因果推論在社會科學與政策評估的應用

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

中研院統計所
統計科學營

Causal Inference

Causal Inference

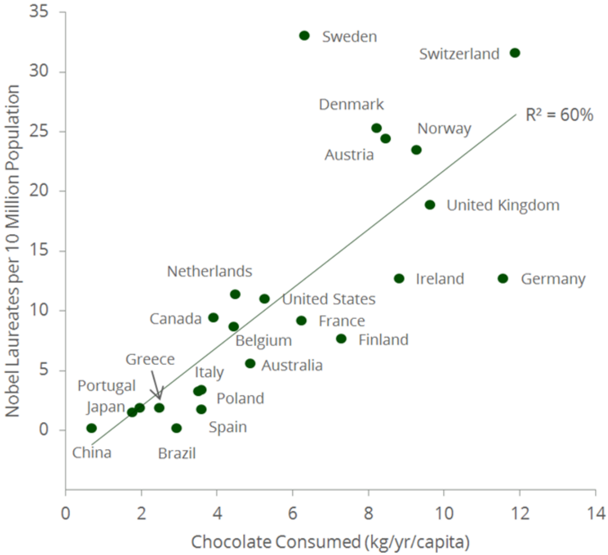
- Social science (Economics) theories are almost always causal in their nature
  - X causes Y
  - A decrease in price of oil causes consumer's demand for oil to increase
  - Raising minimum wage would reduce employment opportunity of low-skilled workers
  - An increase in interest rate can reduce housing price

Causal Inference

- Two key features of causality:
  1. Causes are asymmetrical
     - In general, if X causes Y, Y does not cause X
  2. Causes are effective
     - A cause must be distinguished from an accidental correlation

Correlation is not Causality

Chocolate Consumption and Nobel Laureates



Correlation is not Causality

- In order to increase number of Nobel Laureates (proxy for human capital)
- Should government enforce everyone to eat chocolate everyday?

Correlation is not Causality

$$\begin{tikzcd} & U \arrow[dl] \arrow[dr] & \\ X & & Y \end{tikzcd}$$

- X (Chocolate Consumption) is associated (correlated) with Y (Number of Nobel Laureates)
  - Even if X has no causal effect on Y
  - Since confounding factor U (GDP) can result in the co-movement between X and Y

Causal Inference

- Understanding a causal relationship is useful for making predictions about the consequences of changing circumstances or policies
- Causal inference is a type of statistical methods that help us verify the causal relationship
- In general, a typical causal question is:
  - The effect of a treatment on an outcome
  - Outcome: A variable that we are interested in
  - Treatment: A variable that has the (causal) effect on our outcome of interest

Causal Inference

Example 1

- The effect of getting a master's degree on earnings
  - Ideally, we should get causal effect by comparing the earnings of **the same individuals** with and without receiving a master's degree
  - For each particular individual, we can observe **only one outcome with specific treatment at the same time**:
    - Getting a master's degree
    - Not getting a master's degree
  - The **unobserved outcome** is called the "**counterfactual**" outcome

Causal Inference

Example 1

- The effect of getting a master's degree on earnings
  - What if we compare **observed outcomes**:
    - Earnings of those getting a master's degree
    - Earnings of those choosing not to get it
  - Simply comparing those who are and are not treated may provide a misleading estimate of a causal effect
  - There must be a reason why some people choose to have and some choose to not have a master's degree
    - For example, those who get a master's degree may be from rich families or have high ability
    - Two groups of people might not be comparable
  - We need to isolate casual effect from the effect of other confounding factors

Causal Inference

Example 2

- Macro economists also ask casual questions !
- The effect of quantitative easing (QE) on economic growth
  - Does QE accelerate economic growth?
  - Ideally, we should get causal effect by comparing the GDP growth rate of **the same countries (areas)** with and without adopting QE policy
  - Again, we have an unobserved outcome problem

Causal Inference

Example 2

- The effect of quantitative easing (QE) on economic growth
  - Countries adopting QE v.s. Countries not adopting QE:
    - Two groups are not comparable
  - Why some countries need to implement QE policy?
    - Because they have bad economic performance $\Rightarrow$ underestimate the positive effect of QE

Causal Inference

More Examples

- More examples include:
  - The effect of advertisement on product sales
  - The effect of military service on earnings and employment
  - The effect of unemployment insurance on job search behavior
  - The effect of credit regulation on housing price
  - Does eliminating estate tax increase wealth inequality?
  - Do immigrant workers depress the wages of native workers?
  - Can democracy increase economic growth?
  - The effect of COVID-19 (virus) on world economy

Causal Inference

- The fundamental problem of inferring the causal effect is that:
  - For every unit (e.g. individual, household, state, or country), we fail to observe the outcome if the chosen level of the treatment had been different
- Basically, causal inference is the study of **unobservable counterfactuals**:
  - It tells us what happend in alternative (or “counterfactual”) world
  - What would happened if we were to change this aspect of the world ?

Causal Inference
Unobservable Counterfactuals

Causal Inference

- Since it is impossible to observe the **unobserved** counterfactual outcome
- Causal inferences help us infer the values of these **unobserved counterfactual outcomes** from **observed data** by imposing specific assumptions
- Under specific assumptions, we can obtain the causal effect of treatment

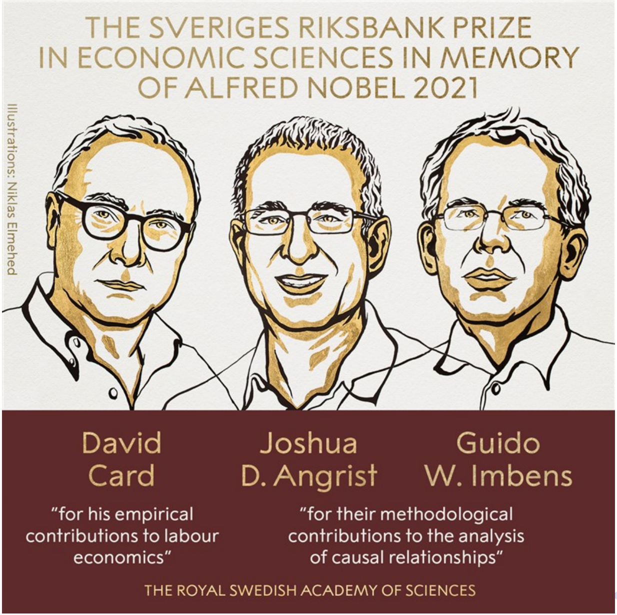
This Course

- In this talk, we will go through three causal inference methods:

    1. Regression discontinuity design
    2. Difference-in-differences design
    3. Synthetic control method (if time permits)

2021 Nobel Laureates

Economics and Causal Inference



Regression Discontinuity Design: Main Idea

Introduction

Selection Bias and RCT

- A major problem of estimating causal effect of treatment is the threat of **selection bias**
- In many situations, individuals can **select into treatment** so those who get treatment could be very different from those who are untreated
- The best to deal with this problem is conducting a **randomized controlled trial (RCT)**

Main Idea of Regression Discontinuity Design

- In an RCT, researchers can eliminate selection bias by **controlling treatment assignment process**
  - An RCT randomizes who receives a treatment –the treatment group - and who does not –the control group
  - Since we randomly assign treatment, the probability of getting treatment is unrelated to other confounding factors
- But conducting an RCT is very expensive and may have ethical issue

Main Idea of Regression Discontinuity Design

- Instead of controlling treatment assignment process, if researchers have **detailed institutional knowledge of treatment assignment process**
- Then we could use this information to create an "experiment"

Main Idea of Regression Discontinuity Design

- Regression Discontinuity Design (RDD) exploits the facts that:
  - Some rules can generate a discontinuity in treatment assignment
    - The treatment assignment is determined based on whether a unit exceeds some threshold on a variable.
    - Such variable is called **assignment variable**
  - Assume other factors do NOT change abruptly at threshold
  - Then any change in outcome of interest can be attributed to the assigned treatment

Main Idea of Regression Discontinuity Design

A Motivating Example

- A large number of studies have shown that graduates from more selective programs or schools earn more than others
  - In Taiwan, many students want to enter elite schools
  - Students graduated from NTU earn more than those graduated from other schools

