + \gamma + \lambda + \delta$

Differences-in-Differences estimate: $(NJ_{Post} - NJ_{Pre}) - (PA_{Post} - PA_{Pre}) = \delta$

| Variable | PA (i) | NJ (ii) | Difference, NJ – PA (iii) |
|---|---|---|---|
| | Stores by state | | |
| 1. FTE employment before, all available observations | 23.33 (1.35) | 20.44 (0.51) | –2.89 (1.44) |
| 2. FTE employment after, all available observations | 21.17 (0.94) | 21.03 (0.52) | –0.14 (1.07) |
| 3. Change in mean FTE employment | –2.16 (1.25) | 0.59 (0.54) | 2.76 (1.36) |

Surprisingly, employment rose in NJ relative to PA after the minimum wage change.

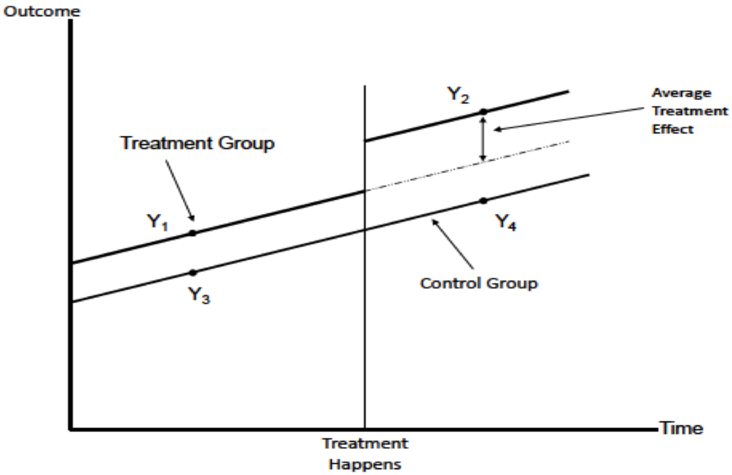
Card and Krueger (1994): Credible Results?

- The results came as quite a shock to most economists who thought employment would fall
- The study has been very controversial but helped to change the common presupposition that a small change in the minimum wage from a low level was bound to cause a significant decrease in employment
- Notice that we can see that prior to the increase in the minimum wage Pennsylvania had higher employment than New Jersey and that it was bound to fall to a lower level
- Minimum wage increase decided in early 1990 and implemented in April '92. The "Before" survey is conducted in February 1992. Could this announcement invalidate the identification strategy?
- Key identifying assumption: Employment trends would be the same in New Jersey and Pennsylvania, absent the change in the minimum wage

**Common (parallel) trends assumption**

- In the absence of treatment, average outcome of the treated group would have changed in the same way as the average outcome of the control group

# Dif-in-Dif: Identifying Assumption



### Checks of DiD strategy

- Use data for prior periods (say period -1) and redo the DiD comparing year 0 and year -1 (assuming there was no policy change between year 0 and year -1)
- If this placebo DiD is non zero, there are good chances that your estimate comparing year 0 and year 1 is biased as well
- Replace $Y$ by another outcome $Y'$ that is not supposed to be affected by the reform. If the DiD using $Y'$ is non-zero, then it is likely that the DiD for $Y$ is biased as well
- Use an alternative control group $C'$. If the DiD with the alternative control is different from the DiD with the original control $C$, then the original DiD is likely to be biased

# Dif-in-Dif: The Setup

- Consider a general case with $G$ groups (e.g. states, age groups, school classes, ...), $T$ periods, and one treatment $T_{gt}$
- Potential outcome without treatment of individual $i$ in group $g$ at period $t$:

$$Y_{0it} = \gamma_g + \lambda_t + u_{it}$$

Hence, $Y$ is determined by the sum of a time-invariant group effect $\gamma_g$ and a year effect $\lambda_t$ that is common across groups

- Potential outcome with treatment:

$$Y_{1gt} = Y_{0gt} + \tau$$

where $\tau$ is the treatment effect (assumed constant)

- Estimation in a $2 \times 2$ case: 2 groups ($T$ and $C$), 2 periods ($t_0$ and $t_1$):

$$Outcome_{it} = \beta_0 + \beta_1 \cdot post_t + \beta_2 \cdot treat_i + \tau \cdot (post * treat)_{it} + \varepsilon$$

- *post* post treatment dummy and captures time effects, *treat* a dummy if the observation is in the treatment group and captures constant differences between the two groups

# Fixed-Effect Specification

- The DiD estimation does not require panel data
- Repeated cross sections, meaning repeated sample from the same population (average data for each group from a different sample at each time period) is sufficient as long as the composition of the population is stable over time
- When panel data are available, however, we can specify fixed-effects models

# Fixed-Effect Specification

Fixed effects can be seen as a generalization of DiD in the case of more than two periods (say S periods) and more than 2 groups (say G groups)

- Suppose that group $g$ in year $t$ experiences a given policy $T$ (for example an income tax rate) of intensity $T_{gt}$. We want to know the effect of $T$ on an outcome $Y$
- OLS in the form of $Y_{gt} = \alpha + \beta T_{gt} + \varepsilon_{gt}$ With no fixed-effects gives a biased estimate of $\beta$ if treatment $T_{gt}$ is correlated with $\varepsilon_{gt}$
- A way to solve this problem is to put time dummies and group dummies in the regression (i.e. within group time variation)

$$Y_{gt} = \alpha + \gamma_t + \delta_g + \beta T_{gt} + \varepsilon_{gt}$$

- The advantage is that we can take into account unobservable characteristics and individual heterogeneity

## A Brief Example

Suppose you are interested in the question whether union workers earn higher wages

- Problem: unionized workers may be different (e.g. higher skilled, more experienced) from non-unionized workers
- Many of these factors will not be observable to the econometrician (standard omitted variable bias problem)
- Therefore the error term and union status will be correlated and OLS will be biased
- We are interested whether $Y_{it}$ (earnings) is affected by $D_{it}$ (union status) which we assume to be randomly assigned
- We also have time varying covariates $X_i$ (such as experience) and unobserved but fixed confounders $A_i$ (e.g. ability)

### A Brief Example

Suppose you are interested in the question whether union workers earn higher wages
Assuming that the causal effect of union membership is additive and constant we also have:

$$E(Y_{0it}|A_i, X_{it}, t) = \alpha + \lambda_t + A_i\gamma + X_{it}\beta$$
$$E(Y_{1it}|A_i, X_{it}, t) = E(Y_{0it}|A_i, X_{it}, t) + \rho$$
$$E(Y_{1it}|A_, X_{it}, t) = \alpha + \lambda_t + \rho D_{it} + A_i\gamma + X_{it}\beta$$

This equation implies the following regression equation:

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + X_{it}\beta + \varepsilon_{it}$$

where $\alpha_i = \alpha + A_i\gamma$

### A Brief Example

Suppose you are interested in the question whether union workers earn higher wages If you simply estimate this model with OLS (without including individual fixed effects):

$$Y_{it} = Constant + \lambda_t + \rho D_{it} + X_{it}\beta + \underbrace{\alpha_i + \varepsilon_{it}}_{u_{it}}$$

As $\alpha_i$ is correlated with union status $D_{it}$ there is a correlation of $D_{it}$ with the error term. This will lead to biased OLS estimates

A fixed effect model would address this problem

- Demeaning: $Y_{it} - \bar{Y}_i = (\lambda_t - \bar{\lambda}) + \rho(D_{it} - \bar{D}_i) + (X_{it} - \bar{X}_i)\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$
- First differencing: $\Delta Y_{it} = \Delta\lambda_t + \rho\Delta D_{it} + \Delta X_{it}\beta + \Delta\varepsilon_{it}$

