Topics in Labor Economics: Empirical Methods and Applications

Prof. Tzu-Ting Yang
楊子霆

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

February 20, 2023

About Me

- Name: Yang, Tzu-Ting (楊子霆)
- Affiliation: Institute of Economics, Academia Sinica
- Other Appointments: Econ, NTU and IMES, NCCU
- Research Fields: Public/Labor Economics and Applied Econometrics
- Office at NTU: Social Science Building 724
- Office Hour: Make an appointment
- Website: causaldatalab.wordpress.com
- Email: ttyang@econ.sinica.edu.tw

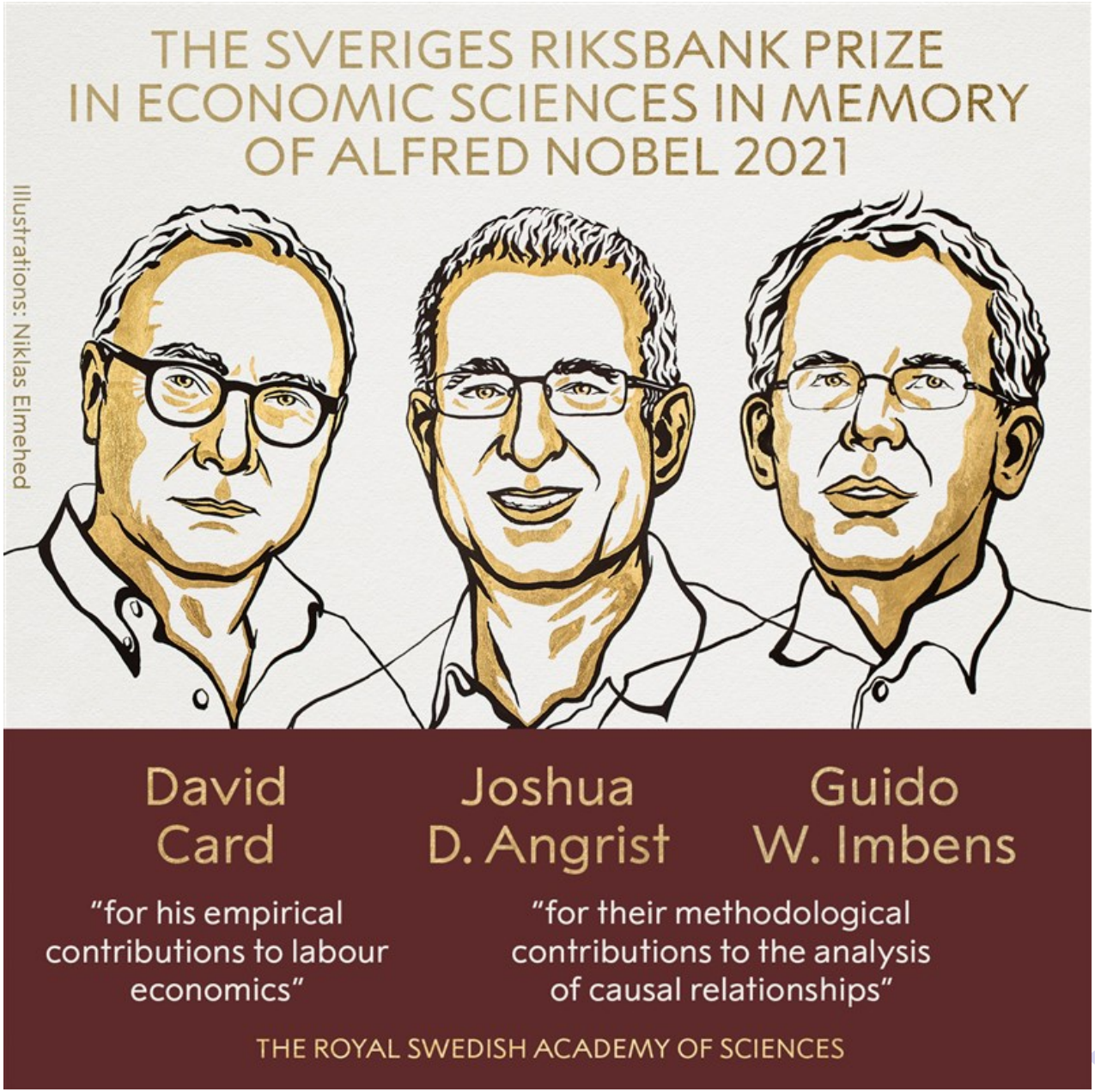
Labor Economics and Empirical Methods

Labor Economics and Empirical Methods

- 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

Labor Economics and Causal Inference



Labor Economics and Empirical Methods

- 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

Labor Economics and Empirical Methods

- 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 labor economics
  - How to conduct an empirical research