Main Idea of Regression Discontinuity Design

A Motivating Example

- But it is difficult to know whether the positive earnings premium is due to
  - true “causal” impact of human capital acquired in the academic program
  - a spurious correlation linked to the fact that good students selected in these programs would have earned more no matter what
- The latter point reflects **selection bias**
- We need to untangle the **causal effect** and **selection bias**

Main Idea of Regression Discontinuity Design

A Motivating Example

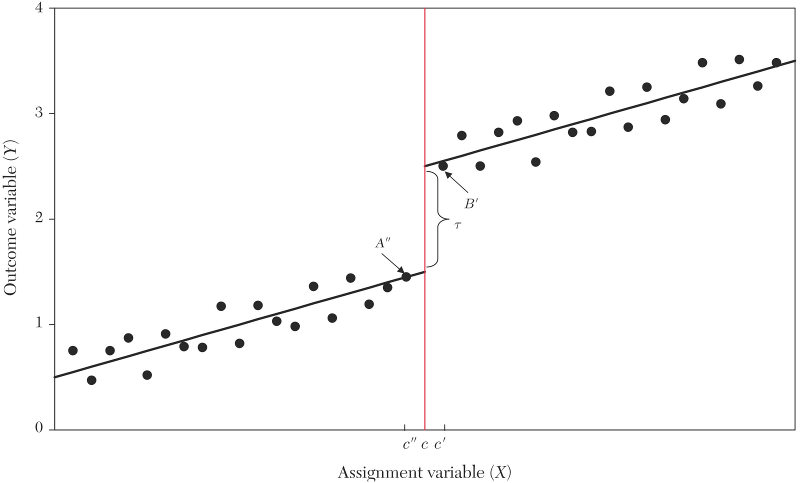
- A great way to answer that question would be to run an experiment:
  - Take students applying both to NTU and NTHU
  - Instead of admitting them the regular way, just flip a coin to decide whether they get into NTU or NTHU
  - Follow them up 10 years later to see whether those admitted to NTU earn more than those admitted to NTHU
- Great idea, but nobody will let me run that experiment...

Main Idea of Regression Discontinuity Design

A Motivating Example

- But say that the entry cutoff for a score of entrance exam is 400 at NTU
- They would perhaps let me flip a coin for those with scores of 399 or 400
- Since the those get 399 and those get 400 are essentially identical
- They get different scores due to some random events
- **RD strategy**: I can do "as well" as in a randomized experiment by tracking down the long term outcomes for the 400 (admitted to NTU) and the 399 (admitted at NTHU)

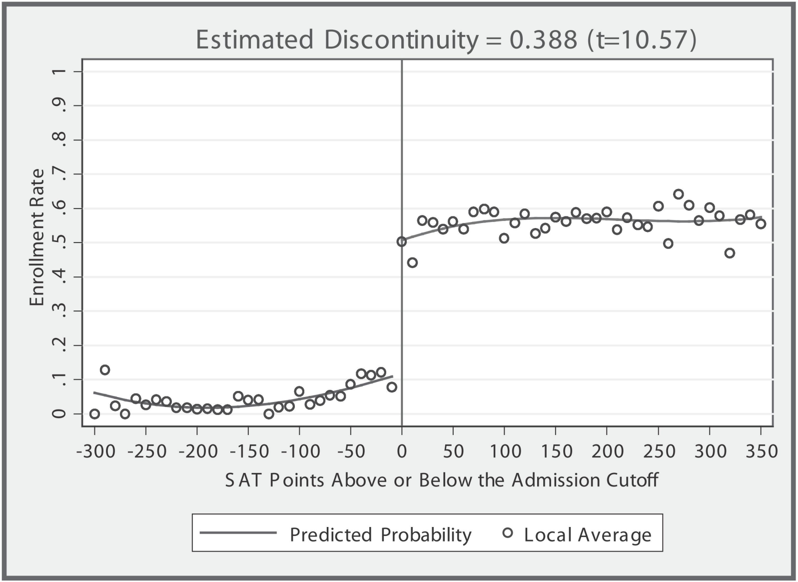
Test Score and Earnings



Source: Lee and Lemieux (2010)

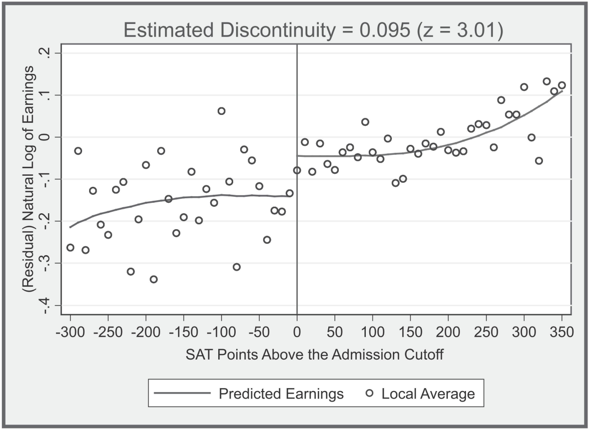
SAT Score and Enrollment

FIGURE 1.—FRACTION ENROLLED AT THE FLAGSHIP STATE UNIVERSITY



SAT Score and Earnings

Figure 2.—Natural Log of Annual Earnings for White Men Ten to Fifteen Years after High School Graduation (Fit with a Cubic Polynomial of Adjusted SAT Score)



Regression Discontinuity Design: Potential Outcomes Framework

RDD and Potential Outcomes

Treatment

- Assignment variable: $X_i \in \mathbb{R}$
- Threshold (cutoff) for treatment assignment: $c \in \mathbb{R}$
- D_i: a dummy that indicate whether individual i receive treatment or not
- Treatment assignment:

$$D_i = \{X_i \geq c\}$$

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

RDD and Potential Outcomes

Potential Outcomes

- Y_i^1: Potential outcome for an individual i if he would receive treatment
- Y_i^0: Potential outcome for an individual i if he would not receive treatment

Observed Outcomes

- Observed outcomes Y_i are realized as:

$$\mathrm{Y}_i = \mathrm{Y}_i^1 D_i + \mathrm{Y}_i^0 (1 - D_i)$$

$$\mathrm{Y}_i = \begin{cases} \mathrm{Y}_i^1 & \text{if } D_i = 1 \ (X_i \geq c) \\ \mathrm{Y}_i^0 & \text{if } D_i = 0 \ (X_i < c) \end{cases}$$

Identification Results for RDD

- Ideally, for each individual *i*, if we could observe two potential outcomes at the same time, we can estimate **average treatment effect (ATE)**:

$$\alpha_{\mathsf{ATE}} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0]$$

- But it is **impossible** to observe two potential outcomes at the same time

Identification Results for RDD

- Instead, we can use RDD to investigate the behavior of the outcome around the threshold:

$$\alpha_{\mathsf{RD}} = \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | X_i = c + \varepsilon] - \lim_{\varepsilon \to 0} \mathrm{E}[\mathrm{Y}_i | X_i = c - \varepsilon]$$

- Under certain assumptions, this quantity identifies the **ATE at the threshold**:

$$\alpha_{\mathsf{ATE\ at\ c}} = \mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0 | X_i = c]$$

Regression Discontinuity Design: An Empirical Example

Regression Discontinuity Design

Example

Chen, Wei-Lin, Ming-Jen Lin, and Tzu-Ting Yang. "**Curriculum and National Identity: Evidence from the 1997 Curriculum Reform in Taiwan**." *Journal of Development Economics* 163 (2023)

Motivation

- Can school curriculum affect the formation of national identity or political behaviors?
- Governments around the world are incentivized to use the education system as an instrument for cultivating national identity
  - The more homogeneous the people, the easier it is to manage a nation
  - Especially, when the countries face military threats
- This issue arises many debates in Taiwan

Challenge

- Causal evidence of curriculum effect is still very rare
  - Reverse causality: Government could change the content of textbook based on social trend

