

定量研究方法II：因果推断 POL8013 2024年春

# 第1讲：课程导论

上海交通大学 国务学院 国际关系系 助理教授

付舒

2023年2月23日



# 2021年诺贝尔经济学奖



Ill. Niklas Elmehed © Nobel Prize Outreach.

**David Card**

Prize share: 1/2



Ill. Niklas Elmehed © Nobel Prize Outreach.

**Joshua D. Angrist**

Prize share: 1/4



Ill. Niklas Elmehed © Nobel Prize Outreach.

**Guido W. Imbens**

Prize share: 1/4



# 为什么要学习因果推断？

- 近年来，因果推断（Causal Inference）在政治学领域大行其道
- 可信度革命（Credibility Revolution），方兴未艾，如火如荼
- 由于进行随机实验的难度较大，采用观察性数据 (observational data) 进行因果推断已是社会科学研究的主流
- 研究发表、找工作等现实需要



# 今天

- 自我介绍 Self-Introductions
- 课程大纲 Syllabus
- 先导：因果推断 Causal Inference 100





# 自我介绍



# 自我介绍：付舒

## ■ 工作经历

- 2023 ~ 上海交通大学 国务学院国际关系系 助理教授
- 2021 ~ 2023 芝加哥大学 政治学系与本科生院 讲师

## ■ 教育背景

- 2015 ~ 2021 芝加哥大学 政治学科系 博士 美国政治 & 量化方法
- 2008 ~ 2015 清华大学 国际关系学系 本科、硕士



# 研究兴趣

## ■ 美国总统与国会

- The Filibuster and Legislative Discussion *Journal of Politics* (Fu & Howell 2023)
- Particularism or Policy *Political Research Quarterly* (Fu 2023)
- Direct Appeals of First Ladies *Presidential Studies Quarterly* (Fu & Savel 2020)
- Interbranch Messaging *Presidential Studies Quarterly* (Fu & Howell 2020)
- U.S. Ambassadors and Home-State Trade (Kim & Fu)
- Moderates on Capitol Hill (Fu)

## ■ 美国选举

- Do Primaries Exacerbate Congressional Polarization? (Fowler & Fu)
- 谁在支持特朗普？实证分析美国民粹主义（付舒）
- 问责与选拔：美国选举政治的逻辑谬误（付舒、罗兆天）

## ■ 中美关系

- 美国涉台舆论分析 《世界经济与政治》（张传杰、付舒 2012）



# 方法训练

## ■ 博弈论模型

## ■ 定量方法

- 数学与统计            Math and Statistics
- 计算社会科学        Computational Social Sciences
- 回归分析              Linear Models
- 最大似然估计        MLE/Model Based Inference
- 因果推断              Causal Inference (Hong + Grimmer + Fowler)
- 量化文本分析        Text-as-Data

## ■ 统计软件

- R, Stata, Python



# 引导者

- INFJ

- 我对教学很有热忱!

- 我的教学理念是做一个引导者，以学生为中心（student-centered），为大家做好的研究提供帮助。

- 讲座 + 讨论

- 主动学习

- 16周之后，真正有所收获（方法、学习方式、兴趣...）

- 唯一的要求：专注!

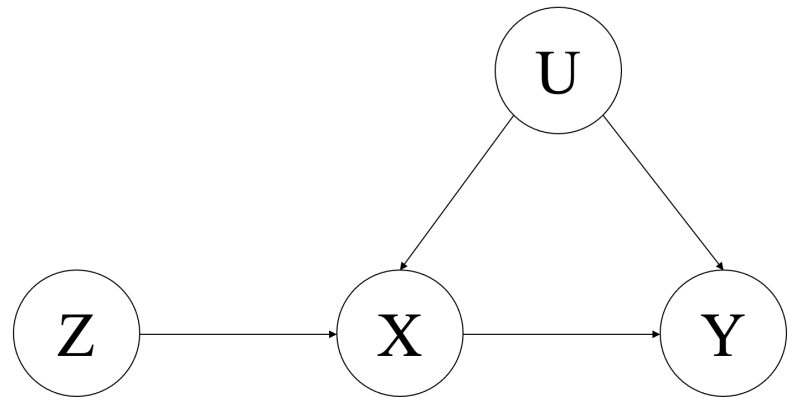


# 破冰：互相认识一下





# 课程大纲



# 定量研究方法II：因果推断 POL8013

- 课程性质：博士/硕士 专业前沿课
- 学时学分：2学分 共16周
- 开课时间：周五3-4节 (10:00-11:40)
- 开课地点：徐汇 工程馆211
  
- 电子邮件：fushu@sjtu.edu.cn
- 办公地点：徐汇 包图112
- 答疑时间：周五 15:00~17:00 (包图103)