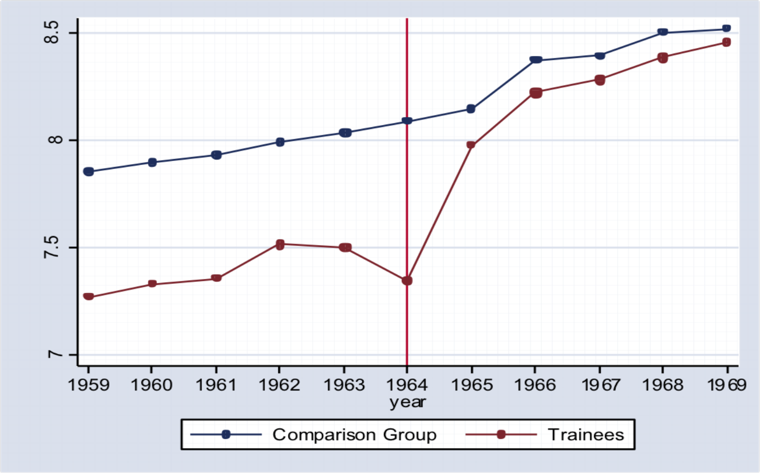
### "Ashenfelter's Dip": Targeting based on differences

- A pre-condition of the validity of the DiD assumption is that the program is not implemented based on the pre-existing differences in outcomes
- The DiD procedure does not control for unobserved temporary individual-specific shocks that influence the participation decision

To illustrate the conditions such inconsistency might arise

- Suppose a training programme is being evaluated in which enrollment is more likely if a temporary dip in earnings occurs just before the programme takes place - the so-called Ashenfelter's dip
- A faster earnings growth is expected among the treated, even without programme participation
- Thus, the DiD estimator is likely to over-estimate the impact of treatment

## "Ashenfelter's Dip": Targeting based on differences



## "Ashenfelter's Dip": Targeting based on differences

- Pre-program "dip" for participants
- if your treatment is selected by participants then only the worst off individuals elect the treatment - not comparable to general effect of policy
- Treatment group are those who received training in 1964
- Control group are random sample of population as a whole
- Simple D-in-D approach would compare earnings in 1965 with 1963
- But earnings of trainees in 1963 seem to show a "dip"; probably because those who enter training are those who had a bad shock (e.g. job loss)
- D-in-D assumption probably not valid

## Differential macro trends

- The identification of ATT using DiD relies on the assumption the treatment and controls experience the same macro shocks
- If this is not the case, the DiD approach will yield a biased and inconsistent estimate of ATT.
- E.g., differential trends might arise in the evaluation of training programs if treated and controls operate in different labour markets

# Common Problems with DiD estimates

**Long-term response versus reliability trade-off:**

- DiD estimates are more reliable when you compare outcomes just before and just after the policy change
- The identifying assumption (parallel trends) is more likely to hold over a short time-window
- With a long time window, many other things are likely to happen and confound the policy change effect
- However, for policy purposes, it is often more interesting to know the medium or long term effect of a policy change

### Inference

- The observations in the control and the treatment group may tend to move together over time.
- there may be a common random effect at the time*group level
- In this case, the standard error of the estimator should take into account this correlation
- If we have enough clusters, we can estimate a cluster covariance matrix

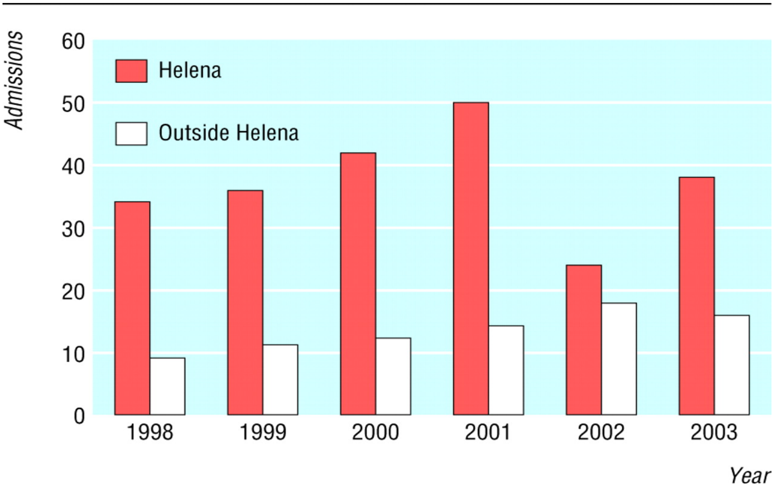
# Dif-in-Dif Papers

Waldinger (2010): the effect of faculty quality on the outcomes of PhD students

- Estimating the effect of faculty quality on PhD student outcomes is challenging because of selection bias, OVB, and measurement error
- address these issues by using the dismissal of scientists in Nazi Germany as an exogenous shock to faculty quality
- The dismissal affected some departments very strongly, while other departments were not affected
- used a panel dataset of all mathematics PhD students graduating from all German universities between 1923 and 1938 and use the dismissal as exogenous variation in faculty quality
- The empirical strategy essentially compares changes in outcomes of PhD students in affected department before and after 1933 to changes in outcomes in unaffected departments

Helena Smoking Experiment

- Question: What is the effect of second hand smoke on the incidence of heart disease?
- There is some medical experiments that show that this may be the case, but the real effects are ambiguous
- Policies under consideration: Banning smoking in bars and restaurants
- Helena, Montana, USA, is a geographically isolated community that imposed such a law from 5 June 2002. Opponents won a court order suspending enforcement of the law on 3 December 2002.
- Examine the association of the policy with admissions for "myocardial infarction" from within Helena (intervention) and from outside Helena, where the policy did not apply (control)


Admissions
60
50
40
30
20
10
0
Helena
Outside Helena
1998
1999
2000
2001
2002
2003
Year

# Dif-in-Dif Papers

Donahue and Levitt "The Impact of Legalized Abortion on Crime" (QJE, 2001)

- This was a paper that got a huge amount of attention in the press at the time
- They show (or claim to show) that there was a large effect of abortion on crime rates
- The story is that the children who were not born as a result of the legalization were more likely to become criminals
- This could be either because of the types of families they were likely to be born to, or because there was differential timing of birth
- Identification comes because 5 states legalized abortion prior to Roe v. Wade (around 1970): New York, Alaska, Hawaii, Washington, and California
- In 1973 the supreme court legalized abortion with Roe v. Wade
- They match the timing of abortion with the age that kids are likely to commence their criminal behavior

# Dif-in-Dif Papers

Jin and Leslie (2003): the role of information in consumers decision making:

- exploit a "natural experiment" that occurred in Los Angeles in 1997
- The city government introduced rules forcing restaurants to post the results of their hygiene report, including a very visible colored grade, on the front of their restaurant door
- The introduction of report cards was staggered in different cities within Los Angeles County. This allows the authors to use a "difference-in-differences" strategy to identify the effect of report cards on both hygiene scores and food related disease
- Positive effect on revenue
- Reduced incidence of food poisoning

# Empirical Methods in Economics

Liyousew G. Borga

February 17, 2016

## Instrumental Variables

# The Endogeneity Problem

What if we want to estimate the effect of an inherently endogenous variable? For example "What is the effect of education on wages?"