Main Idea

- In 1997, Taiwanese government implement a new curriculum (認識台灣) for the students who attend junior high school after September 1997
- That is, those who were born after September 1984 had to read new textbook, which focused on Taiwanese history, geography, and society.
- Those who were born before September 1984 would read old textbook, which exclusively focused on China
- Use regression discontinuity design
  - Compare the national identity of those born right before and those born after September 1984

Comparison of the Textbooks

- Old Textbook focused on history of mainland China
  - Students have to learned the history of China during their first two year
  - Only 16 pages on Taiwan
    - Describe how to develop Taiwan as a base for recovering China

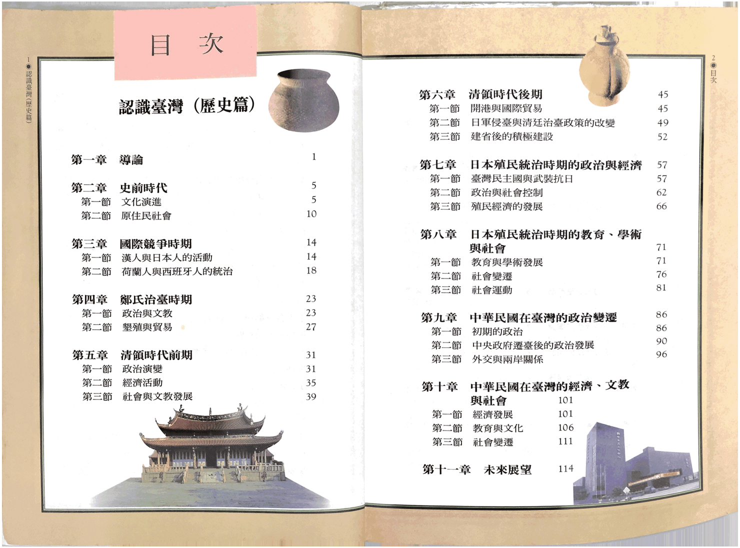
Old Textbook: 16 Pages about Taiwan

One chapter and a section

Comparison of the Textbooks

- New Textbook (*Knowing Taiwan* series) content focuses on history of Taiwan
  - Students have to learned the history of Taiwan during their first year
  - About 116 pages on Taiwan
    - It has eleven chapters and each chapter described how ancestors of different ethnic groups made developments in Taiwan

New Textbook: 116 pages about Taiwan



1 ● 認識臺灣（歷史篇）

目　次

認識臺灣（歷史篇）

2 ● 目次

Comparison of the Textbooks

- Term usage also changed
- Examples:
  - Old textbook: 'our country' for both China and Taiwan
  - New textbook: 'China' 'Taiwan'

Old Textbook: Our Country=China

大概的活動情形了。

史前時代的人類，先是體質特徵介於人形猿和現代人之間的「猿人」，後有「眞人」的出現。我們從地下發現的史前人類化石之中，知道了很多猿人和眞人的情形。目前已經發現的猿人，比較有名的至少有四種：一是在非洲東岸發現的「東非人」；二是在印尼發現的「爪哇人」；三是在我國發現的「北京人」；四是在德國發現的「尼安德人」。在各種眞人當中，有兩種比較著名：一是在我國發現的「山頂洞人」；另一是在法國發現的「克魯麥囊人」。猿人的形態比起現代人尚有一段距離，而眞人則已具有現代人的形態了。

我國歷史的特色

我國歷史悠久，是東亞文明的主體，在世界歷史上占有重要的地位。歸納起來，其特色有下列四點：

一、時間悠久　我國的歷史，從黃帝建國算

New Textbook: China

用火取暖、照明和燒烤食物。

至於中國境內舊石器時代晚期的人類，則以「山頂洞人」為代表。「山頂洞人」距今約兩萬年，體質已和現代人差不多。他們已知埋葬死者，還會用獸骨作成骨針，用獸齒製成裝飾品，生活比「北京人」進步得多。

由這些舊石器時代人類化石的發現，可知中國是人類的主要起源地之一；但他們和現代中國人有無直接關係，目前仍無法確定。要追究中國文化的源頭，比較可靠的線索是新石器時代的考古發現。

「北京人」頭蓋骨
從復原的「北京人」頭蓋骨化石可知，「北京人」兩眉相連，前額低平，腦容量不如現代人。

Old Textbook: Our Country=Taiwan

破。

重大建設：六十二年十一月，行政院院長蔣經國鄭重宣布：政府除積極興建核能發電廠外，並限期五年內完成南北高速公路、臺中港、北迴鐵路、蘇澳港、石油化學工業、大煉鋼廠、大造船廠、鐵路電氣化及桃園國際機場等重要建設；合稱十大建設，皆陸續完成。六十八年，政府又推動交通、工業、農業等十二項建設。七十三年，又推出十四項重要建設計畫，多為前述十大、十二項建設的延續，具有前瞻性的大工程。

目前，我國由於雄厚的工業基礎與外貿潛力，躋身「亞洲四小龍」之列，經濟方面的成就，已為舉世所公認。

文教建設　普及教育方面：民國三十九學年度，臺灣地區六至十一歲學齡兒童的就學率為百分之八十左右；至七十九學年度業已近於百分之

New Textbook: Taiwan

back

圖10-7
花蓮國際港開放

化方面，積極致力於對外貿易、金融、產業經營的自由化。具體作法分別為解除進口管制、大幅降低關稅稅率、取消銀行利率的管制、大幅放寬外匯管制①、開放民間設立銀行，以及推動公營事業民營化等。

國際化方面，具體作法包括放寬外國公司在臺投資、設立臺灣境外金融中心、致力使新臺幣國際化等。近年又籌設亞太營運中心，期使臺灣成為亞太地區的運輸、金融、資訊的重鎮。

Data and Sample

- Taiwan Social Change Survey
  - Repeated cross-sectional data, representative sample of total population, aged 18 above
  - Sample for main results: 2003–2005 (age 18-23)
  - Sample for long run effect: 2009–2015 (age 23-32)
- Key feature:
  - Ask respondents their birth year and month
- Individuals born close to September 1984
  - Four education cohorts: September 1982 – September 1986

The Identity Question

"In our society, somebody call themselves "Taiwanese," some body call themselves "Chinese," and somebody call themselves "both." Do you consider yourself as "Taiwanese," "Chinese," or "both"?"

$$\text{Taiwanese Identity} = \begin{cases} 1 & \text{if Taiwanese,} \\ 0 & \text{Both or Chinese} \end{cases}$$

- Less than 5% of the sample respond with Chinese
- Stronger vs Weaker Taiwanese identity

Regression Discontinuity Design

$$Identity_i = \alpha_0 + \alpha_1 TextBook_i + f(m;\beta) + \gamma X_i + \eta_j + \delta_t + \epsilon_i$$

- $Identity_i$: a dummy variable indicating Taiwanese identity for an individual i
- $TextBook_i$: a dummy variable indicating an individual i born after September 1984
- $f(m;\beta)$: first-order polynomial of birth cohort m interacting fully with $TextBook$

Regression Discontinuity Design

$$Identity_i = \alpha_0 + \alpha_1 TextBook_i + f(m;\beta) + \gamma X_i + \eta_j + \delta_t + \epsilon_i$$

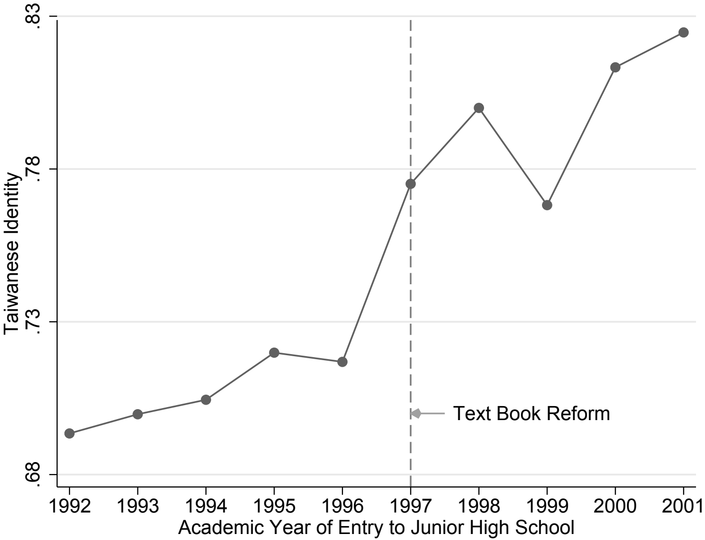
- X_i: gender, parents' edu, parents' ethnicity, Hoklo people ratio (dummy) Sample
- η_j: home county fixed effect
- δ_t: survey year fixed effect
- Clustered s.e: birth cohort (birth year-month)
- Bandwidth: 24 months (2 academic year)