Causal Inference

Causal Inference

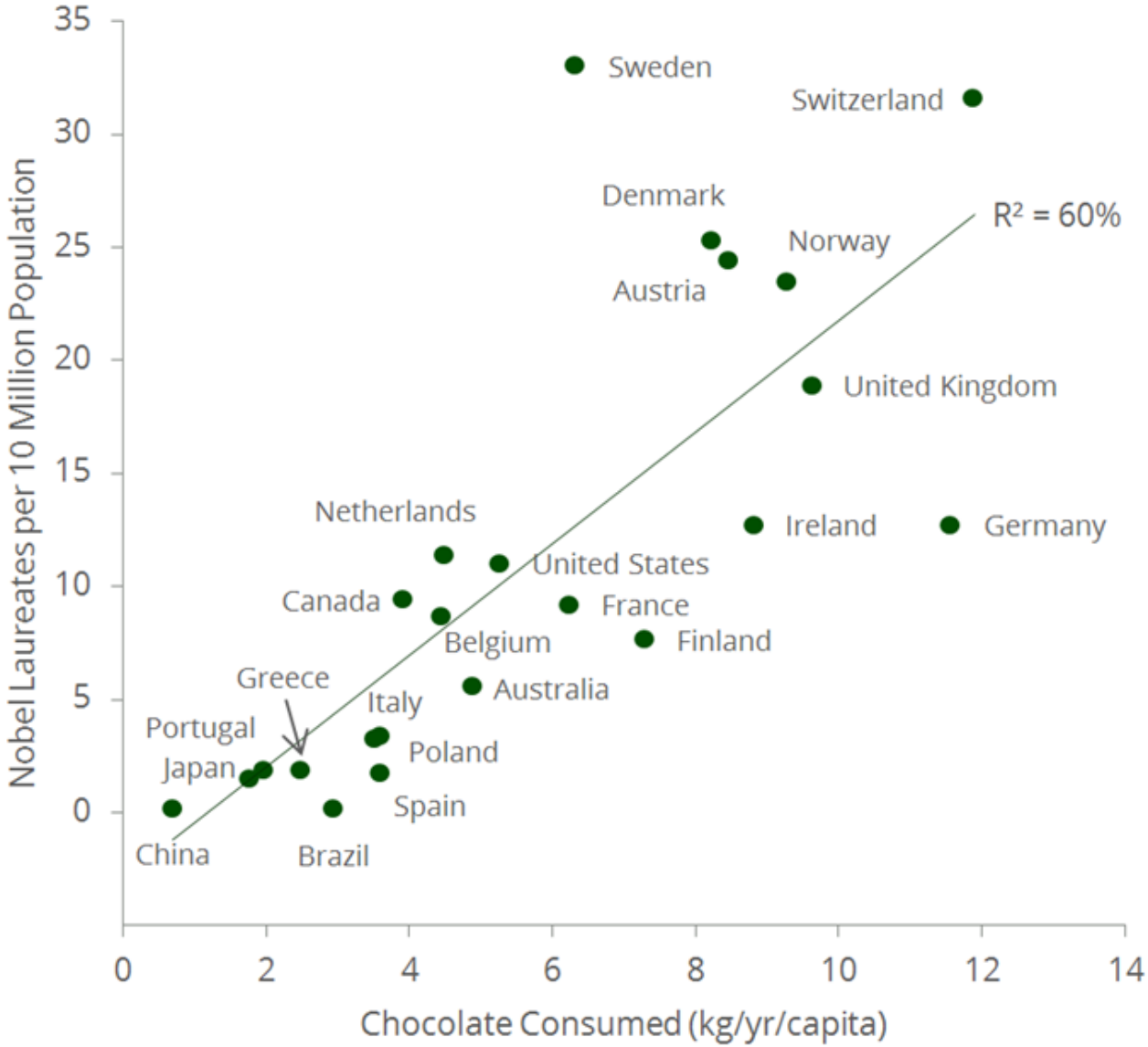
- Social science (Economics) theories are almost always causal in their nature
  - X causes Y
  - An increase in price of oil causes consumer's demand for oil to decrease
  - An increase in schooling years can raise people's productivity (wage)
  - Raising minimum wage would reduce employment opportunity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