- RCT? No, assignment to "treatment" is not random
- Matching? No, selection happens on unobservables
- Regression? No, not enough control variables, hence omitted variable bias
- Dif-in-Dif? No, the "common trends" assumption is not credible

**The ability bias problem**

We want to estimate

$$Y = \alpha + \beta X + \varepsilon$$

- but what does $\beta$, the coefficient for $X$ (education), then tell us?
- An individual with higher ability has a more positive error term since ability is not included in the regression and it affects $Y$ (wages) positively
- The individual also has higher education due to higher ability

# The Endogeneity Problem: Illustration

Possible sources of endogeneity are:

- Omitted variables: Some determinants of $Y$ (e.g. ability) are unobserved, and thus remain in the error term. Endogeneity arises if the omitted variable(s) is correlated with $X$ (i.e. $E(\varepsilon|X) \neq 0$)

Possible sources of endogeneity are:

- Omitted variables: Some determinants of $Y$ (e.g. ability) are unobserved, and thus remain in the error term. Endogeneity arises if the omitted variable(s) is correlated with $X$ (i.e. $E(\varepsilon|X) \neq 0$)
- Measurement error: The true variable is $X^*$, but we only observe $X = X^* + \eta$; the true model is $Y = \alpha + \beta X^* + \varepsilon$, but we estimate $Y = \alpha + \beta X + \nu$ where $(\nu = \varepsilon - \beta \eta)$

# The Endogeneity Problem: Illustration

Possible sources of endogeneity are:

- Omitted variables: Some determinants of $Y$ (e.g. ability) are unobserved, and thus remain in the error term. Endogeneity arises if the omitted variable(s) is correlated with $X$ (i.e. $E(\varepsilon|X) \neq 0$)
- Measurement error: The true variable is $X^*$, but we only observe $X = X^* + \eta$; the true model is $Y = \alpha + \beta X^* + \varepsilon$, but we estimate $Y = \alpha + \beta X + \nu$ where $(\nu = \varepsilon - \beta\eta)$
- Simultaneity: $X$ is jointly determined with $Y$ in the same economic model. The regression equation forms part of a system of simultaneous equations

# The IV solution

To overcome the endogeneity problem we can use the Instrumental Variables (IV) approach

**The ability bias problem again**

$$Y_i = \alpha + \rho X_i + \gamma A_i + \nu_i$$

where $Y_i$ is log of earnings, $X_i$ schooling measured in years, and $A_i$ individual ability

- We need at least one variable, $Z$, that is correlated with education, but uncorrelated with the wage received other than through education
- we then estimate a regression for education with all $X$ variables and $Z$ as explanatory variables and we get predicted education
- This predicted value is then put in the original equation instead of actual education and gives us a non biased estimate
- Problem solved?

**The Intuition: How does IV work?**

Variation in $X$ can be decomposed into

- **Endogenous variation**: Determined within the model and hence correlated with the error term (e.g. *educ* and *ability*)
- **Exogenous variation**: Determined outside the model and hence uncorrelated with the error term (e.g. an exogenous shock)

IV estimates $\beta$ by using exogenous variation in $X$ (the part that comes through $Z$)

## Conditions for an IV

A valid instrument, $Z$, needs to satisfy three conditions:

1. Z is as good as randomly assigned
2. Z satisfies the exclusion restriction, i.e. it does not appear as a separate regressor in the original regression we like to run
3. Z is relevant, i.e. affects the endogenous regressor

Of these, only condition 3 can be tested. Conditions 1 and 2 have to be argued based on knowledge from outside the data we have

# Conditions for an IV

A valid instrument, $Z$, needs to satisfy three conditions:

1. Z is as good as randomly assigned
2. Z satisfies the exclusion restriction, i.e. it does not appear as a separate regressor in the original regression we like to run
3. Z is relevant, i.e. affects the endogenous regressor

Of these, only condition 3 can be tested. Conditions 1 and 2 have to be argued based on knowledge from outside the data we have

Formally, in a general setting:

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- $Z$ is relevant if $Cov(z,x) \neq 0$
- $Z$ is exogenous if $Cov(z,\varepsilon) = 0$

# IV: Three Causal Effects

In IV procedure, there are three causal effects we can think about:

1. The causal effect of $Z_i$ on $educ_i$
2. The causal effect of $Z_i$ on $Y_i$
3. The causal effect of $educ_i$ on $Y_i$

The last one is the one we are ultimately interested in (the return to schooling, $\beta_1$)

Consider the ability bias example (the long regression)

$$Y_i = \alpha + \beta_1 educ_i + \beta_2 ability_i + \varepsilon_i$$

The three causal effects are formalized by three equations:

- **Structural equation**: The regression of earnings on schooling (causal effect 3)

$$Y_i = \alpha + \beta_1 educ_i + \eta_i$$

  where $\eta_i = \beta_2 ability_i + \varepsilon_i$
- **First Stage**: The regression of schooling on the instrument (causal effect 1)

$$educ_i = \pi_{10} + \pi_{11} Z_i + \mu_{1i}$$

- **Reduced form**: The regression of earnings on the instrument (causal effect 2)

$$Y_i = \pi_{20} + \pi_{21} Z_i + \mu_{2i}$$

# IV: Indirect Least Squares

The coefficients in the three equations are linked. Substitute the first stage into the structural equation:

$$\begin{aligned} Y_i &= \alpha + \beta_1 educ_i + \eta_i \\ &= \alpha + \beta_1 [\pi_{10} + \pi_{11} Z_i + \mu_{1i}] + \eta_i \\ &= (\alpha + \beta_1 \pi_{10}) + \beta_1 \pi_{11} Z_i + (\mu_{1i} + \eta_i) \\ &= \pi_{20} + \pi_{21} Z_i + \mu_{2i} \end{aligned}$$

# IV: Indirect Least Squares

The coefficients in the three equations are linked. Substitute the first stage into the structural equation:

$$\begin{aligned} Y_i &= \alpha + \beta_1 educ_i + \eta_i \\ &= \alpha + \beta_1 [\pi_{10} + \pi_{11} Z_i + \mu_{1i}] + \eta_i \\ &= (\alpha + \beta_1 \pi_{10}) + \beta_1 \pi_{11} Z_i + (\mu_{1i} + \eta_i) \\ &= \pi_{20} + \pi_{21} Z_i + \mu_{2i} \end{aligned}$$

Hence, the reduced form coefficients are:

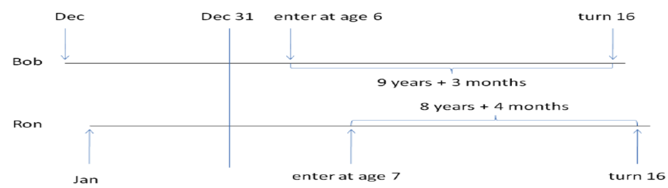
$$\begin{aligned} \pi_{20} &= \alpha + \beta_1 \pi_{10} \\ \pi_{21} &= \beta_1 \pi_{11} \\ \beta_1 &= \frac{\pi_{21}}{\pi_{11}} \end{aligned}$$

i.e. the IV estimate is equal to the ratio of the reduced form coefficients on the instrument to the first stage coefficients. This is called **indirect least squares**

# How IV Works: An Example

Angrist & Krueger (1991): IV in the return to educations literature

- A very influential study where they used quarter of birth as an instrumental variable for schooling
- Their concern is years of schooling may be endogenous, with pre-schooling levels of ability affecting both schooling choices and earnings given education levels
- They exploit variation in schooling levels that arise from differential impacts of compulsory schooling laws
- School districts typically require a student to have turned six by January 1st of the year the student enters school
- Since students are required to stay in school till they turn sixteen, those born in the first quarter have lower required minimum schooling levels than the ones born in the last quarter



