Applied Econometrics: An Introduction
應用計量經濟學: 課程介紹

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
AGEC/STAT, NTU

September 4, 2024

About Me

- Name: Yang, Tzu-Ting (楊子霆)
- Affiliation:
  - Institute of Economics, Academia Sinica
  - Office at Academia Sinica: 中央研究院經濟研究所 B308
- Other Appointments:
  - NTU-AGEC, NTU-ECON, and NCCU-IMES
  - Office at NTU: 農業綜合館 218-B5 / 社科院 724
- Research Fields: Public/Labor Economics and Applied Econometrics
- Website: https://sites.google.com/view/cpelab/
- Email: ttyang@g.ntu.edu.tw

Teaching Assistants

- 黎宏濬 (農經碩二)
  - Email: `r11627065@ntu.edu.tw`
- 吳子欣 (農經碩二)
  - Email: `r12627007@ntu.edu.tw`

This Course

- The goal of this course is equip students with a comprehensive set of statistical tools that are useful in conducting high-quality empirical research in economics
- Specifically, the course places a strong emphasis on **causal inference** and understanding their applications
- We will especially focus on the practical implementation of these empirical methods by writing a term paper
  - How to conduct an empirical research
  - Provide a good start for your thesis

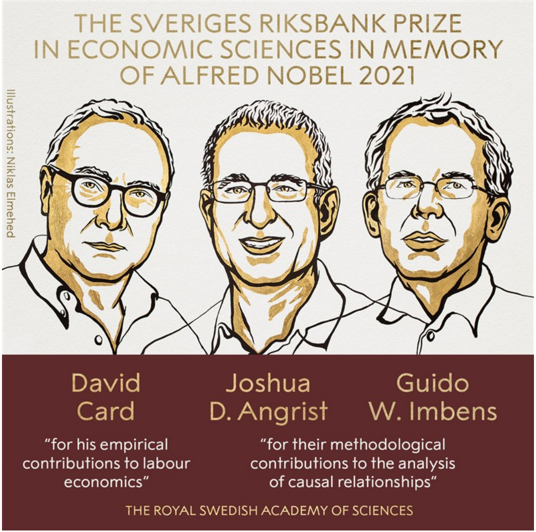
Economics and Causal Inference

Economics and Causal Inference

- Empirical research is experiencing two methodological "revolutions" over the past few decades
- On the one hand, there is the "credibility revolution"
  - A movement that emphasizes the goal of empirical research is to understand causality

2021 Nobel Laureates

Causal Inference in Economics



Economics and Causal Inference

- On the other hand, there is the "big data revolution"
  - A movement that emphasizes how our increasing ability to collect and analyze vast amounts of data can transform our understanding of the human behaviors
- Recent trend in empirical research
  - Use large scale dataset to identify causal relationship

Economics and Causal Inference

- Economic theory plays an important role in the causal analysis of large data sets with complex structure
    - It can be difficult to study this type of data or even to decide which variables to construct
    - Economic models can provide conceptual frameworks to point out what are key variables or what kind of relationship we should care about
- Better data and more credible empirical methods can help researchers test economic theories that had previously been difficult to assess

This course

- This course will go through several useful techniques based on recent methodological developments in empirical methods
  - Focus on **causal inference** and its applications in economics

Causal Inference

Causal Inference

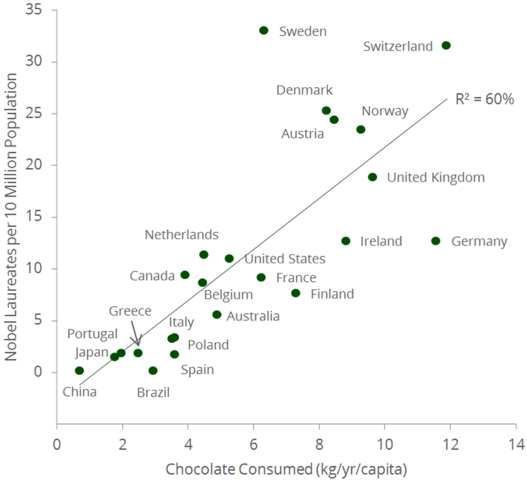
- Social science (Economics) theories are almost always causal in their nature
  - X causes Y
  - An increase in price of oil causes consumer's demand for oil to decrease
  - An increase in schooling years can raise people's productivity (wage)
  - Implementation of a carbon pricing incentivizes firms to adopt more environmentally friendly practices