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

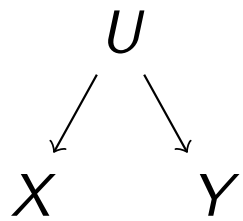
Chocolate Consumption and Nobel Laureates



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



- 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 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 machine learning method to decide which observable characteristics is important when we construct a comparison group
  - Double selection method

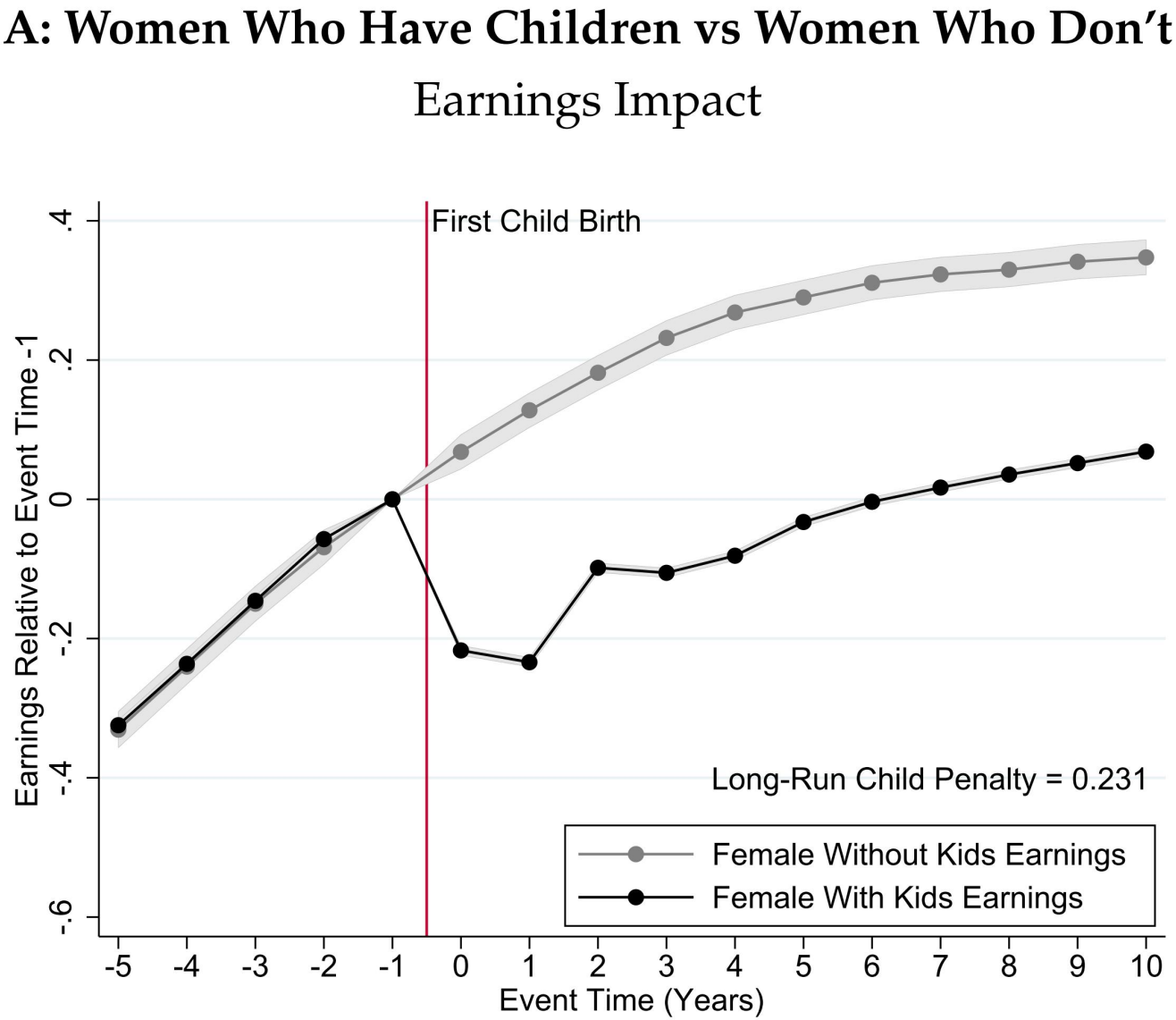
4 Differences-in-Differences (DID)