Taiwanese Identity and School Entry Year

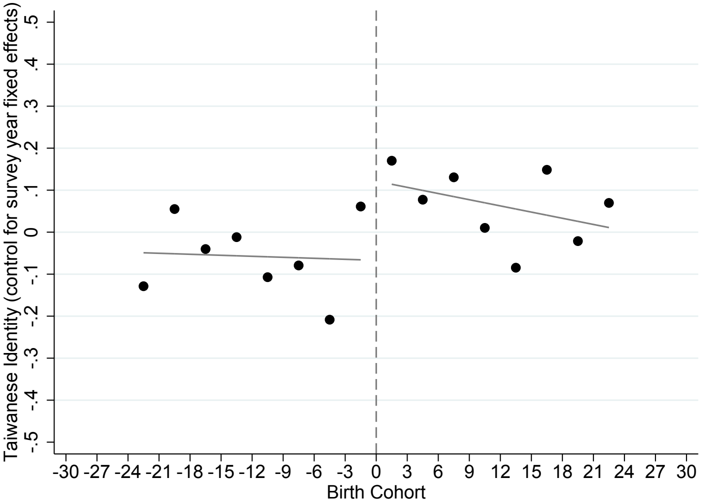
Table 1: Descriptive Statistics for Treatment Group and Control Group

	Born after September 1984	Born before September 1984	Difference (after - before)
Female	0.449	0.448	0.001
	(0.499)	(0.498)	(0.050)
Age	19.598	20.954	-1.355***
	(0.658)	(0.991)	(0.081)
Years of schooling (self)	13.928	14.254	-0.326
	(2.107)	(1.981)	(0.206)
Years of schooling (father)	10.725	10.424	0.301
	(3.256)	(3.368)	(0.341)
Years of schooling (mother)	9.946	9.668	0.278
	(3.412)	(3.262)	(0.335)
Proportion of Hoklo in the hometown	0.709	0.736	-0.027
	(0.230)	(0.207)	(0.022)
Hoklo father	0.784	0.768	0.016
	(0.412)	(0.423)	(0.042)
Hoklo mother	0.820	0.828	-0.008
	(0.385)	(0.378)	(0.0382)
# of individuals	167	250	

Taiwanese Identity and School Entry Year



Taiwanese Identity and Birth Quarter



Summary of Results

Main Results

- Compared to people studying old textbooks, those studying new textbooks hold stronger Taiwanese identity
- The share of reporting themselves as Taiwanese increases by 18 percentage points
- Heterogeneity: Education track
  - **Hard working** (academic track) students are affected, while vocational track students are not
- Heterogeneity: Hometown Ethnic distribution
  - Students lived in areas with **less Taiwanese identity** (less Hoklo people) are affected more

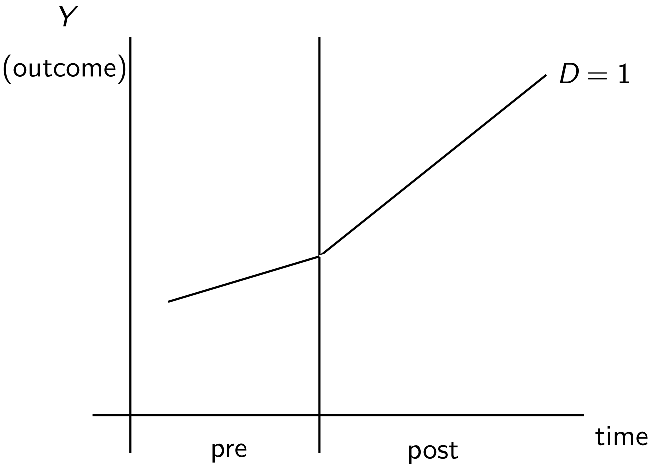
Difference-in-Differences Design: Main Idea

Main Idea of Difference-in-Differences (DID)

- If we can observe **group-level** outcomes several times
  - At least before and after treatment
- Assume **in the absence of treatment**, outcomes of treatment and control group **move in parallel way**
- Then, we can construct the **counterfactual trend in outcomes of treatment group** by using
  - **Trend in outcomes of control group**
- Comparing observed trend with counterfactual trend in outcome of treatment group, we can get causal effect of treatment