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

- 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

- Macro economists also ask causal questions!
- The effect of changes in interest rates on house prices
  - Does increasing interest rates cause house prices to decrease?
  - Ideally, we should get causal effect by comparing the house prices of **the same economy** with and without the interest rate change
  - Again, we have an unobserved outcome problem

- The effect of changes in interest rates on house prices
  - Countries with low interest rates v.s. Countries with high interest rates:
    - Two groups are not directly comparable
  - Why do central banks change interest rates?
    - They might lower rates during economic downturns $\Rightarrow$ may underestimate the positive effect of low interest rates on house prices
    - Or, they might raise rates when the economy is overheating $\Rightarrow$ may underestimate the negative effect of high interest rates on house prices

- More examples include:
    - The effect of advertisement on product sales
    - The effect of military service on earnings and employment
    - The effect of climate change on crop yields
    - Do renewable energy subsidies lead to increased adoption of clean technologies?
    - Does eliminating estate tax increase wealth inequality?
    - Do immigrant workers depress the wages of native workers?
    - Can democracy increase economic growth?

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 are able to construct a comparison group that can represent counterfactual outcomes
- Then, we can obtain the causal effect of treatment

Course Content: Causal Inference

Randomized Experiment

- In this course, we will introduce at least 7 methods of causal inference:

1. Randomized Experiment
   - Randomly assign treatment ensures that every observation has the same probability of being assigned to the treatment group
   - The characteristics of treatment and comparison groups are similar since receiving treatment is unrelated to any other confounding factors
   - Then, we can obtain causal effect of treatment by simply comparing outcomes between treatment and comparison groups

2 Matching Methods

- Assume key differences between treatment and comparison groups are **observable**
- Construct a comparison group that have similar **observable** characteristics as treatment group

3 Regression and Causal Machine Learning

- Use regression to control for observable confounding factors
- Use machine learning method to decide which observable characteristics is important so that we should include in regression
    - Post-Double selection method

Differences-in-Differences

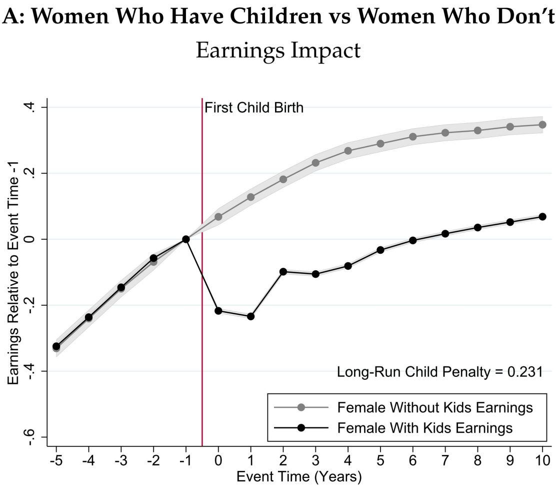
4 Differences-in-Differences (DID)

- If treatment and comparison group's outcomes move in parallel in the absence of treatment
- Then, we can use trend in outcome of a comparison group to represent counterfactual trend for the treatment group

Differences-in-Differences

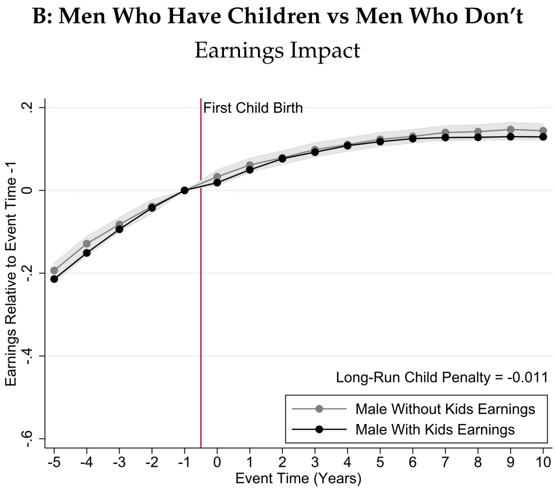
- Example: The effect of having children on female earnings
  - Despite considerable gender convergence over time, substantial gender inequality persists in all countries
  - Henrik Kleven et. al (2019) uses Danish administrative data from 1980-2013 and an DID approach
  - They show that most of the remaining gender inequality in earnings is due to children
  - The arrival of children creates a gender gap in earnings of around 20% in the long run