- 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

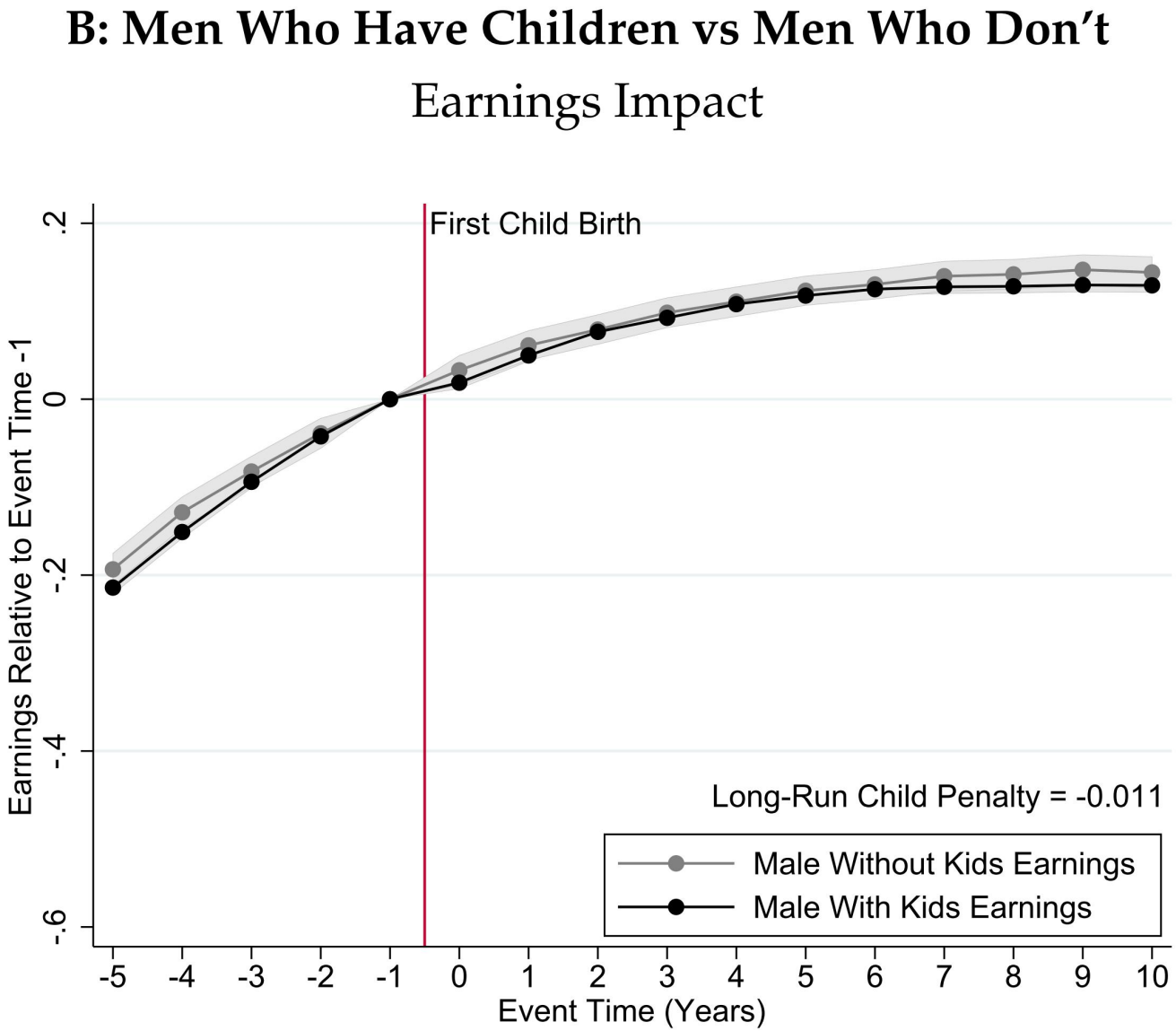
- Example: The effect of having children on female earnings
  - Despite considerable gender convergence over time, substantial gender inequality persists in all countries
  - Henrik Kleven et. al (2019) uses Danish administrative data from 1980-2013 and an DID approach
  - They show that most of the 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