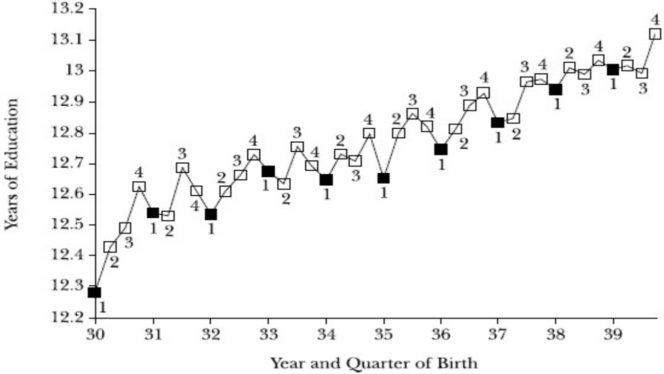
# How IV Works: An Example

Any substantive arguments why quarter of birth need not be a valid instrument?

- Random assignment: Are birthdays random with respect to the counterfactual earnings for different schooling levels?
- Do birthdays satisfy the exclusion restriction, or could birthdays be correlated with earnings for other reasons than their effect on schooling?
- Do birthdays indeed affect schooling?

## First stage:

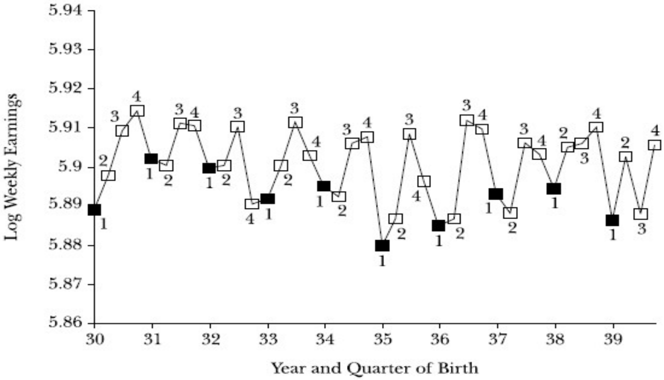
The regression of the causal variable of interest on covariates and the instrument(s)




- The education-quarter-of-birth pattern for men in the 1980 Census who were born in the 1930s
- Men born earlier in the calendar year had lower average education

### Reduced form

The reduced form relationship between the instruments and the dependent variable (i.e. the regression of the dependent variable on any covariates in the model and the instrument)




- Those born early in the calendar year had lower earnings

# Several Instruments: 2SLS

- Indirect least squares only works when there is one endogenous regressor and one instrument (**just identified** model)
- If there are multiple instruments for a single endogenous regressor the model is **over-identified**
- Since there is no unique way to solve the model for $\beta_1$, we use an alternative method called Two-Stage-Least-Squares (2SLS)
  - Find instruments $Z_1$ and $Z_2$ for the endogenous variable $X$, in the equation of interest

$$Y = \alpha + \beta X + \varepsilon$$

  - First stage: Regress $X$ on all instruments

$$X = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + v$$

  - Second stage: Use the predicted values $\hat{X}$ from the 1st stage as explanatory variables in the outcome equation

$$Y = \alpha + \beta \hat{X} + \varepsilon$$

# How (Where) do you Find IVs?

**Be creative!**

The relationship between fertility and labor supply:

- Angrist & Evans (1998) wanted to estimate the effect of children on parents' labor supply
- Endogeneity concern: mothers with weak labor force attachment or low earnings potential may be more likely to have children
- They conjecture that in the US, most parents prefer to have both daughters and sons
- Implication: Those with two daughters or two sons are more likely to have a third child than those with one boy and one girl
- At the same time, having two same-sex children should be uncorrelated with other determinants of labor supply
- Having two same-sex children looks like a promising instrument for a third child!

# How (Where) do you Find IVs?

The relationship between fertility and labor supply:

- Basic regression:

$$Y = \alpha + \beta More + \varepsilon$$

  $Y$ is labor market outcome (income, hours worked); *More* is an indicator for whether the person has more than two children
- Instrument for *More*: Dummy that switches on if the first two children are of the same gender
- Additional controls: Mother's age, age at 1st birth, girl/girl or boy/boy, race
- Variation: Use "2nd birth was a twin" as an instrument for more than 2 children
- 1st stage: Parents of same-sex siblings are 6.7 percentage points more likely to have a third birth
- 2nd stage: Those with a third child are 13.5% less likely to be employed, work 6 weeks less per year, and 5.5 hours less per week

# How (Where) do you Find IVs?

Are the Instruments Exogenous?

- Intuition: The gender of the first two children is plausibly random
- Thus the instrument is plausibly uncorrelated with other observable and unobservable characteristics of households
- Same-sex children are less costly (re-use toys and clothing), this could have labor-market effects
- Probably a different implication in a developing country setup?

# How (Where) do you Find IVs?

## Use Policy Change

Angrist (1990) Veteran Draft Lottery: the effects of military service on earnings

- Angrist (1990) uses the Vietnam draft lottery as in IV for military service
- In the 1960s and early 1970s, young American men were drafted for military service to serve in Vietnam
- Concerns about the fairness of the conscription policy lead to the introduction of a draft lottery in 1970
- From 1970 to 1972 random sequence numbers were assigned to each birth date in cohorts of 19-year-olds
- Men with lottery numbers below a cutoff were drafted while men with numbers above the cutoff could not be drafted
- The draft did not perfectly determinate military service. Many draft-eligible men were exempted for health and other reasons. Exempted men volunteered for service

# How (Where) do you Find IVs?

- The instrument is thus defined as follows:

  $Z_i = 1$ if lottery implied individual $i$ would be draft eligible
  $Z_i = 0$ 1 if lottery implied individual $i$ would not be draft eligible

- First stage results: Having a low lottery number (being eligible for the draft) increases veteran status by about 16 percentage points (the mean of veteran status is about 27 percent)
- Second stage results: Serving in the army lowers earnings by between $2,050 and $2,741 per year

# How (Where) do you Find IVs?

## Know some history

The dismissal of German Jew scientists & allied bombing as a natural experiment

- Relative importance of human and physical capital in the scientific knowledge production function
- human capital shock (dismissal of Jews scientists by the Nazis in 1933)
- Physical capita shock (bombing by allies during the war)
- Exogeneous Bombings: Universities were never listed as targets in any of the Allied bombing directives and similar documents
- Nonetheless, many universities facilities were destroyed by bombings which could never be precisely aimed until the end of WWII
- Because of these targeting problems bombs often fell relatively randomly within cities and there was thus large variation in destruction across different university buildings

# How (Where) do you Find IVs?

Research question: what is the role of physical capital in the production of scientific knowledge?