# 课程目标

## 基础目标

- 帮助博士研究生掌握可应用于自己研究领域的前沿量化方法
- 掌握常用的因果推断方法，并达到**应用**程度

## 高阶目标

- 掌握因果推断的**统计**表达，明晰什么是因果效应，理解各因果识别策略的假定
- 能够**评估**研究设计和量化方法



# 课程特色

- “祛魅”：因果推断并不神秘，消除对量化方法的复杂迷思
- “手把手”和“做中学”

课程结构：Lecture + Lab + Problem Sets

- 讲座：讲授统计知识和方法概念
  - 实践：讨论方法的在具体研究中如何应用，并在电脑上进行实际操作
  - 作业：加深学生对方法的理解与应用熟练程度
- 统计理论 + 研究应用
  - 国际接轨、国内少有



# 先修要求

- 定量研究方法I: 回归分析
- 概率论与数理统计、线性回归、熟悉Stata
- 对于不熟悉的内容, 有意愿**努力**学会
- 走两步?

-  $\beta =$

-  $Var(X + Y) =$



# 先修要求

- 定量研究方法I: 回归分析
- 概率论与数理统计、线性回归、熟悉Stata
- 对于不熟悉的内容, 有意愿**努力**学会
- 走两步?
  - $\beta = (X'X)^{-1}X'Y$
  - $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X + Y)$



# 软件与阅读材料

- 课程将使用Stata作为课程统计软件
- Angrist and Pischke 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.
- 陈强, 《计量经济学及Stata应用》, 第二版, 高等教育出版社, 2023年12月

(所有阅读资料均可在Canvas中获取)



# Canvas

## 单元

- 进度安排
- 阅读材料
- 讲座PPT (每节课后上传)

## 讨论

- 鼓励学生使用讨论版块，讨论课程内容，分享讨论内容

## 作业

- 所有作业的布置和提交

### ▾ 课程导论 Course Introduction

📎 Mostly Harmless Econometrics.pdf

### ▾ 实验 Experimental Ideals

第2周 (3.1) : 潜在结果框架 Potential Outcome Framework

第3周 (3.8) : 随机实验 Randomized Experiments

第4周 (3.15) : 课堂实践 Lab Session

📎 GerberGreenLarimer2008APSR.pdf

📎 Olken2007JPE.pdf

### ▸ 基于观测的选择 Selection on Observables

### ▸ 因果识别策略 Identification Strategy #1

### ▸ 因果识别策略 Identification Strategy #2

### ▸ 因果识别策略 Identification Strategy #3

### ▸ 因果推断前沿



# 考核方式

- 没有考试！ 没有论文！ 没有presentations！



# 考核方式

## 1. 出勤 (10%)

- 如有生病请假, 理解, 提前邮件告知

## 2. 讨论 (15%) :

- 实践课前, 完成应用文章的阅读
- 鼓励学生积极在课上参与讨论和互动

## 3. 作业 (75%) :

- 一学期共5次作业
- 作业旨在巩固统计知识的理解 (stats) , 熟练方法的应用 (replications)