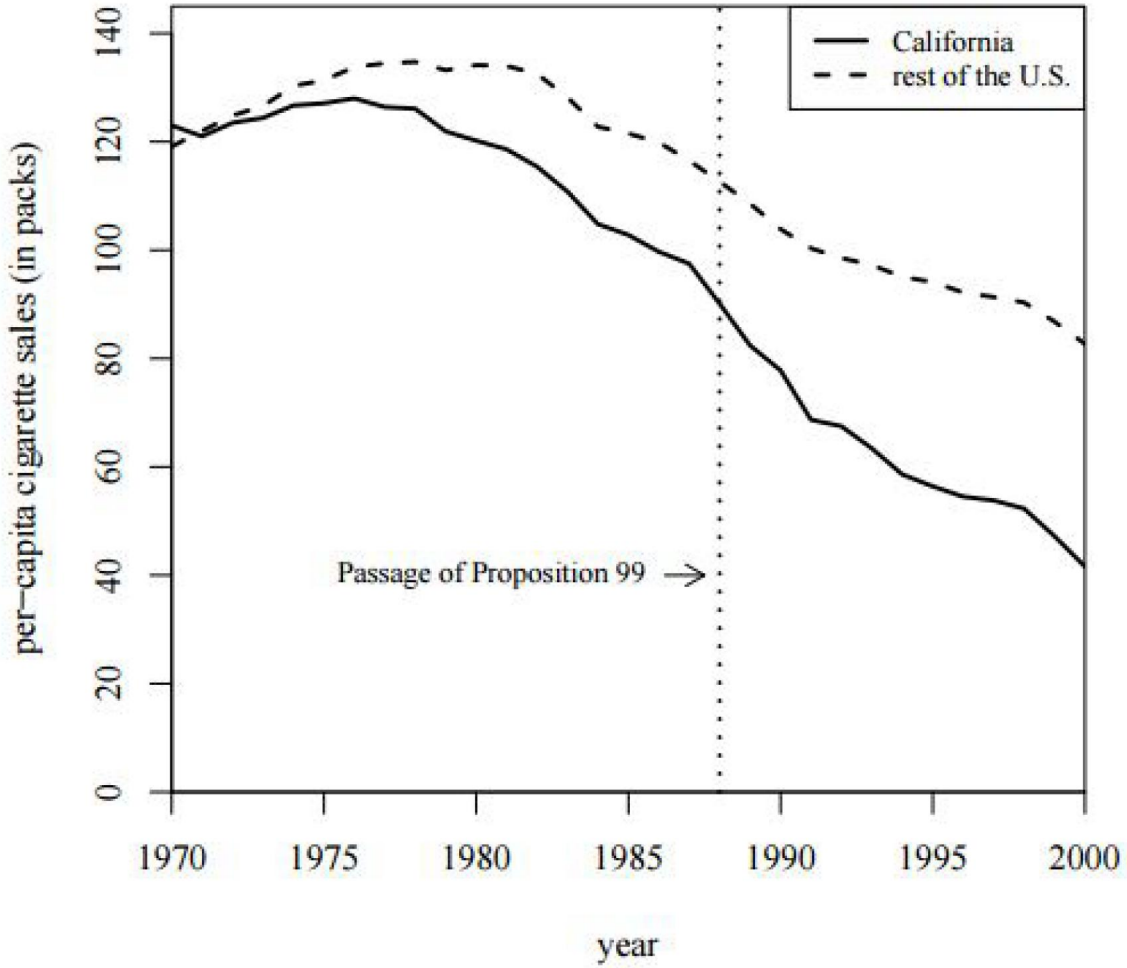
Source: Henrik Kleven et. al (2018)