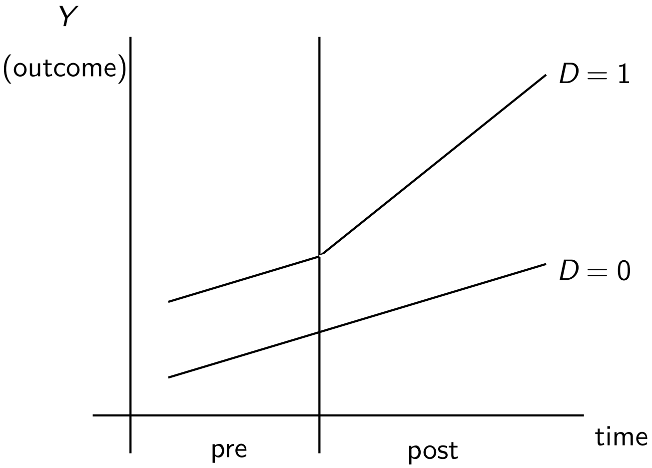
Main Idea of Difference-in-Differences

Graph



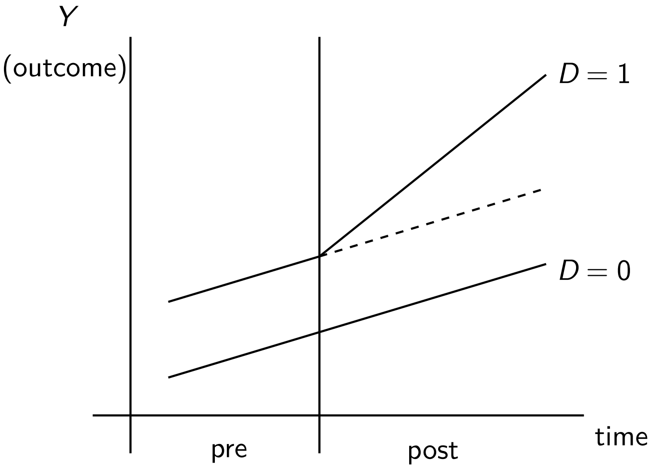
Main Idea of Difference-in-Differences

Graph



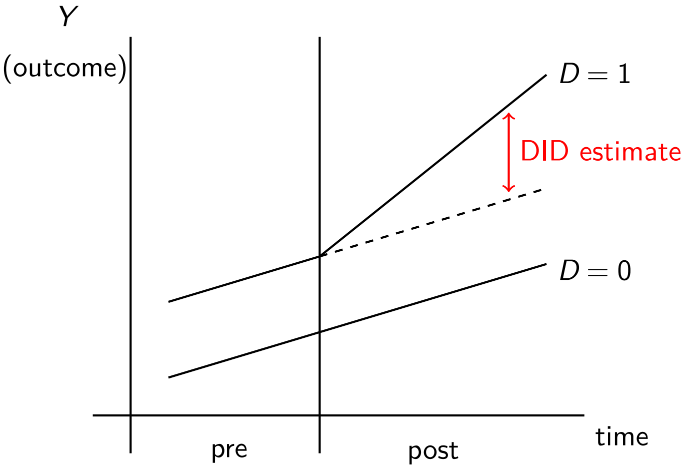
Main Idea of Difference-in-Differences

Graph



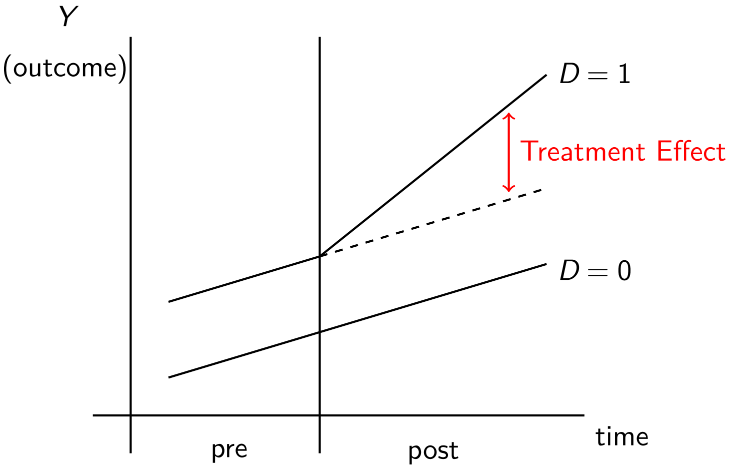
Main Idea of Difference-in-Differences

Graph



Main Idea of Difference-in-Differences

Graph



Difference-in-Differences Design: Potential Outcomes Framework

DID and Potential Outcomes Framework

- Basic setup: two time periods, two groups
- Two periods
  - In period $t = 1$: one of the groups is treated
  - In period $t = 0$: neither group is treated
- Two groups
  - $D_i = 1$: those that are treated at $t = 1$ (treatment group)
  - $D_i = 0$: those that are always untreated (control group)

DID and Potential Outcomes Framework

- Potential Outcomes
  - Y_{it}^{1}: the potential outcome for unit *i* if he would receive treatment at time *t*
  - Y_{it}^{0}: the potential outcome for unit *i* if he would NOT receive treatment at time *t*

DID and Potential Outcomes Framework

- Observed Outcomes
  - Y_{it} is the observed outcome for unit i at time t
    - Observed outcomes at $t = 0$:

$$Y_{i0} = \mathrm{Y}_{i0}^{0}$$

    - Observed outcomes at $t = 1$:

$$Y_{i1} = \mathrm{Y}_{i1}^{0}(1 - D_i) + \mathrm{Y}_{i1}^{1} D_i$$

Identification Results for DID

- Our main interest is average treatment effect on treated (ATT):
$$\alpha_{\mathsf{ATT}} = E[\mathrm{Y}_{i1}^{1} - \mathrm{Y}_{i1}^{0} | D_i = 1]$$

- Missing data problem: $E[\mathrm{Y}_{i1}^{0} | D_i = 1]$ is unknown

- DID design can help us identify ATT if common trend assumption holds

Identification Results for DID

Identification Assumption

Common Trend Assumption

$$E[\mathrm{Y}_{i1}^0 - \mathrm{Y}_{i0}^0 | D_i = 1] = E[\mathrm{Y}_{i1}^0 - \mathrm{Y}_{i0}^0 | D_i = 0]$$
$$= E[\mathrm{Y}_{i1} - \mathrm{Y}_{i0} | D_i = 0]$$

- The treatment group and control group would have exhibited the same trend in the absence of the treatment
- We can use common trend assumption to construct a counterfactual for treatment group at $t = 1$

$$\mathrm{E}[\mathrm{Y}_{i1}^0 | D_i = 1] = \mathrm{E}[\mathrm{Y}_{i0}^0 | D_i = 1] + \mathrm{E}[\mathrm{Y}_{i1}^0 - \mathrm{Y}_{i0}^0 | D_i = 0]$$
$$= \mathrm{E}[\mathrm{Y}_{i0} | D_i = 1] + \mathrm{E}[\mathrm{Y}_{i1} - \mathrm{Y}_{i0} | D_i = 0]$$

- We can use **observed outcomes** to represent **unobserved** $\mathrm{E}[\mathrm{Y}_{i1}^0 | D_i = 1]$

Identification Results for DID

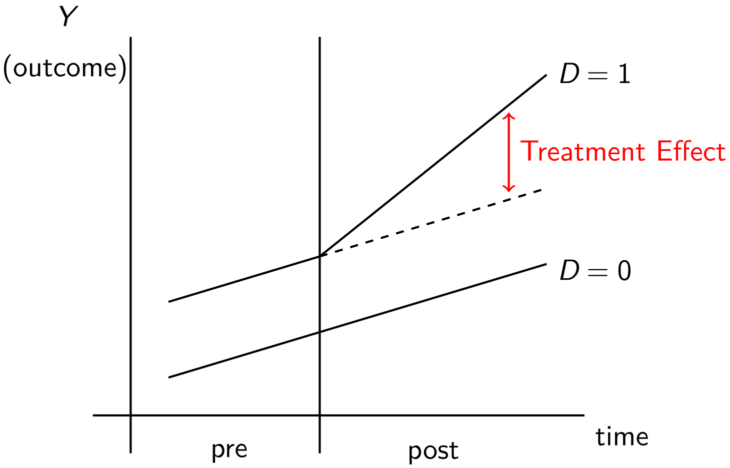
- Apply common trend assumption:

$$\begin{aligned}
\alpha_{\mathsf{ATT}} &= \mathrm{E}[\mathrm{Y}^1_{i1} - \mathrm{Y}^0_{i1} | D_i = 1] \\
&= \mathrm{E}[\mathrm{Y}^1_{i1} | D_i = 1] - \mathrm{E}[\mathrm{Y}^0_{i1} | D_i = 1] \\
&= \mathrm{E}[\mathrm{Y}^1_{i1} | D_i = 1] - \mathrm{E}[\mathrm{Y}^0_{i0} | D_i = 1] - \mathrm{E}[\mathrm{Y}^0_{i1} - \mathrm{Y}^0_{i0} | D_i = 0] \\
&= \mathrm{E}[\mathrm{Y}^1_{i1} - \mathrm{Y}^0_{i0} | D_i = 1] - \mathrm{E}[\mathrm{Y}^0_{i1} - \mathrm{Y}^0_{i0} | D_i = 0] \\
&= \mathrm{E}[\mathrm{Y}_{i1} - \mathrm{Y}_{i0} | D_i = 1] - \mathrm{E}[\mathrm{Y}_{i1} - \mathrm{Y}_{i0} | D_i = 0] = \alpha_{\mathsf{DID}}
\end{aligned}$$

- The **average treatment effect on treated (ATT)** can be identified by difference in trend of outcome between treatment and control groups

Identification Results for DID

Graphical Interpretation



Difference-in-Differences Design: An Empirical Example

Hsing-Wen Han, Kuang-Ta Lo, Yung-Yu Tsai, and Tzu-Ting Yang, "**The Effect of Financial Resources on Fertility: Evidence from Administrative Data on Lottery Winners**", Working Paper

Motivation

- During the past fifty years, fertility rates in developed countries have declined dramatically
- Low fertility rate leads to the growth of an aging population, workforce shortages, and reductions in tax revenue.
- Many countries initiated child-related cash transfer policies to encourage childbearing.
  - On average, the public spending of child-related cash benefits accounts for 1.1% of GDP in OECD countries.
- The rationale behind these policies is that people do not have enough income to afford the expense of raising children, so the government needs to subsidize them.

Motivation

- Empirically, there is an endogenous problem between income and fertility.
  - Reverse Causality
  - Income effect confounds with substitution effect
    - Both working and raising children are time-consuming activities
    - A sudden increase in wage income can increase the relative price of having children
    - Higher wage income would make people work more and reduce demand for children

DID Event-Study Design

- This paper examines the fertility impact of the large and permanent income shock generated by winning lottery prizes.
- We implement an DID event-study design to examine the causal effect of large income shock on fertility.
- Compare the trend in fertility before and after receiving a windfall gain between:
  - Households winning 1,000,000 NT$ from lottery prizes.
  - Households winning less than 10,000 NT$.