- Waldinger (2012) uses the % of destruction caused by Allied WWII bombings as an exogeneous shock to physical capital of German universities
- Are you convinced? What potential problems could there be?
- Human capital shocks are persistent due to: Peer effects; Permanent drop in department size; Effect of lower quality faculty on PhD student outcomes (German universities like to hire their own graduates); Difficult to attract high quality people once the average quality has dropped

# How (Where) do you Find IVs?

## Learn a bit of Medicine

- Thomas et al (2006) use an iron supplementation program to estimate the returns to health for rubber tree tappers in Indonesia
- Workers randomly received 120mg of iron every week for a year; controls received a placebo
- Their two main questions are:
  - Does iron supplementation improve workers' health? Yes, as measured by anemia levels
  - Did the program increase their income? Yes: Self-employed males earned on average $40 more (if iron-deficient: $200 more) with iron supplementation, at a program cost of $6 per year
- The program should have no direct effect on income, other than through health (i.e. to the extent that iron supplementation affects income, it must be operating through improved health)
- We can use this to test whether health is important for income

## Bad instruments

- When the instruments are not valid (remember, this cannot be tested)
- Overidentification test (if there are more instruments, use only one of the instruments in 1st stage and use the other as controls in the 2nd stage. Then test if the “control” instruments are equal to zero, which they should be)
- these tests only help prove that an instrument is bad (e.g. all instruments may be bad, then the test is useless)
- The best tool to evaluate exogeneity is common sense

## Weak instruments

- An instrument is weak if the correlation with the endogenous variable is low
- Relevance is easy to test: regress $x$ on $z$ and check whether the coefficients are significant (i.e. first stage)
- Rule of thumb: 1st stage F-stat at least 10

## Ugly instruments

- It is not really the instruments that are ugly, but rather the interpretation of the instruments
- It may be the case that the causal effect is not the same for all individuals and the instrument works differently for different groups
- What are we really measuring?

**Quarter of birth and earnings: Angrist and Krueger (1991)**

- Think of the variation in education predicted by quarter of birth. To whom is it relevant?
- Those who would like to drop out but are prevented from dropping out by compulsory schooling laws
- Another way to say this: QOB is irrelevant to anyone who doesn't want to drop out before the minimum age in their particular state
- This means that the variation we are using in education is not the variation we might be interested in if we are trying to predict the relationship between higher education and earnings

### Vietnam Draft: Angrist (1990)

- In the draft lottery example: IV estimates the average effect of military service on earnings for the subpopulation who enrolled in military service because of the draft but would not have served otherwise
- This excludes volunteers and men who were exempted from military service for medical reasons for example
- The instrument affects treatment, which is this application amounts to entering the military service. The researcher observes treatment status as follows:

  $D_i = 1$ if individual $i$ served in the Vietnam war (veteran)
  $D_i = 0$ if individual $i$ did not serve in the Vietnam war (not veteran)
- Now define potential outcomes for $D_i$

  $D_{0i} = 0$ if individual $i$ would not serve in the military if not draft eligible
  $D_{0i} = 1$ if individual $i$ would serve in the military even though not draft eligible
  $D_{1i} = 0$ if individual $i$ would not serve in the military even though draft eligible
  $D_{1i} = 1$ if individual $i$ would serve in the military if draft eligible

# Local Average Treatment Effects (LATE)

LATE: the average is not for all treated, but for an instrument specific subpopulation

## Compliance types

The LATE framework partitions any population with an instrument into potentially 4 groups. For a binary treatment $T$, binary instrument $Z$ case, we have:

- Compliers: do what they have been told to
- Never takers: do not want the treatment
- Always takers: do want the treatment
- Defiers: do the opposite of what they have been told

Never takers and always takers don't contribute to the IV estimate; defiers and compliers contribute, the IV estimate is the sum of those two effects

# Local Average Treatment Effects (LATE)

- The LATE theorem says that (under the assumptions) IV estimates the average causal effect of treatment on the subpopulation of compliers.
  - **Independence assumption:** The instrument is as good as randomly assigned
  - **Exclusion restriction:** The potential outcomes are only affected by the instrument through the treatment
  - **Relevance (First stage):** The average causal effect of the instrument on treatment is not zero
  - **Monotonicity:** No defiers
- If all 4 assumptions are satisfied, IV estimates LATE
- LATE is the average effect of X on Y for those whose treatment status has been changed by the instrument Z

**The LATE Cup is Half Full (Empty)**

# Empirical Methods in Economics

Liyousew G. Borga

February 17, 2016

## Regression Discontinuity Design

# RD Designs: Making Use of Arbitrary Rules

Regression discontinuity designs (RDD) exploit natural experiments generated by arbitrary rules

- Students receive a scholarship if their GPA is above 3.0
- Legislators are elected if they receive over 50% of the vote
- Children are allowed to start school if they turn 6 by 31 December that year
- Welfare relief is only given to those with less than 40 dollars per month

# RD Designs: Making Use of Arbitrary Rules

So, the idea is to estimate the treatment effect using individuals just below the threshold as a control for those just above. RDDs exploit the fact that:

- treatment assignment is based on the value of a continuous variable (the selection variable) $X$
- the participation rate is discontinuous at least at one known value of that selection variable
- On either side of the common threshold, individuals have very close characteristics, but some are treated and some are not

# RDD: Sharp and Fuzzy

The key to the RD design is that we have a deep understanding of the mechanism which underlies the assignment of treatment $D_i$. In this case, assignment to treatment depends on a single variable $X_i$.

There are two designs of an RD method, the sharp and the fuzzy design:

- Sharp Design: the assignment to treatment is determined through a known deterministic decision rule
  - selection on observables
- Fuzzy Design: Discrete jump in the *probability* of being treated, but not deterministic.
  - instrumental variable setup

# Sharp RD

Assignment to treatment depends on a single deterministic and discontinuous function of a variable $X$

$$D_i = \begin{cases} 1 & \text{if } X_i \geq X_0 \\ 0 & \text{if } X_i < X_0 \end{cases}$$

- $X_i$ is called the running variable
- $X_0$ is a known as the threshold or cutoff
- This assignment mechanism is a deterministic function of $X_i$; because once we know $X_i$ we know $D_i$.
- It's a discontinuous function because no matter how close $X_i$ gets to $X_0$, treatment is unchanged until $X_i = X_0$

A major advantage of the RD design over competing methods is its transparency, which can be illustrated using graphical methods.

# Sharp RD: An Example

Thistlethwaite & Campbell (1960):

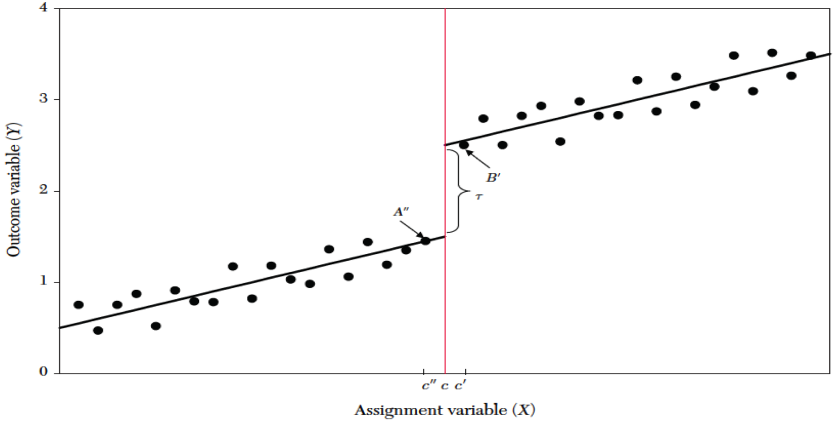
- American high school students are awarded the National Merit Scholarship on the basis of PSAT scores
  - Test taken by most college-bound high school juniors
  - Scholarships are awarded based on a PSAT score threshold
- This was the first use of RD, and motivated by the question: Are students who win these awards more likely to finish college?
- Consider OLS:

$$finish\_college_i = \alpha + \beta \cdot scholarship_i + \varepsilon_i$$

  Clearly, those who qualify for scholarships are more likely to graduate even without the scholarship (ability, hard work); hence, OLS is inconsistent.
- Instead, let the receipt of treatment be denoted by the dummy variable $D \in \{0,1\}$, so that we have $D = 1$ if $X \geq c$ and $D = 0$ if $X < c$

$$Y = \alpha + D\tau + X\beta + \varepsilon$$

# Sharp RD: An Example



# Sharp RD: An Example

Consider an individual whose score $X$ is exactly $c$:

- To get the causal effect for a person scoring $c$, we need guesses for what her $Y$ would be with and without receiving the treatment
- If it is "reasonable" to assume that all factors (other than the award) are evolving "smoothly" with respect to X, then $B'$ would be a reasonable guess for the value of $Y$ of an individual scoring $c$ (and hence receiving the treatment).
- Similarly, $A''$ would be a reasonable guess for that same individual in the counterfactual state of not having received the treatment.
- It follows that $B' - A''$ would be the causal estimate.
- This illustrates the intuition that the RD estimates should use observations "close" to the cutoff (e.g., in this case at points $c'$ and $c''$).