Differences-in-Differences



Source: Henrik Kleven et. al (2018)

Differences-in-Differences

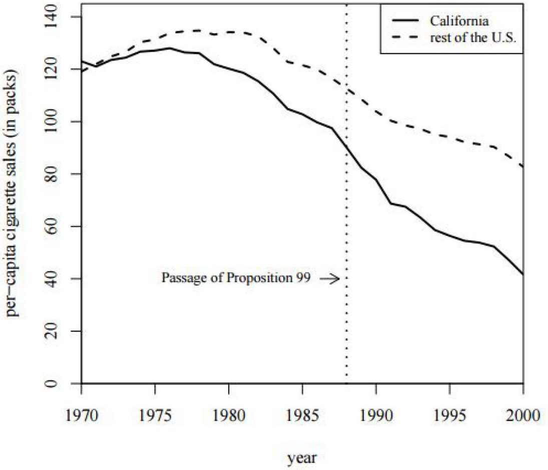


Source: Henrik Kleven et. al (2018)

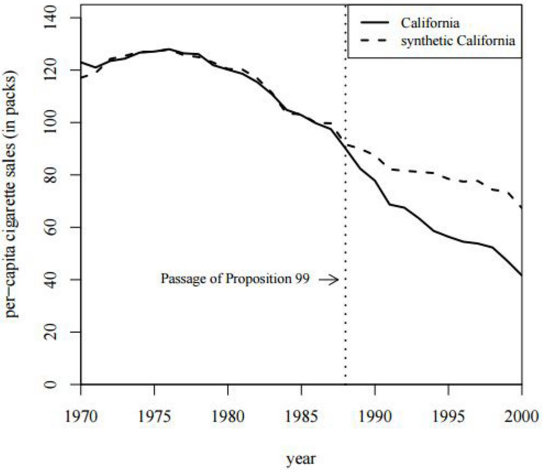
5 Synthetic Control Method (SCM)

- In some situations, treatment and comparison group's outcomes do not move parallelly before a treatment happens
- Use **data-driven procedure** and **a small number of non-treated units** to build a suitable counterfactual outcome

Synthetic Control Method



Synthetic Control Method



6 Regression Discontinuity Design (RDD)

- When a treatment is applied depending on some thresholds
  - Assume the choices of thresholds are arbitrary
- We can estimate causal effects by comparing outcomes for those just above threshold and those just below threshold
  - Two groups should be similar since they are around threshold

Regression Discontinuity Design

- Example: The effect of college major on early-career wages
  - Forty-year-old US workers with undergraduate degrees in economics earned median wages of $90,000 in 2018
    - By comparison, college graduates with any major other than economics earned $66,000
  - However, average wage differences between majors do not necessarily reflect the causal effect of choosing one major over another
    - Most students self-select their college major, and many universities use grade requirements to restrict entry into certain majors
    - Observational wage differences across majors may reflect other confounding factors

Regression Discontinuity Design

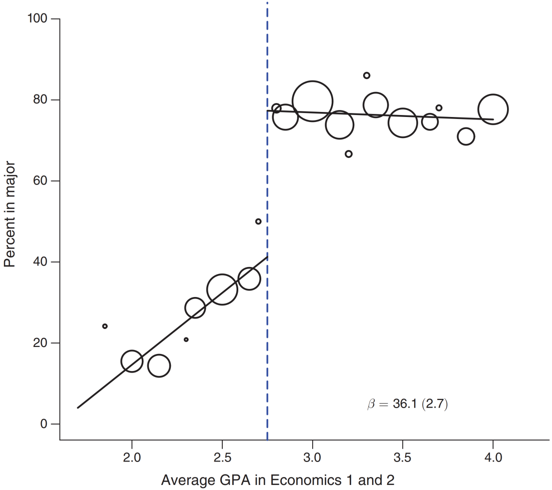


FIGURE 1. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON MAJORING IN ECONOMICS

Source: Bleemer, Zachary, and Aashish Mehta (2022)