DID Event-Study Design

- We estimate the following regression:

$$Y_{it} = \alpha + \beta D_i + \sum_{k=-3}^{6} \delta_k \mathbf{I}[t - E_i = k] + \sum_{k=-3}^{6} \gamma_k D_i \cdot \mathbf{I}[t - E_i = k] + X'_{it}\theta + \varepsilon_{it},$$

  - D_i represents treatment group dummy.
  - Treatment Group:
    - Households who earn more than 1,000,000 NT$ by winning lotteries in a given year
  - Control group:
    - Households who earn less than 10,000 NT$ from winning lotteries during sample period

DID Event-Study Design

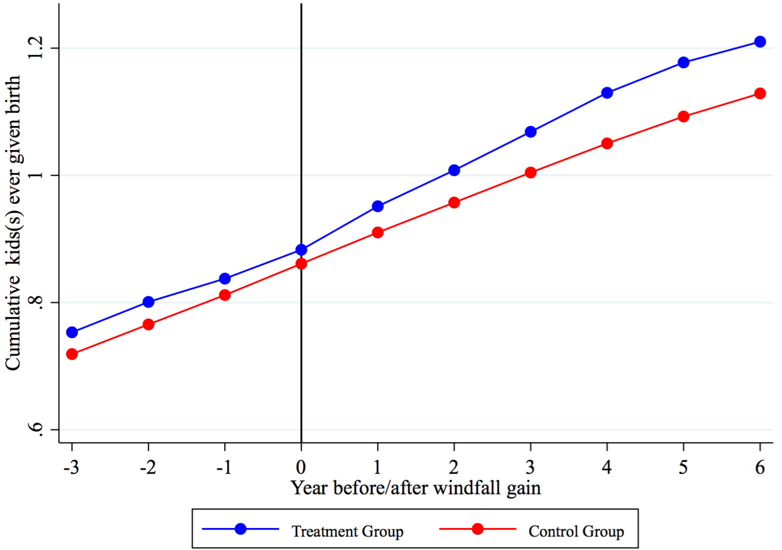
- We estimate the following regression:

$$Y_{it} = \alpha + \beta D_i + \sum_{k=-3}^{6} \delta_k \mathbf{I}[t - E_i = k] + \sum_{k=-3}^{6} \gamma_k D_i \cdot \mathbf{I}[t - E_i = k] + X_{it}'\theta + \varepsilon_{it},$$

  - Outcome variable Y_{it}:
    - Cumulative number of children for household i in the year t
  - E_i is the lottery-winning year
  - $\mathbf{I}[t - E_i = k]$ denotes dummy variables for the year before and after winning lottery.
  - For example, $\mathbf{I}[t - E_i = 1]$ represents a dummy for the first year after winning lottery.
  - Note that we use one year before lottery-winning year as the baseline year (i.e. $k = -1$).

Test Common Trend Assumption

Raw Data: Cumulative Number of Children



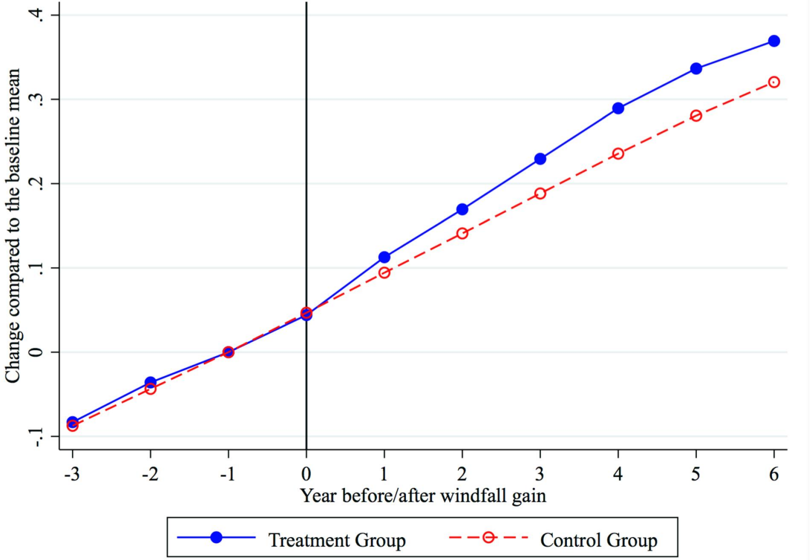
Test Common Trend Assumption

Raw Data: Cumulative Number of Children

- Since we focus on trend rather than level, we sometimes subtract the baseline mean ($k = -1$) from the outcome at each time period

Test Common Trend Assumption

Subtract the Baseline Mean: Cumulative Number of Children



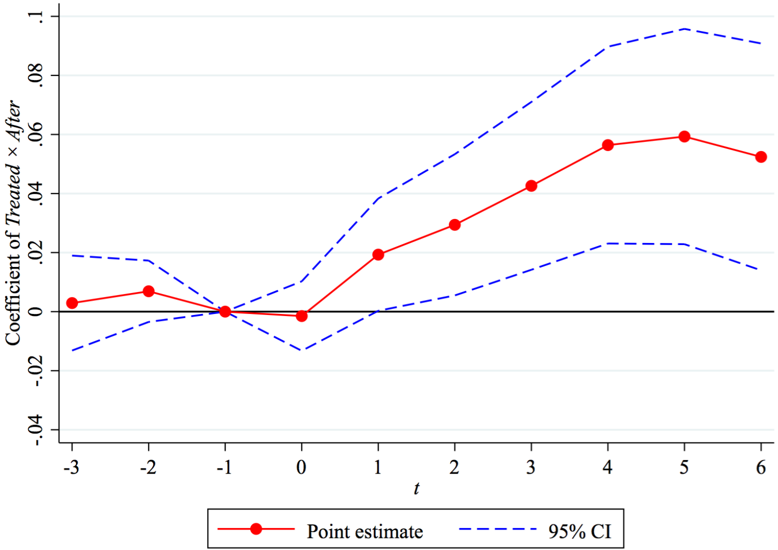
Test Common Trend Assumption

Raw Data: Cumulative Number of Children

- We can formally examine common trend assumption by showing the estimated coefficients $\gamma_{-2}, \gamma_{-3}, ..., \gamma_6$
- If common trend assumption is valid, γ_{-2}, γ_{-3} should be close to zero
- $\gamma_0, \gamma_1, ..., \gamma_6$ represent the treatment effects of winning lotteries

Test Common Trend Assumption

DID Event-Study Design: Cumulative Number of Children



Summary of Results

- Lottery wins of over 1 million NT$ can significantly increase the number of children households have by 0.06
- In other words, for every 100 affected individuals, 6 more children are born within 6 years of the windfall than would have been without the lottery prize
- Large cash windfalls increase fertility primarily by inducing childless households to have their first child
  - Lottery wins have a negligible impact on subsequent births for those who already have children
- We find a lottery win of 10 million NT$ increases marriage rates by 4 percentage points and implemented a causal mediation analysis
  - Around one-third of the overall fertility effect can be attributed to increased marriage rates

Synthetic Control Method: Main Idea

Main Idea of Synthetic Control Method

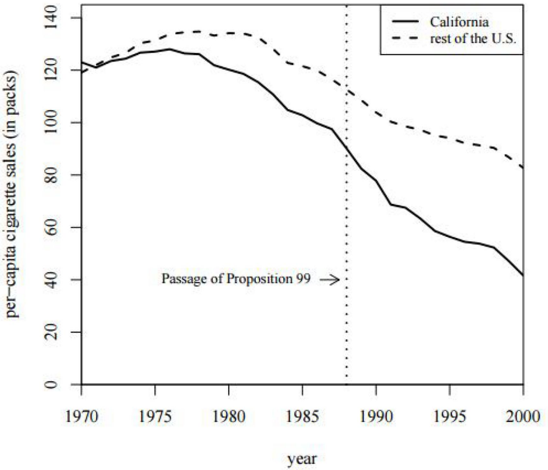
- Synthetic Control (SC) is a method to evaluate the causal effect of treatment.
- SC is quite popular in social science due to the following features:
  1. It can evaluate treatment effects on **one (or very few) treated unit**
     - Aggregate outcomes (e.g. county-level crime rate)
  2. Use a **data-driven procedure** and **a small number of non-treated units** to build the suitable counterfactuals