# 课堂礼仪

- 别迟到
- 随时可以打断，发言不用起立
- Old Fashioned: 买个笔记本，把阅读打印出来



# 教学内容进度安排

## 第一部分 实验 The Experimental Ideals

- 潜在结果框架、随机实验

## 第二部分 基于观测的选择 Selection on Observables

- 匹配方法、回归

## 第三部分 因果识别策略 Identification Strategies

- 双重差分、面板数据
- 断点回归
- 工具变量

## 第四部分 前沿研究

- 安慰剂
- 文本分析与因果推断

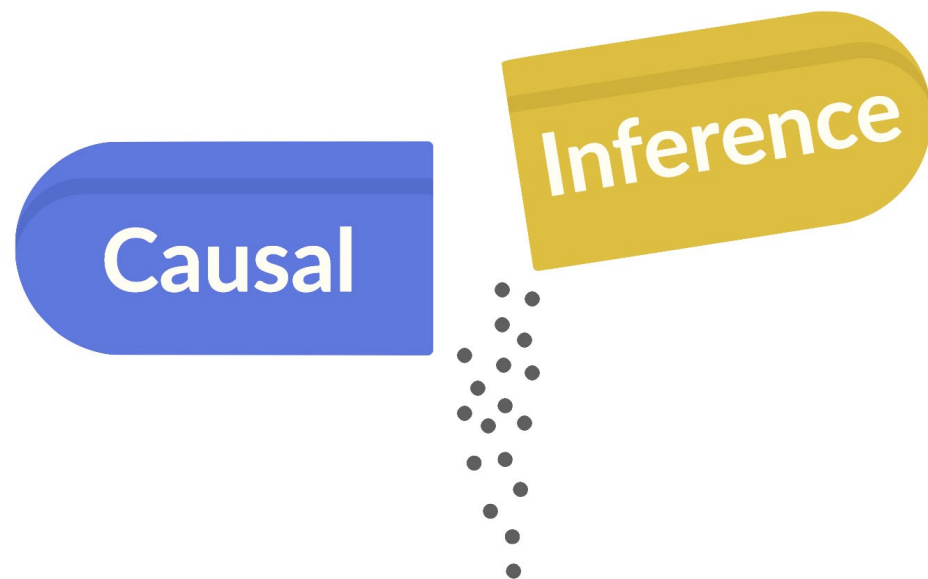
(详见课程大纲)



疑问?



# 因果推断 先导



# 因果推断

统计学 (Statistics) 有很多用途:

- 发现 Discovery
- 测量 Measurement
- 因果推断 Causal Inference
  - 在统计学中属于较新的领域
  - 跨学科, 发展迅速



# 两种因果关系

## 1. Effects of Causes

- 正向因果问题 (forward causal questions)
- 已知干预变量D, 问D的效应是什么?
- 如, 经济形势不好, 特朗普没能连任



## 2. Causes of effects

- 反向因果问题 (reverse causal inference)
- 已知结果变量Y, 问Y由什么导致?
- 如, 特朗普没能连任, 因为什么?



# 两种因果关系

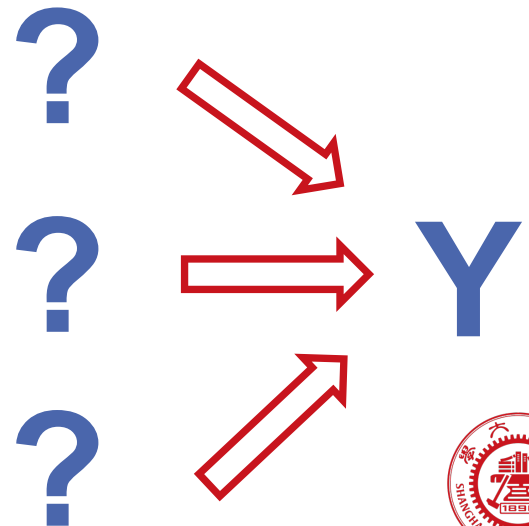
## 1. Effects of Causes

- 正向因果问题 (forward causal questions)
- 已知干预变量D, 问D的效应是什么?
- 如, 经济形势不好, 特朗普没能连任



## 2. Causes of effects

- 反向因果问题 (reverse causal inference)
- 已知结果变量Y, 问Y由什么导致?
- 如, 特朗普没能连任, 因为什么?



# 回归的局限

$$Y = \alpha + \tau D + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon$$

回归分析的关键局限在哪里？

- 内生性和遗漏变量偏差 Endogeneity and omitted variable bias
- 函数形式的误设 Misspecified functional form
- 效应的异质性 Heterogeneous Treatment Effect