Regression Discontinuity Design

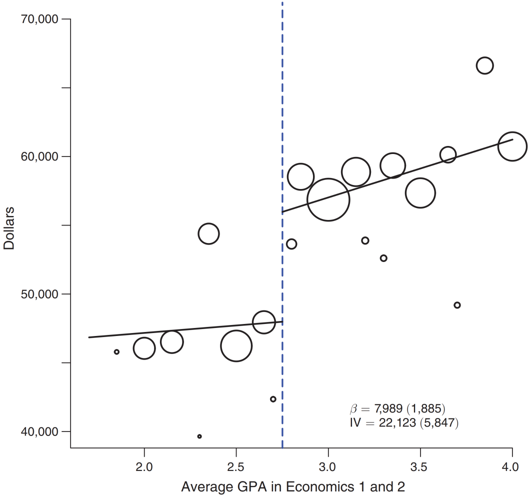


FIGURE 2. THE EFFECT OF THE UCSC ECONOMICS GPA THRESHOLD ON ANNUAL WAGES

Source: Bleemer, Zachary, and Aashish Mehta (2022)

Regression Discontinuity Design

- Because students with GPAs just below 2.8 are generally similar to students with GPAs just above 2.8, on average
  - The only difference could be their major
  - Those just above 2.8 GPA threshold have much higher wage
  - This suggests that the economic major causes this effect
- Students who just met the 2.8 GPA threshold to major in economics earned $22,000 (46%) higher early-career wages than they would have in their second-choice majors

7 Instrumental variables

- The instrumental variable (IV) is:
  - An exogenous source of variation that drives the treatment
  - But it is unrelated to other confounding factors that affect outcome
- Intuitively, IV breaks variation of the treatment into two parts
  1. A part that might be correlated with other confounding factors
  2. A part that is not (driven by IV)
- We can use the variation in treatment that is driven by IV to estimate causal effect of the treatment

Advanced Topics

1. Shift-Share IV Design
   - Utilizes an instrument based on national trends in the treatment exposure that are unrelated to local confounders
2. Staggered DID Design
   - Treatment adoption that occurs at different times across units
3. Synthetic DID Design
   - Combine synthetic control method and DID Design

Advanced Topics

4 Spatial RD Design

- Estimate treatment effects by comparing observations just above and below a geographic boundaries for treatment assignment

5 Causal Forest

- A machine learning technique used to estimate heterogeneous treatment effects

Course Content: Data Analysis

Data Analysis

- A good causal inference requires a well-established DATA
- Create an "analysis-ready" dataset is a challenging task, especially for large-scale data or unstructured data
- A lot of data analysis time is spent data cleaning and preparing data, up to 80% of the time.
- In this course, you will learn how to clean data, create your own dataset and visualize your data
  - You might also learn how to collect your own data

Data Analysis

- Economists had a long tradition of utilizing the evidence from data to verify their theories
- In the past, the major data sources were the government surveys
- The data revolution of the past decade have a further and profound effect on economic research

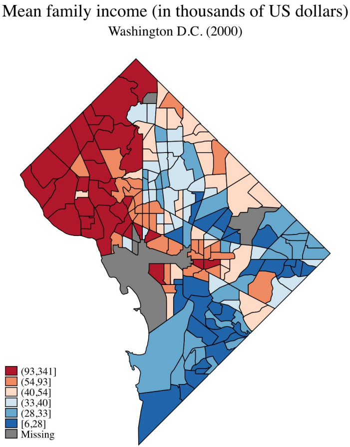
Data Analysis

- Increasingly, economists make use of newly available large-scale administrative data with near-universal population coverage
  - Health insurance claims data:
    - Record every Taiwanese's healthcare utilization whenever they visit doctors
  - Tax return data:
    - Record income and wealth of each taxpayer
  - Housing transactions Data:
    - Record all housing and land transactions in Taiwan

Data Analysis

- Due to the growth of the internet, economists also begin to use new data formats (unstructured data)
    - Online document
    - Social media
    - Geolocations
- In this course, I will also teach you how to handle with these new types of datasets
    - Geographic data

Geographic data



Source: Maurizio Pisati (2012)

Course Structure

1. Focus on how to implement various empirical methods of drawing causal inference
2. Discuss the applications in economics
3. Let you know how to use statistical softwares to conduct data analysis

Reading Materials

- Lecture slides: posted on my website
- Suggested Readings:
  - **The Effect: An Introduction to Research Design and Causality** by Huntington-Klein
  - **Causal Inference: The Mixtape** by Scott Cunningham
    - New textbook and cover more methods
    - Provide STATA and R examples
  - **Econometric Methods for Program Evaluation** by Alberto Abadie and Matias D. Cattaneo
    - This is an academic paper not a textbook
    - It can help you understand causal inference methods in a short time