Main Idea of Synthetic Control Method

- Main idea:
  - Use (long) panel data to build the **weighted average of non-treated units**
    - The **weighted average of non-treated units** is the **synthetic unit**
    - Synthetic unit can best reproduce characteristics of the **treated unit** over time in pre-treatment period
  - Causal effect of treatment can be quantified by:
    - A simple difference in the post-treatment period: **treated unit vs synthetic unit**

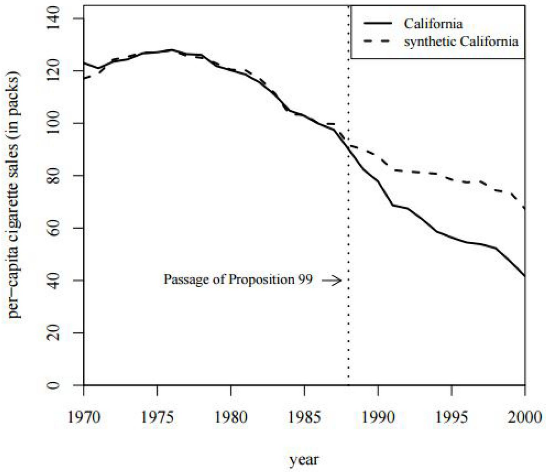
Main Idea of Synthetic Control Method

Graphical Representation



Main Idea of Synthetic Control Method

Graphical Representation



DID and SC

- DID and SC are often viewed as targeting different types of empirical applications
- DID methods are applied in cases:
  - Usually need a substantial number of units that are exposed to the treatment
  - Require a "parallel trends" assumption
- In contrast, SC methods are suitable in cases:
  - Only a single (or small number) of units exposed to the treatment
  - Seek to compensate for the lack of parallel trends by reweighting units to match their pre-exposure trends

Synthetic Control Method: Potential Outcome Framework

Basic Setup: Single Treated Model

- Suppose we observe $J + 1$ units over $t = 1, ..., T$ periods
- A "treatment" occurs at period $T_0 + 1$
  - Unit 1 being treated
  - Units $\{2, ..., J + 1\}$ being unaffected
  - Pre-treatment period: $1.....T_0$
  - Post-treatment period: $T_0 + 1......T$
- We aim to measure the causal effect of the treatment on the treated unit 1

SC and Potential Outcomes

- Treatment
  - $D_{it} = 1$: the units that are treated from periods $T_0 + 1$ until T
  - $D_{it} = 0$: the units that are always untreated

SC and Potential Outcomes

- Potential Outcomes
  - Y_{it}^1: the potential outcome we *would* observe for unit i at time t if unit i receives the treatment
    - Note that the treated unit would receive treatment from periods $T_0 + 1$ until T
  - Y_{it}^0: the potential outcome we *would* observe for unit i at time t if unit i does not receives the treatment
- Note that **unit in synthetic control method is usually aggregate level**: country, state, county, or region

SC and Potential Outcomes

- Observed Outcomes
  - Y_{it} is the observed outcome for unit i at time t
    - Observed outcomes before period $T_0 + 1$:

$$Y_{it} = \mathrm{Y}_{it}^{0}$$

    - Observed outcomes after period $T_0 + 1$:

$$Y_{it} = \mathrm{Y}_{it}^{0}(1 - D_{it}) + \mathrm{Y}_{it}^{1} D_{it}$$

SC and Potential Outcomes

- Since only unit 1 is treated, we aim to estimate the causal effect of treatment over time ($T_{0+1}, ..., T$) for the treated unit 1

$$\alpha_{1t} = (\alpha_{1T_0+1}, ..., \alpha_{1T})$$

where for $t > T_0$:

$$\alpha_{1t} = Y^1_{1t} - Y^0_{1t} = \underbrace{Y^1_{1t}}_{\text{observed}} - \underbrace{Y^0_{1t}}_{\text{counterfactual}}$$

- Therefore, we need to construct the **unobserved counterfactual**

SC Estimation

$$\alpha_{1t} = Y^1_{1t} - Y^0_{1t} = \underbrace{Y^1_{1t}}_{\text{observed}} - \underbrace{Y^0_{1t}}_{\text{counterfactual}}$$

- SC method suggests treatment effect can be estimated by the simple difference:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{i=2}^{J+1} w_i^* Y_{it}$$

SC Estimation

- Choose $W = (w_2^*, ..., w_{J+1}^*) \in [0,1]$ to minimize difference in pre-treatment characteristics X between treated and weighted average of non-treated units
  - Minimize $||X_1 - X_i W||$
  - X includes observed characteristics Z and pre-treatment outcomes $Y_1, ... Y_{T0}$
  - subject to $\sum_{i=2}^{J+1} w_i^* = 1$
- Thus, different weights W gives different synthetic units

Synthetic Control Method: An Empirical Example

Synthetic Control Method
Example

Timo Mitze, Klaus Wälde, Reinhold Kosfeld, and Johannes Rode
"**Face Masks Considerably Reduce COVID-19 Cases in Germany: A Synthetic Control Method Approach**"
Proceedings of the National Academy of Sciences (2020)

- The authors estimate the causal effect of face masks on the spread of Covid-19 using synthetic control method

Motivation

- Many countries have experimented with several public health measures to mitigate the spread of Covid-19.
  - One particular measure that has been introduced are face masks.
- The effect of face masks worn in public on the spread of Covid-19 has not been systematically analyzed so far.

The Timing of Mandatory Mask Wearing in Germany

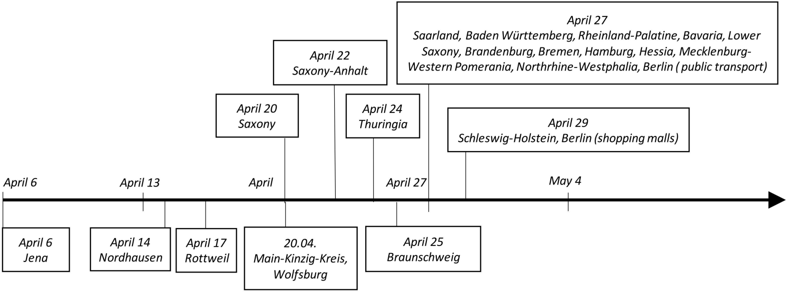


Figure 1: The timing of mandatory mask wearing in federal states (top) and individual regions (below)

Background

- Jena was the first region to introduce face masks in public transport and sales shops on 4/6
- By 4/29, all German regions had introduced face masks.
- Jena is a representative case for studying the Covid-19 development:
  - On 4/5, the cumulative number of registered Covid-19 cases in Jena was 144 (the median of 155 for Germany)
  - Similarly, the cumulative number of Covid-19 incidences per 100,000 inhabitants was 126.9 in Jena compared to a mean of 119.3 in Germany

Data

- Official German statistics on reported Covid-19 cases from the Robert Koch Institute
- They build a balanced panel for 401 regions.
- Sample period: 95 days spanning the period from January 28 to May 1, 2020

SC Estimation

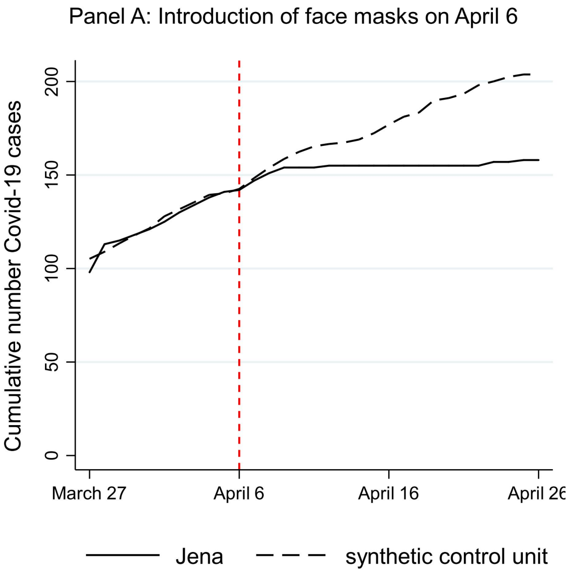
$$\alpha_{1t} = Y^1_{Jena,t} - Y^0_{Jena,t} = \underbrace{Y^1_{Jena,t}}_{\text{observed}} - \underbrace{Y^0_{Jena,t}}_{\text{counterfactual}}$$