Synthetic Control Method

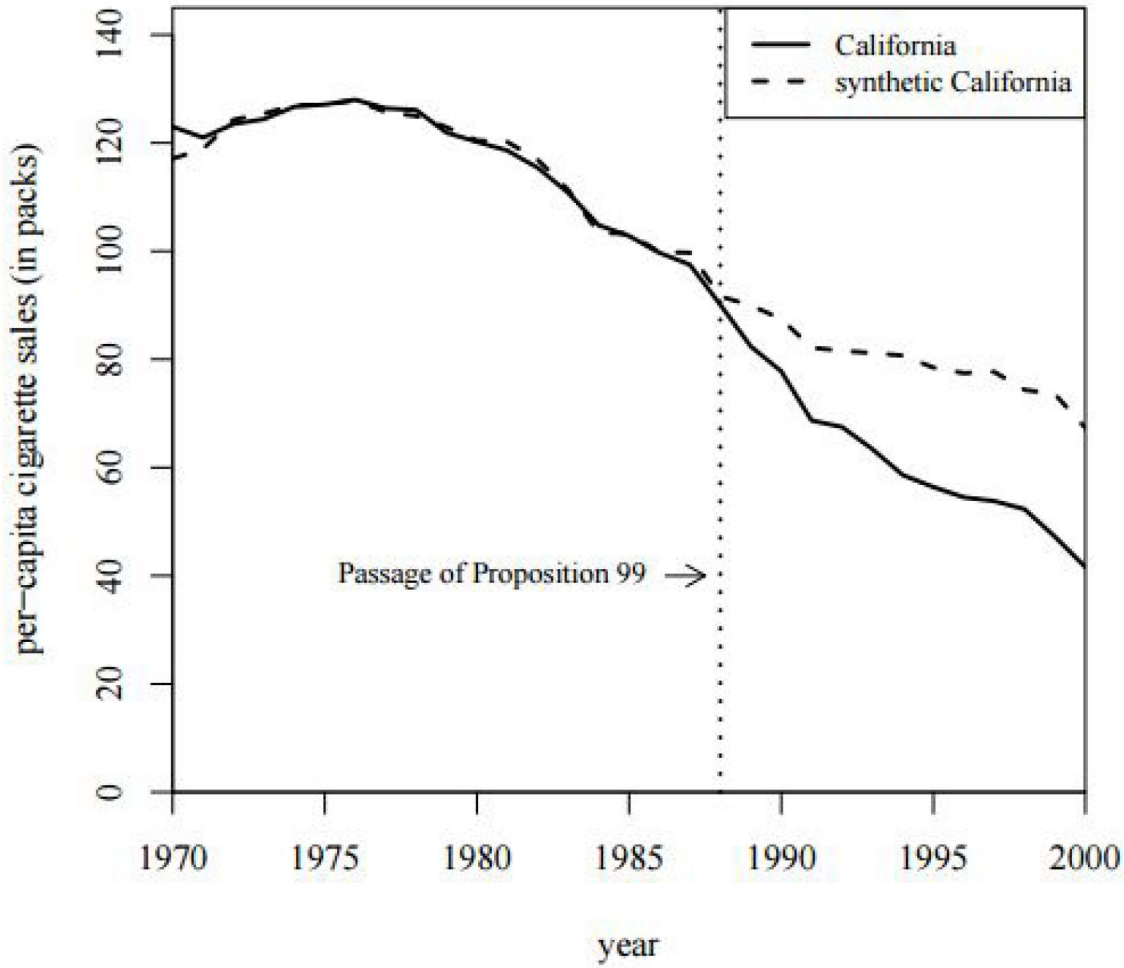
5 Synthetic Control Method (SCM)

