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

March 26, 2026

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-ECON, NTU-AGEC and NCCU-IMES
  - Office at NTU: 社科院 724 / 農業綜合館 218-B5
- Research Fields: Public/Labor Economics and Applied Econometrics
- Website: https://sites.google.com/view/cpelab/
- Email: ttyang@g.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 (labor) 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 or writing sample

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
  - Labor economists play an crucial role in this movement

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
    - Emerging role of generative AI in automating data processing, coding, and workflow design
    - AI-assisted tools reduce technical barriers, but do not replace identification strategy or economic reasoning

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

Causal Inference

Causal Inference

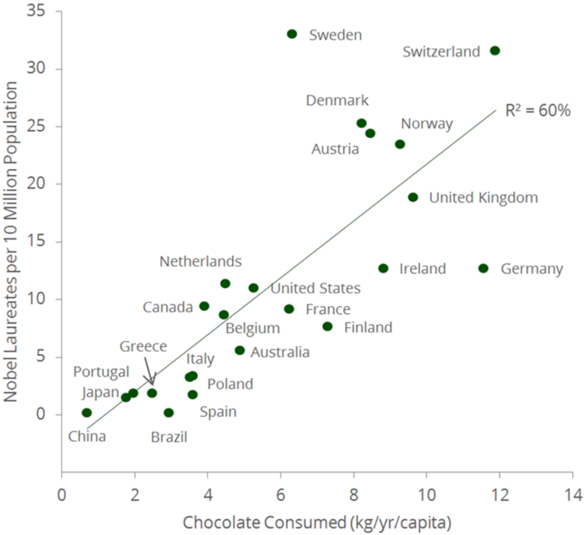
- Social science (Economics) theories are almost always causal in their nature
  - X causes Y
  - An increase in price of oil causes consumer's demand for oil to decrease
  - An increase in schooling years can raise people's productivity (wage)
  - Raising minimum wage would reduce employment opportunity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

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

- 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

- 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 fiscal stimulus spending on economic growth
  - Does government stimulus increase GDP growth?
  - Ideally, we want to compare the **same country** with and without stimulus
  - But we cannot observe both outcomes

Causal Inference

Example 2

- Compare countries with stimulus vs. without stimulus?
    - Governments often implement stimulus during recessions
    - Countries with weak growth are more likely to adopt stimulus
    - $\Rightarrow$ Simple comparison may underestimate the true effect

- 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 prices
  - Do immigrant workers depress the wages of native workers?
  - Does eliminating estate tax increase wealth inequality?
  - Can democracy increase economic growth?
  - What is the effect of **AI adoption** on firm productivity?
  - Does the introduction of **generative AI tools** change worker wages or employment?
  - What is the long-term impact of COVID-19 on the global 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

0 Randomized Experiment (RCT)

- 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

Gold Standard

Randomized Experiment

- RCT is the gold standard of causal inference
- However, it has some limitations
  - Often costly, unethical, or infeasible
  - Many important economic questions cannot be randomized
- In this course:
  - We briefly introduce RCT logic
  - Main focus: the methods when we can not implement RCT

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

4 Differences-in-Differences Design (DID)

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

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

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

7 Instrumental Variables Design

- 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

8 Shift-Share IV Design

- Utilizes an instrument based on national trends in the treatment exposure that are unrelated to local confounders

9 Spatial RD Design

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

10 Causal Forest

- A machine learning technique used to estimate heterogeneous treatment effects

Course Content: Data Analysis

Data Analysis and Research Workflow

- Credible causal inference requires a well-constructed dataset
- Creating an "analysis-ready" dataset is often the most time-consuming step
    - Data cleaning, merging, reshaping, and validation
    - Often accounts for 70–80% of research time
- In this course, you will do:
    - Data cleaning and transformation
    - Collecting/Constructing your own dataset
    - Visualization and exploratory analysis