- SC method suggests treatment effect can be estimated by the simple difference:

$$\hat{\alpha}_{1t} = Y_{Jena,t} - \sum_{i=1}^{401} w_i^* Y_{it}$$

- Choose $W = (w_1^*, ..., w_{401}^*) \in [0,1]$ to minimize difference in pre-treatment characteristics X between treated and weighted average of non-treated units
  - Minimize $||X_1 - X_i W||$
  - X includes observed characteristics Z and pre-treatment outcomes $Y_1, \ldots Y_{T0}$
  - subject to $\sum_{i=1}^{401} w_i^* = 1$

Results


Panel A: Introduction of face masks on April 6
Cumulative number Covid-19 cases
0
50
100
150
200
March 27
April 6
April 16
April 26
Jena
synthetic control unit

Results

Table A2: Pre-treatment predictor balance and RMSPE for SCM in Figure 2

Treatment:	**Introduction of face masks**		**Announcement/ start of campaign**	
	Jena	Synthetic control group	Jena	Synthetic control group
Cumulative number of registered Covid-19 cases (one and seven days before start of treatment, average)	129.5	129.2	93	92.7
Number of newly registered Covid-19 cases (last seven days before the start of the treatment, average)	3.7	3.8	5	5.2
Population density (Population/km^2)	38.4	22.8	968.1	947.9
Share of highly educated population (in %)	968.1	1074.3	38.4	26.3
Share of female in population (in %)	50.1	50.1	50.1	50.1
Average age of female population (in years)	43.5	43.7	43.5	43.9
Average age of male population (in years)	40.5	40.6	40.5	40.8
Old-age dependency ratio (in %)	32.1	29.3	32.1	29.8
Young-age dependency ratio (in %)	20.3	19.6	20.3	19.5
Physicians per 10,000 of population	20.5	19.8	20.5	20.8
Pharmacies per 100,000 of population	28.8	28.7	28.8	28.6
Settlement type (categorial variable)	1	1.3	1	1.9
RMSPE (pre-treatment)	**3.145**		**4.796**	

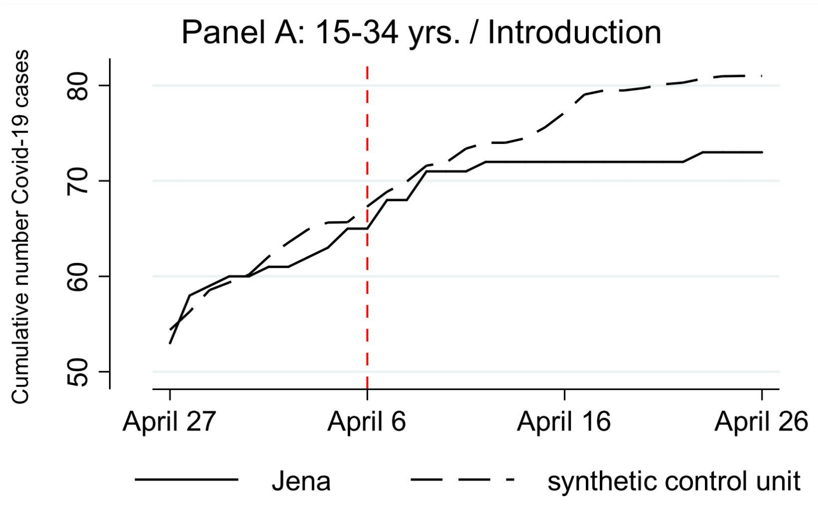
Notes: Donor pool includes all other German NUTS3 regions except the two immediate neighboring regions of Jena (Weimarer Land, Saale-Holzland-Kreis) as well as the regions Nordhausen and Rottweil since the latter regions introduced face masks in short succession to Jen on April 14 and April 17.

Results

Table A3: Distribution of sample weights in donor pool for synthetic Jena

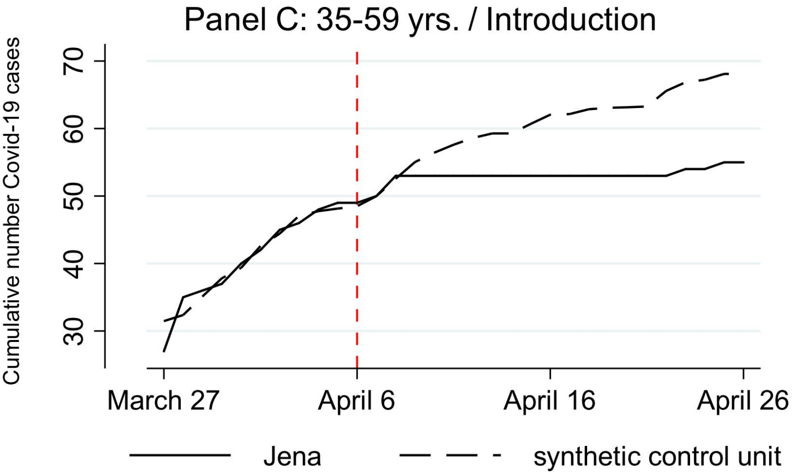
Introduction of face masks (Panel A in Figure 2)		
ID	**NUTS 3 region**	**Weight**
13003	Rostock	0.326
6411	Darmstadt	0.311
3453	Cloppenburg	0.118
7211	Trier	0.117
6611	Kassel	0.082
5370	Heinsberg	0.046

Results

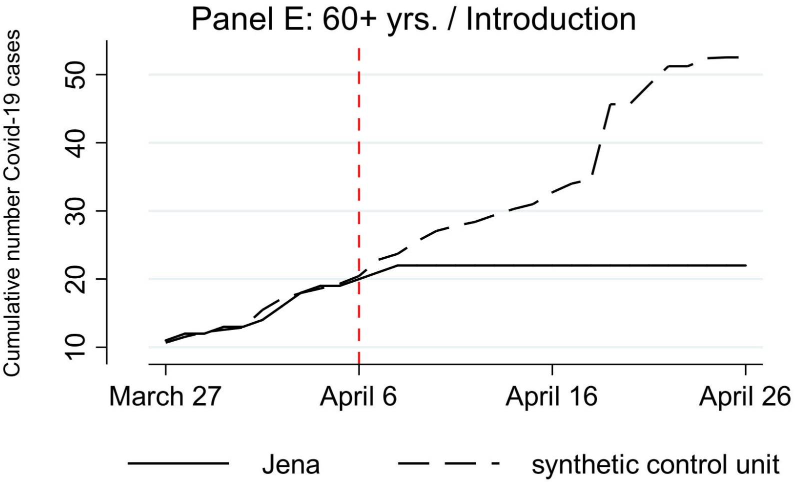


Panel A: 15-34 yrs. / Introduction
Cumulative number Covid-19 cases
80
70
60
50
April 27
April 6
April 16
April 26
Jena
synthetic control unit

Results


Panel C: 35-59 yrs. / Introduction
Cumulative number Covid-19 cases
30
40
50
60
70
March 27
April 6
April 16
April 26
Jena
synthetic control unit

Results


Panel E: 60+ yrs. / Introduction
Cumulative number Covid-19 cases
10
20
30
40
50
March 27
April 6
April 16
April 26
Jena
synthetic control unit

Summary of Results

- The early introduction of face masks in Jena has resulted in a reduction of almost 25% in the cumulative number of reported Covid-19 cases after 20 days.
  - The drop is greatest, larger than 50%, for the age group 60 years and above.
- This corresponds to a reduction in the average daily growth rate of the total number of reported infections by 1.32 percentage points.
  - This represents 60% reduction.

Concluding Remarks

- Estimating "causal effects" is a challenging task
  - It might confound with **selection bias** and the effects of other factors
  - Thus, many studies can only get **correlation** not **causality**
- Once you can estimate causal effect convincingly
- Your finding will improve our understanding of human behaviors and society