# Sharp RD: An Example

- There is, however, a limitation to the intuition that "the closer to $c$ you examine, the better"
- In practice, one cannot "only" use data close to the cutoff. The narrower the area that is examined, the less data there are.
- In our example, examining data any closer than $c'$ and $c''$ will yield no observations at all!
- Thus, in order to produce a reasonable guess for the treated and untreated states at $X = c$ with finite data, one has no choice but to use data away from the discontinuity
- Indeed, if the underlying function is truly linear, we know that the best linear unbiased estimator of $\tau$ is the coefficient on $D$ from OLS estimation

# Sharp RD: An Example

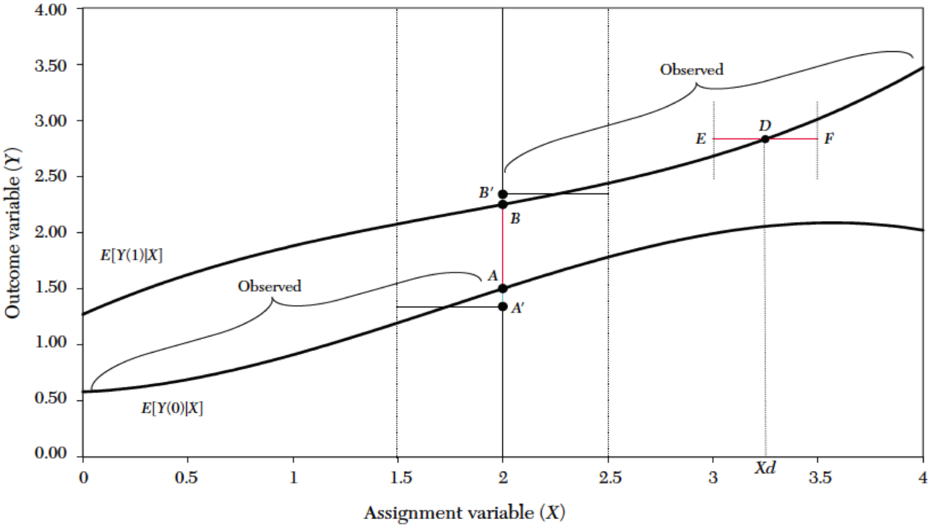
This simple heuristic presentation illustrates two important features of the RD design:

- First, in order for this approach to work, "all other factors" determining $Y$ must be evolving "smoothly" with respect to $X$.
  - If the other variables also jump at $c$, then the gap $\tau$ will potentially be biased for the treatment effect of interest
- Second, since an RD estimate requires data away from the cutoff, the estimate will be dependent on the chosen functional form
  - the method relies on extrapolation across covariate values

# RD Designs and the Potential Outcomes Framework

It is typically imagined that, for each individual $i$, there exists a pair of "potential" outcomes:

- $Y_i(1)$ for what would occur if the unit were exposed to the treatment and $Y_i(0)$ if not exposed
- The causal effect of the treatment is represented by the difference $Y_i(1) - Y_i(0)$
- The fundamental problem of causal inference is that we cannot observe the pair $Y_i(1)$ and $Y_i(0)$ simultaneously
- We therefore typically focus on average effects of the treatment, that is, averages of $Y_i(1) - Y_i(0)$ over (sub-)populations, rather than on unit-level effects
- In the RD setting, we can imagine there are two underlying relationships between average outcomes and $X$,



4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
Outcome variable (Y)
0
0.5
1
1.5
2
2.5
3
Xd
3.5
4
Assignment variable (X)
Observed
Observed
E[Y(1)|X]
E[Y(0)|X]
A
A′
B
B′
D
E
F

# RD Designs and the Potential Outcomes Framework

- By definition of the RD design, all individuals to the right of the cutoff ($c = 2$ in this example) are exposed to treatment and all those to the left are denied treatment
- Therefore, we only observe $E[Y_i(1)|X]$ to the right of the cutoff and $E[Y_i(0)|X]$ to the left of the cutoff
- With what is observable, we could try to estimate the quantity:

$$B - A = \lim_{\varepsilon \downarrow 0} E[Y_i|X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i|X_i = c + \varepsilon]$$

  which would equal

$$E[Y_i(1) - Y_i(0)|X = c]$$

- This is the "average treatment effect" at the cutoff $c$
- This inference is possible because of the continuity of the underlying functions $E[Y_i(0)|X]$ and $E[Y_i(1)|X]$
- This continuity condition enables us to use the average outcome of those right below the cutoff (who are denied the treatment) as a valid counterfactual for those right above the cutoff (who received the treatment)

# Sharp RD: Estimation Procedure

We have now defined a causal effect as the difference of two functions at a point. How do we estimate that? There are 3 general approaches:

1. Compare means
2. OLS (with Polynomials)
3. Local Linear Regression

### First approach: Compare means

- In the data, we never observe observe $E[Y(0)|X = c]$, that is there are no units at the cutoff that don't get the treatment, but in principle it can be approximated arbitrarily well by $E[Y(0)|X = c - \varepsilon]$
- Therefore we estimate:
$$E[Y|X = c + \varepsilon] - E[Y|X = c - \varepsilon]$$
- This is the difference in means for those just above and below the cutoff
- This is a nonparametric approach. A great virtue is that it does not depend on correct specification of functional forms
- In practice, however, this depends on having lots of data within $\varepsilon$ of the cutoff

### Second approach: OLS (with Polynomials)

- What if the trend relation, $E[Y_i(0)|X]$, is nonlinear?
- Suppose the nonlinear relationship is $E[Y_i(0)|X] = f(X)$ for some reasonably smooth function $f(X)$
- In that case we can construct RD estimates by fitting:

$$Y = \alpha + \tau T + f(X) + \eta$$

  where $f(X)$ is a smooth nonlinear function of $X$
- Perhaps the simplest way to approximate $f(X)$ is via OLS with polynomials in $X$
- Common practice is to fit different polynomial functions on each side of the cutoff by including interactions between $T$ and $X$
- Modeling $f(X)$ with a pth-order polynomial in this way leads to

$$Y = \alpha + \beta_{01}X + \beta_{02}X^2 + \cdots + \beta_{0p}X^p + \tau T + \beta_{11}TX + \beta_{12}TX^2 + \cdots + \beta_{1p}TX^p + \eta$$