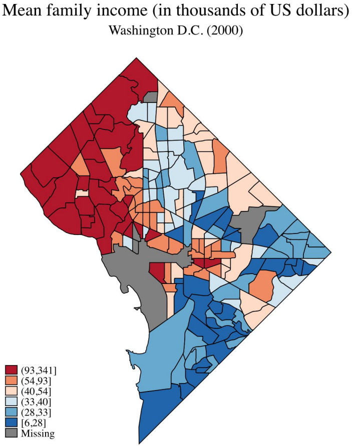
Data Sources

- Traditional sources
  - Government surveys
  - Official statistics
- Data revolution in the past decade
  - Large-scale administrative data
    - Near-universal population coverage
  - Examples in Taiwan
    - Health insurance claims
    - Tax return data
    - Housing transaction records

Unstructured and High-Dimensional Data

- Increasing use of non-traditional data formats
  - Text documents
  - Social media
  - Geolocation data
  - Satellite and image data

Geographic data



Source: Maurizio Pisati (2012)

Generative AI in Data Processing

- Generative AI as a coding assistant
  - Writing and debugging code on demand
  - Translating data tasks described in plain language into working scripts
  - Accelerating iteration when exploring data or testing specifications
- AI agents for semi-autonomous data workflows
  - Executing multi-step, repetitive operations across files or datasets
  - Parsing and restructuring unstructured or irregularly formatted raw data

Course Structure

1. Focus on how to implement various empirical methods of drawing causal inference
2. Discuss the applications in labor economics
3. Learn how to use statistical software and generative AI tools to conduct data analysis/empirical research

Applications in Labor Economics

- AI Impact on Labor Market
- 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:
  - **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
  - **Labor Economics** by Pierre Cahuc, Stephane Carcillo and Andre Zylberberg (graduate level)
  - **Labor Economics** by George Borjas (undergraduate level)

Reading Materials

- Suggested Readings:
  - **Korinek (2023), Generative AI for Economic Research**
    - JEL article on how LLMs assist economists across research tasks
  - **Korinek (2025), AI Agents for Economic Research**
    - Update on AI agent frameworks and autonomous workflows

Course 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 & Two Oral Exams (30%)
  - Each homework is followed by an individual oral exam
  - The oral exam will assess your understanding of your own code and empirical design
- Reading group presentation (15%)
- Term paper presentation (15%)
- Term paper (40%): milestones throughout the semester

Course Requirements

- You should use **Latex** to type your term paper in Chinese or English
  - **Latex** is a tool for typesetting professional-looking documents
- You can use "homework" to practice the above "requirements"

Important Dates

- Homework 1: 4/12
  - Oral Exam 1: 4/22
- Homework 2: 5/17
  - Oral Exam 2: 5/27 Week
- Reading group presentation: 5/28 and 6/4
- Term paper presentation: 6/11
- Term paper deadline: 6/18

Oral Exams

4/22 week and 5/27 Week

- Each oral exam will last about 20–30 minutes
- You will be asked to explain:
  - The code you wrote in your homework
  - The empirical question you want to study
  - The regression equation you plan to estimate
  - Why your empirical specification makes sense
- The goal is to assess your understanding of:
  - Empirical methods you learn in this courses
  - Statistical programming languages: R/Stata

Oral Exams

4/22 week and 5/27 Week

- After the oral exam, we will discuss your term paper progress
  - Discuss/Refine your research question
  - Clarify identification strategy
  - Discuss possible datasets and regression specifications
- You can prepare in advance
  - All questions are based on your own submitted work
  - If you understand your code and empirical design clearly, the oral exam will be straightforward
- This format reflects the rise of generative AI tools and emphasizes genuine understanding

Reading group presentation

5/27 and 6/3

- 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
6/10

- 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

6/17

- Feel free to discuss your term paper with me before the deadline

Guideline for Writing a Term Paper

Guideline for Writing a Term Paper

- You should start early; the paper is due on 6/17
- Short paper style: less than 3,000 words
  - See **Economics Letters**
  - See **AER: Insights**
- Typed, double-spaced, using one-inch margins and 12-point font

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