- 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



Regression Discontinuity Design

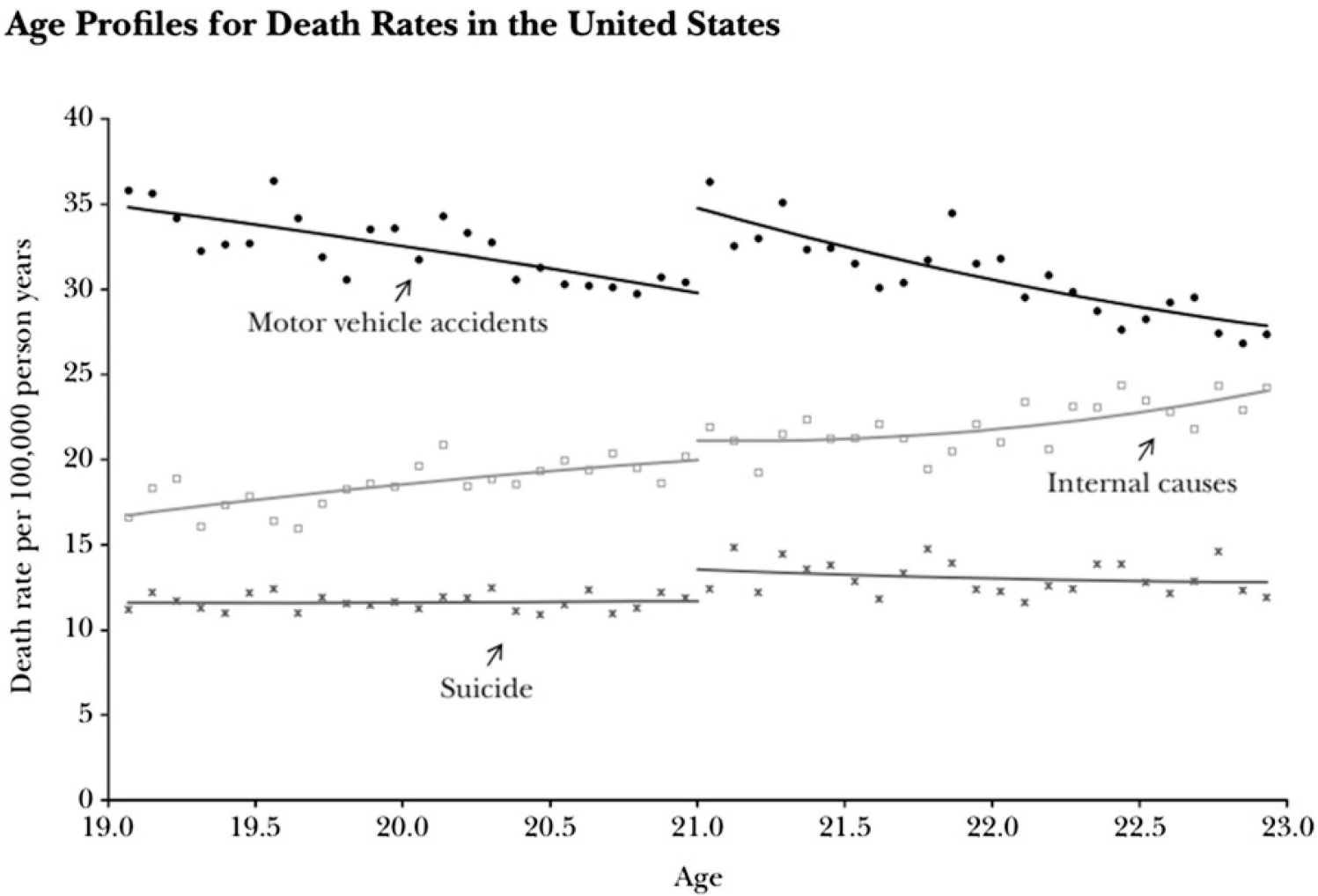
6 Regression Discontinuity Design (RDD)

- 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 alcohol consumption on mortality
  - In the US, young people can drink legally after their 21st birthday
    - Minimum legal drinking age
  - Due to this law, a small change in age (measure in months or even days) generates a big change in legal access to alcohol
  - This difference leads to substantial jump in mortality rate of motor vehicle accidents right after 21st birthday
    - See Christopher Carpenter and Carlos Dobkin (2009)

Regression Discontinuity Design



Source: Christopher Carpenter and Carlos Dobkin (2009)

Regression Discontinuity Design

- Since those just below age 21 is not that different from those just above age 21, on average
  - The only difference could be alcohol consumption
- Those just above age 21 have much higher death rates from automobile accidents
- This suggests that the change in alcohol consumption causes this effect