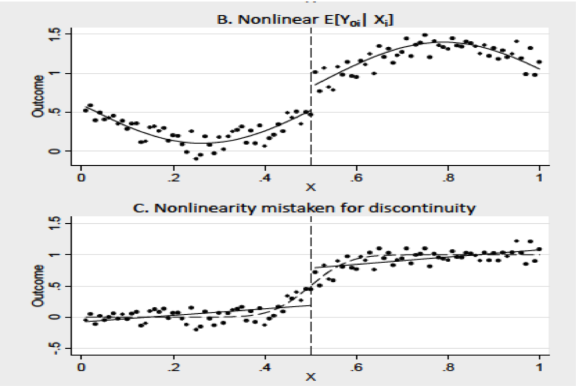
- The polynomial method suffers from the problem that you are using data that is far away from the cutoff to estimate the $f(X)$ function

### Third approach: Local Linear Regression

- Instead of locally fitting a constant function (e.g., the mean), fit linear regressions to observations within some bandwidth of the cutoff
- Local linear regressions provide a nonparametric way of consistently estimating the treatment effect in an RD design
- In non-parametric estimation we restrict the sample to individuals close to the cutoff (i.e. Strictly local estimation at the threshold (left & right limits))
- It is less sensitive to functional form assumptions; but we need enough observations around the threshold

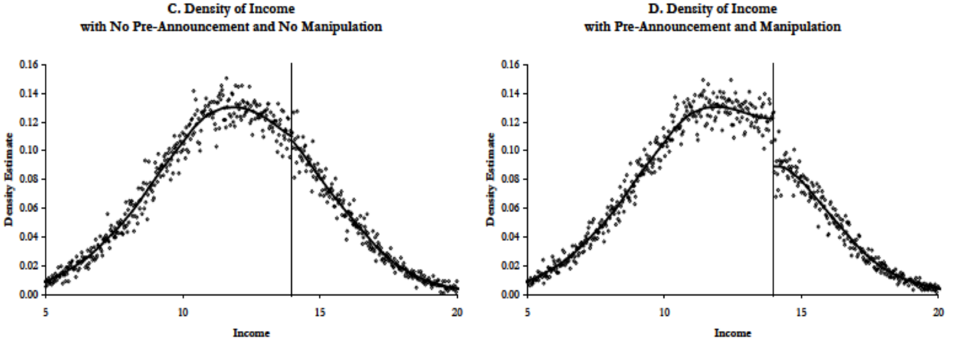
# RD Pitfall: Mistaking Nonlinearity for Discontinuity

- Consequences of using an incorrect functional form are potentially more severe for RD than for other methods we studied so far
- This is because there is no value of $X_i$ at which we get to observe both treated and control observations
- We need to extrapolate how $Y$ behaves as $X$ approaches our threshold!



# RD Pitfall: Manipulation

- If individuals have control over the assignment variable, then we should expect them to sort into (out of) treatment if treatment is desirable (undesirable)
- If individuals have precise control over the assignment variable, we would expect the density of X to be zero just below the threshold but positive just above the threshold (assuming the treatment is desirable)
- For example consider an income support program in which those earning under $14,000 qualify for support (McCrary 2008):



# RDD: Checklist

## Assumptions of RD

- The treatment is determined at least in part by the assignment variable
- There is a discontinuity in the level of treatment at some cutoff value of the assignment variable (selection on observables at the cutpoint)
- Units cannot precisely manipulate the assignment variable to influence whether they receive the treatment or not
- Other variables that affect the treatment do not change discontinuously at the cutoff

## Internal & External Validity

- The strength of the RD design is its internal validity, arguably the strongest of any quasi-experimental design
- External validity may be limited: Sharp RD provides estimates for the subpopulation with $X = c$, i.e. those right at the cutoff of the assignment variable
- You need to justify extrapolation to other subpopulations

# RDD: Checklist

### Threats to an RD Analysis

- Other variables change discontinuously at the cutoff
- There are discontinuities at other values of the assignment variable
- Manipulation of the assignment variable

### Steps for Sharp RD Analysis

- Graph the data by computing the average value of the outcome variable over a set of bins
- Estimate the treatment effect by running linear regressions on both sides of the cutoff point
- The robustness of the results should be assessed by employing various specification tests

## Evaluating an RD Paper

Does the author show convincingly that:

- Treatment changes discontinuously at the cutpoint?
- Outcomes change discontinuously at the cutpoint?
- Other covariates do not change discontinuously at the cutpoint?
- Pre-treatment outcomes do not change at the cutpoint?
- There is no manipulation of the assignment variable (bunching near the cutpoint)?
- The results are robust to different functional form assumptions about the assignment variable?

## Graphical Analysis in RD Designs

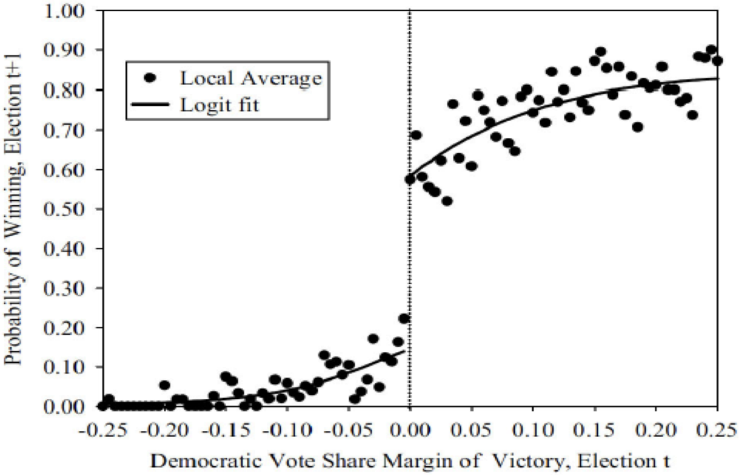
- Outcome by forcing variable (Xi ): The standard graph showing the discontinuity in the outcome variable
- Construct a similar graph, but using a covariate as the "outcome"; There should be no jump in other covariates
- Plot the density of the forcing variable to investigate whether there is a discontinuity in the distribution of the forcing variable at the threshold

# Sharp RD: An Example

Sharp Design: Lee (2008)

- Effect of party incumbency on re-election probability in the U.S. House of Representatives
- Does a candidate for parliament have better chances of being elected if his party won the seat the last time? Do representatives use the privileges and resources of their office to gain an advantage for themselves or their party?
- Lee uses a sample of 6,558 elections over the 1946-98 period
- The assignment variable in this setting is the fraction of votes awarded to Democrats in the previous election
  - When the fraction exceeds 50 percent, a Democrat is elected and the party becomes the incumbent party in the next election
  - Clearly, vote share in the last election could simply capture who is more well-liked, better at representing the people, etc
- We'd expect a relationship between vote share in $t$ and re-election in $t+1$ even if everything is legit
- But those with 49% and 51% in $t$ should have similar chances in $t+1$. Right?
- Both the share of votes and the probability of winning the next election are considered as outcome variables

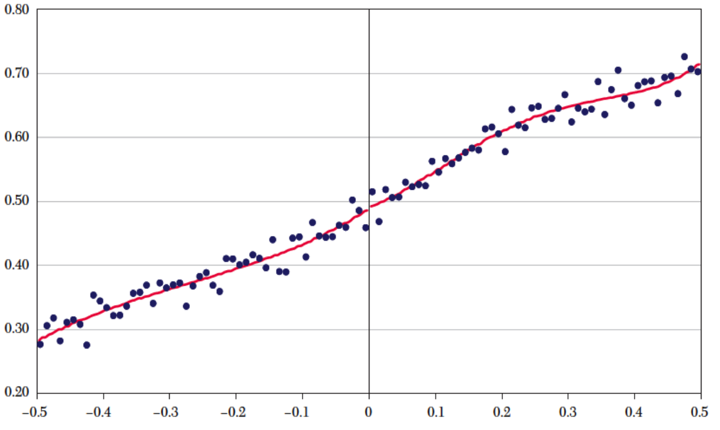
Lee analyzes the probability of winning the election in year $t+1$ by comparing candidates who just won compared to candidates who just lost the election in year $t$