Reading Materials

- Suggested Readings:
  - **Mastering Metrics: The Path from Cause to Effect** by Angrist and Pischke
    - Chatty, opinionated, but intuitive approach to causal inference
  - **Mostly Harmless Econometrics** by Angrist and Pischke
    - More advanced
  - **An Introduction to Statistical Learning with Applications in R** by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
    - An introductory book for machine learning

Course Goals, Grading Policy, and Requirements

Course Goals

- Get a solid understanding of the empirical methods to estimate causal effect and conduct data analysis
  - Be able to implement a good empirical research
  - Be able to critically evaluate empirical studies
- Be familiar with techniques and tricks of data management and visualiztion
  - Use STATA
  - Use R
- Have a good start of your thesis/writing sample

Grading Policy

- Two empirical homework (20%)
- Reading group presentation (20%)
- Term paper presentation (20%)
- Term paper (40%): milestones throughout the semester
- You will get extra 5 points in your grade for term paper (equivalent to 2 points in final grade) if you upload your codes and related files to *GitHub* for replication.

Course Requirements

- You should use **Latex** to type your term paper in Chinese or English
  - **Latex** is a tool for typesetting professional-looking documents
- In addition, you are encouraged to upload your code to **GitHub** for replication (Bonus!)
  - **GitHub** is an online repository that store and share your source code projects
- You can use "homework" to practice the above "requirements"

Important Dates

- Compulsory Office Hour: 10/9 week and 11/20 week
- Homework 1: 10/20
- Homework 2: 11/24
- Reading group presentation: 11/20, 11/27 and 12/4
- Term paper presentation: 12/4, 12/11 and 12/18
- Term paper deadline: 12/30

Two Compulsory Office Hour

10/9 week and 11/20 week

- We will have two compulsory office hour
    - Help you find a research topic: brainstorming
- Before each office hour, please send me an **research questions slide** (1-5 pages)
- Describe 1-2 research ideas
- For each idea, you should briefly describe causal relationship you are interested in and possible dataset you can use
- If possible, you should try to point out possible empirical methods
- 5 minutes presentation

Reading group presentation

11/20, 11/27 and 12/4

- Present one of the paper that applies causal inference from reading list
- Students in a group of **3-4** persons will give a presentation
  1. Introduction and Background
  2. Data and Empirical strategy
  3. Results and Conclusion
- Around 30-40 minimutes

Term paper presentation

12/4, 12/11 and 12/18

- Present the research progress of your term paper
- 10 minutes presentation
  - Introduce your research question
  - Discuss your empirical methods
  - Describe the data you use and summary statistics of estimated sample
  - Discuss your preliminary results

Term paper deadline

- Feel free to discuss your term paper with me before the deadline
- Send your term paper to me through email
- Email: ttyang@g.ntu.edu.tw

Guideline for Writing a Term Paper

Guideline for Writing a Term Paper

- You should start early, the paper is due on 12/30
- Letter style: roughly 5-10 pages including tables, figures, footnotes, appendices, and references
- Word count: less than 3,000 words
  - See **Economics letters**
  - See **AER: Insight**
- Typed, double-spaced, and using one-inch margins and 12 point type

Guideline for Writing a Term Paper

- For senior graduate students, you cannot just submit your thesis as a term paper
    - Let me know if you have any question about this issue

Guideline for Writing a Term Paper

- Use credible causal inference methods to answer an empirical question
    - Test economics (social science) theory
    - Estimate policy effect
    - Any interesting questions regarding to human behavior/social phenomenon
- **Don't worry if you don't find anything significant as long as your methods are credible and you have interpreted the results well**

How to Find Research Topics

Approaches to Find Research Topics

- There are two main approaches to identifying research topics:
  1. Starting from your own interests and curiosities
  2. Doing an extensive literature review first
- These approaches are not mutually exclusive but iterative, with different starting points.

Approaches to Find Research Topics

Starting from your own interests and curiosities

- I personally prefer the first approach
  - It allows you to arrive at topics you are really interested in
  - You can start by asking questions based on your personal experience
- Then, examine the current literature to see the state of knowledge and feasibility given accessible resources for answering the research question
- However, the risk is higher as the topic may be unimportant or boring for other people
- Requires personal judgment

Approaches to Find Research Topics

Doing an extensive literature review first

- This is more common approach
  - Review important literature in your broad area first
  - Focus on high quality papers (e.g. NBER working paper, top journals)
- Then, identify extensions or gaps in knowledge
- Examine feasibility given accessible resources for answering the research question