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

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

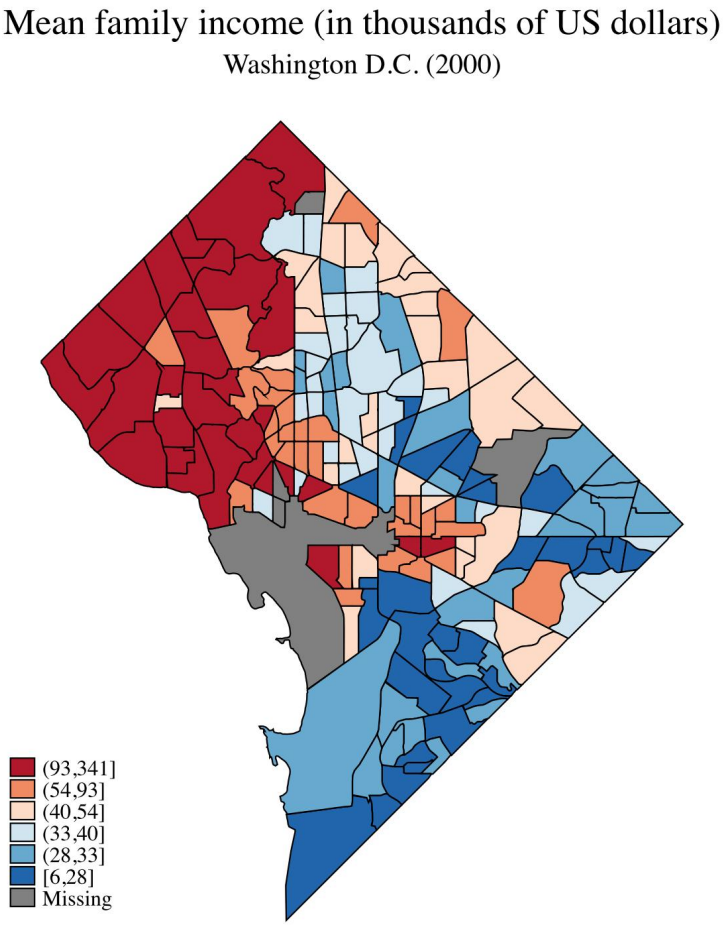
Data Analysis

- 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

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)

Applications in Labor Economics

- Labor Market Discrimination
- Human Capital and Earnings
- Corporate tax and Labor Demand
- Minimum Wage and Labor Demand
- Unemployment Insurance and Job Search
- Pension and Labor Supply
- Fertility and Female Labor Supply

Reading Materials

- Lecture slides: posted on my website
- Suggested Readings:
  - **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
  - **Labor Economics** by Pierre Cahuc, Stephane Carcillo and Andre Zylberberg (graduate level)
  - **Labor Economics** by George Borjas (undergraduate level)

Course Goals, 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 compulsory office hours (10%)
  - Discuss your research topics during office hour
- Two empirical homework (20%)
- Research progress presentation (10%)
- Term paper presentation (15%)
- Feedback on your peer's research (5%)
- Term paper (40%): milestones throughout the term

Course Requirements

- You should use **Latex** to type your term paper in English
  - **Latex** is a tool for typesetting professional-looking documents
- In addition, you also should upload your code to **GitHub** for replication
  - **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: 3/27 and 5/1
- Homework 1: 4/10
- Homework 2: 5/22
- Research progress presentation: 5/8
- Term paper presentation: 6/5
- Term paper deadline: 7/1

Two Compulsory Office Hour

3/27 and 5/1

- We will have two compulsory office hour
  - They will be held online during the weeks of 3/27 and 5/1
  - 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

Research progress presentation

5/8

- Present your research topic within 10 minutes
- Introduce your research question
- Briefly discuss related previous literature
- Describe the data and your empirical methods you will use

Term paper presentation

6/5

- Present your term paper within 15 minutes
- Introduce your research question
- Discuss your empirical methods
- Describe the data you use and summary statistics of estimated sample
- Discuss your preliminary results

Feedback on your peer's research

- You need to attend one of your classmates' office hour and give them some suggestions on their research
- You should also give comments or ask questions during your peer's presentation

Term paper deadline

7/1

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

Guideline for Writing a Term Paper

- You should start early, the paper is due on 7/1
- Letter style: roughly 5-10 pages including tables, figures, footnotes, appendices, and references
- Word count: less than 2,000 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**

Course Structure

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