$X$ is the margin of victory ($X > 0$ if won, $X < 0$ if lost)

# Sharp RD: An Example

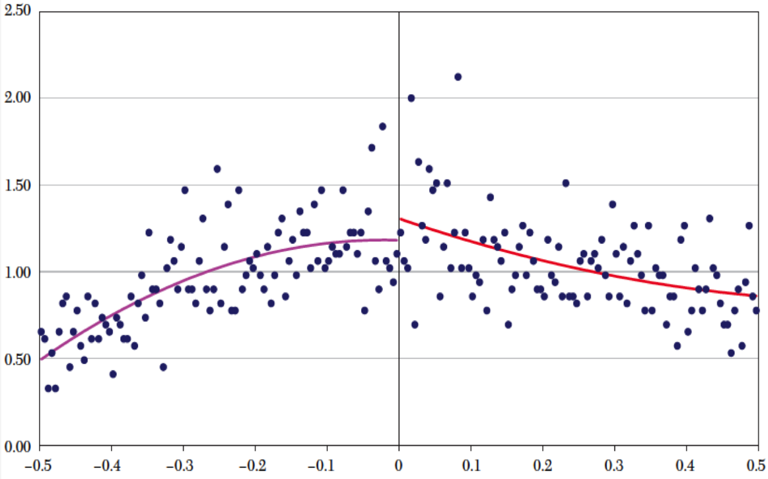
An alternative approach for testing the validity of the RD design is to examine whether the observed baseline covariates are "locally" balanced on either side of the threshold




RD graph for a baseline covariate, the Democratic vote share in the election prior to the one used for the assignment variable (four years prior to the current election)

# Sharp RD: An Example

Inspection of the Histogram of the Assignment Variable




This is an indirect test of the identifying assumption that each individual has imprecise control over the assignment variable (i.e., no manipulation/sorting)

# Interpreting RD

RD relies on regression, yet RD identification is distinct

- In regression (matching) we hope that treatment is as good as randomly assigned after conditioning on controls
- There will be units with the same values of the controls (matches) but with different treatment status
- In RD, there is no value of $X_i$ at which you observe both treatment and control observations
- But for observations very close to the discontinuity we effectively have an experiment
- RD estimates are local to the cutoff
- For this reason we cannot be agnostic about regression functional form in RD

# Fuzzy RD

Fuzzy RD exploits discontinuities in the probability of treatment conditional on a covariate

- Sometimes the threshold rule is not deterministic for treatment assignment, but there may only be a change in the probability of treatment at the cutoff
- The discontinuity becomes an instrumental variable for treatment status
- $D_i$ is no longer deterministically related to crossing a threshold but there is a jump in the probability of treatment at $X_0$

$$Pr(D_i = 1) = p(X_i)$$

$$\lim_{X_i \uparrow X_0} p(X_i) \neq \lim_{X_i \downarrow X_0} p(X_i)$$

- The probability of treatment $p(X_i)$ is a continuous function, except at $X_0$
- You may think of Fuzzy RD as IV: call "compliers" those who are affected by the treatment in the Fuzzy RD

### RDD: Example

Abdulkadiroglu, Angrist, and Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools", (Econometrica, 2014)

- A three bedroom house on the northern edge of Newton, Massachusetts costs $412,000 (in 2008 dollars); while across the street, in Waltham, a similar place cost only $316,000
- 92% of Newton's high school students are graded proficient in math, while only 78% are proficient in Waltham
- This suggests, "something" changes at school district boundaries
- Parents looking for a home are surely aware of achievement differences between Newton and Waltham, and many are willing to pay a premium to see their children attend what appear to be better schools

## RDD: Example

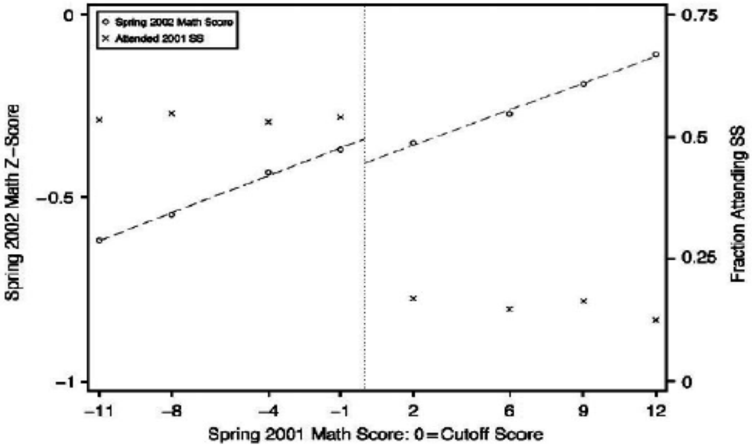
- Question: Do schools affect student performance?
- Problem: Selection into schools is nonrandom
- Solution: Use a regression discontinuity design around elite school admissions criteria
  - Within a small interval around the threshold for admittance, test scores are as good as random
  - So compare two similar groups - one who got treatment (admitted to elite school) and another who didn't
- Data: from Boston and New York public schools (Standardized test scores for elementary and high school, College Board test files; demographic info)
- The outcome variable: High School/College Board test scores
- the running variable: the criteria for admittance to the exam schools (basically the schools' rankings of students)
- Identification: Fuzzy RD; look at students in area around school admission cutoff
- Results: No measurable impact of elite schools on college admissions or test scores

# Fuzzy RDD

An Example: Matsudaira (2008)

- What is the effect of attending a mandatory summer school on test scores in the next year?
- Rule: Students (3rd grade or higher) who fail one or more end-of-year exams in 2001 are required to take remedial summer classes
- Imperfect compliance: Besides passing the test-score in reading and math, there are other criteria for promotion like attendance which is not recorded in these data. This makes the assignment to summer school fuzzy around the test cutoff
- The forcing variable $X_i$ is the minimum of the 2001 reading score and the 2001 mathematics score, minus the threshold for passing
- The outcome variable $Y_i$ is the standardized mathematics score in 2002

Maths score 2001 positively predicts maths score 2002, with a discrete negative jump at the passing threshold




Those who failed 2001 are more likely to attend summer school

## RDD: Example

Sandra Black, "Do Better Schools Matter? Parental Valuation of Elementary Education", (QJE, 1999)

- Idea: Can parents "buy" better schools for their children by living in a neighborhood with better public schools? How do we measure the willingness to pay?
- Problem: Just looking in a cross section, richer parents probably live in nicer houses in areas that are better for many reasons
- Solution: Black uses the school border as a regression discontinuity
  - take two families who live on opposite side of the same street, but are zoned to go to different schools
  - The difference in their house price gives the willingness to pay for school quality
- Data: All home sales from 1993-1995 in 3 counties in suburban Boston, MA (single-family residences only); Elementary school boundaries because they are small; Census block data for demographic controls
- Identification: Sharp RD; look at houses on two sides of street that creates boundary between schools
- Results: People pay for better schools; 5% increase in test scores results in 2.1% increase in home price ($3,948)