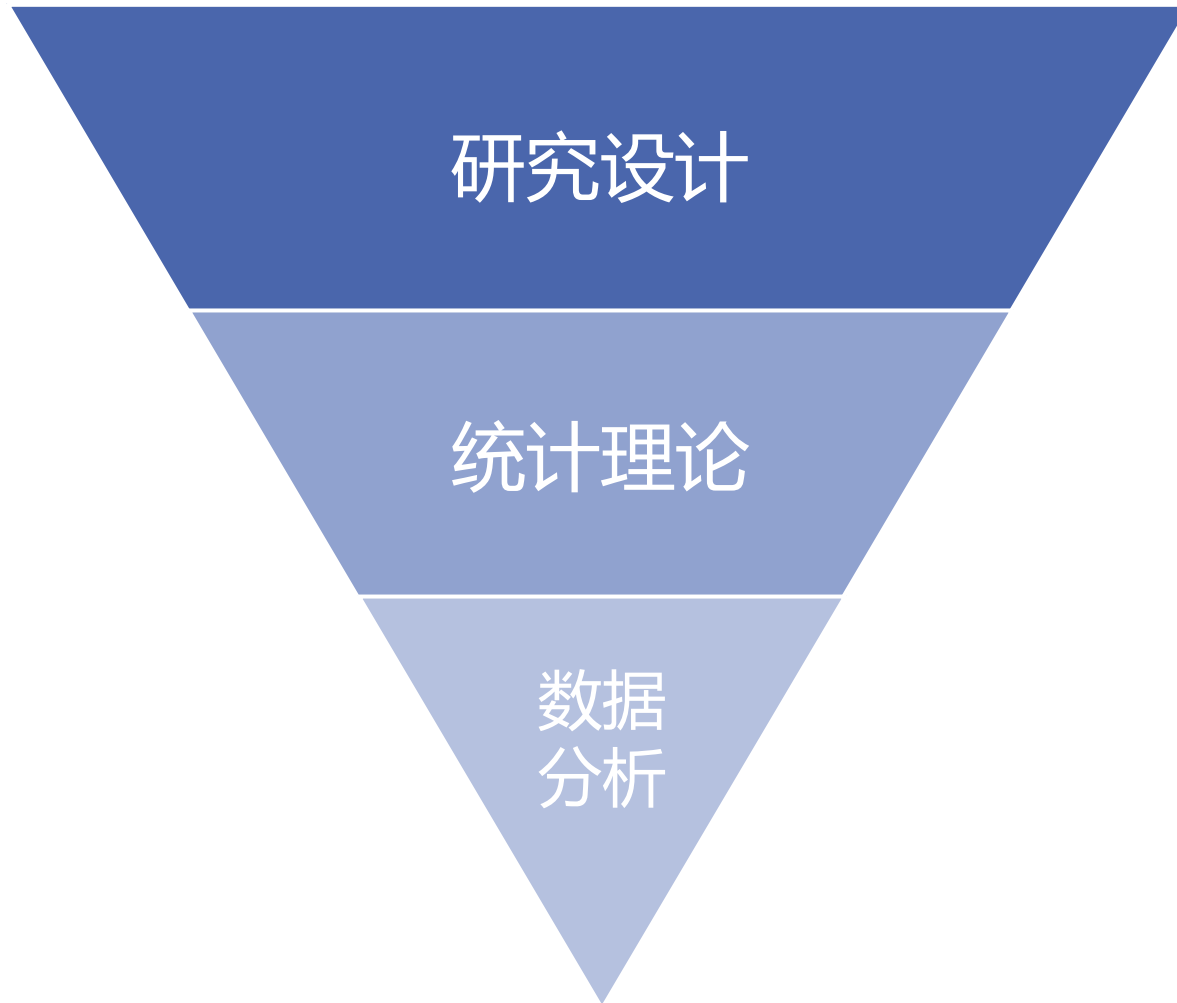


# 因果识别 Causal Identification

- 识别 (Identification) 指的是从有限的观察中获取结论
- **因果识别**, 指通过一系列的假定和估计, 从数据中获得对与因果效应的结论。
- Causal identification describes the combination of assumptions, estimation approach, and data necessary to reach a conclusion about a causal effect. (Manski 1995)
- 假定 + 数据 → 因果结论  
assumptions + data → causal conclusions



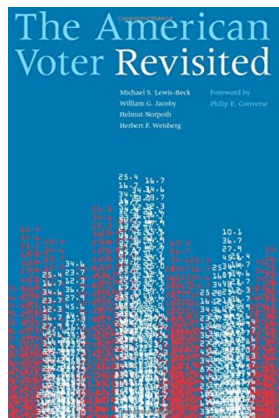
# 如何做因果推断？



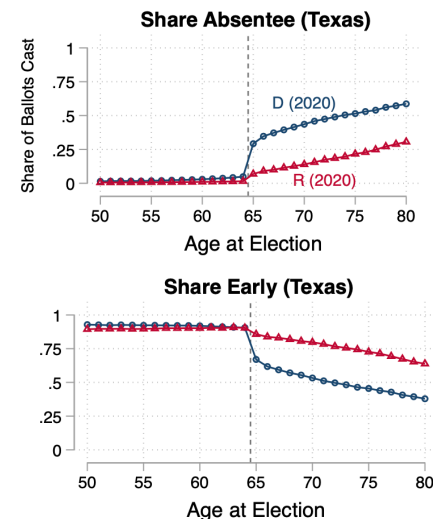
# 可信度革命 Credibility Revolution

- 因果推断与可信度革命，如火如荼，方兴未艾
- 简单的相关性，不代表因果性
- Increasing awareness across many social science fields that naïve observational studies can provide seriously misleading answers to questions about causality.

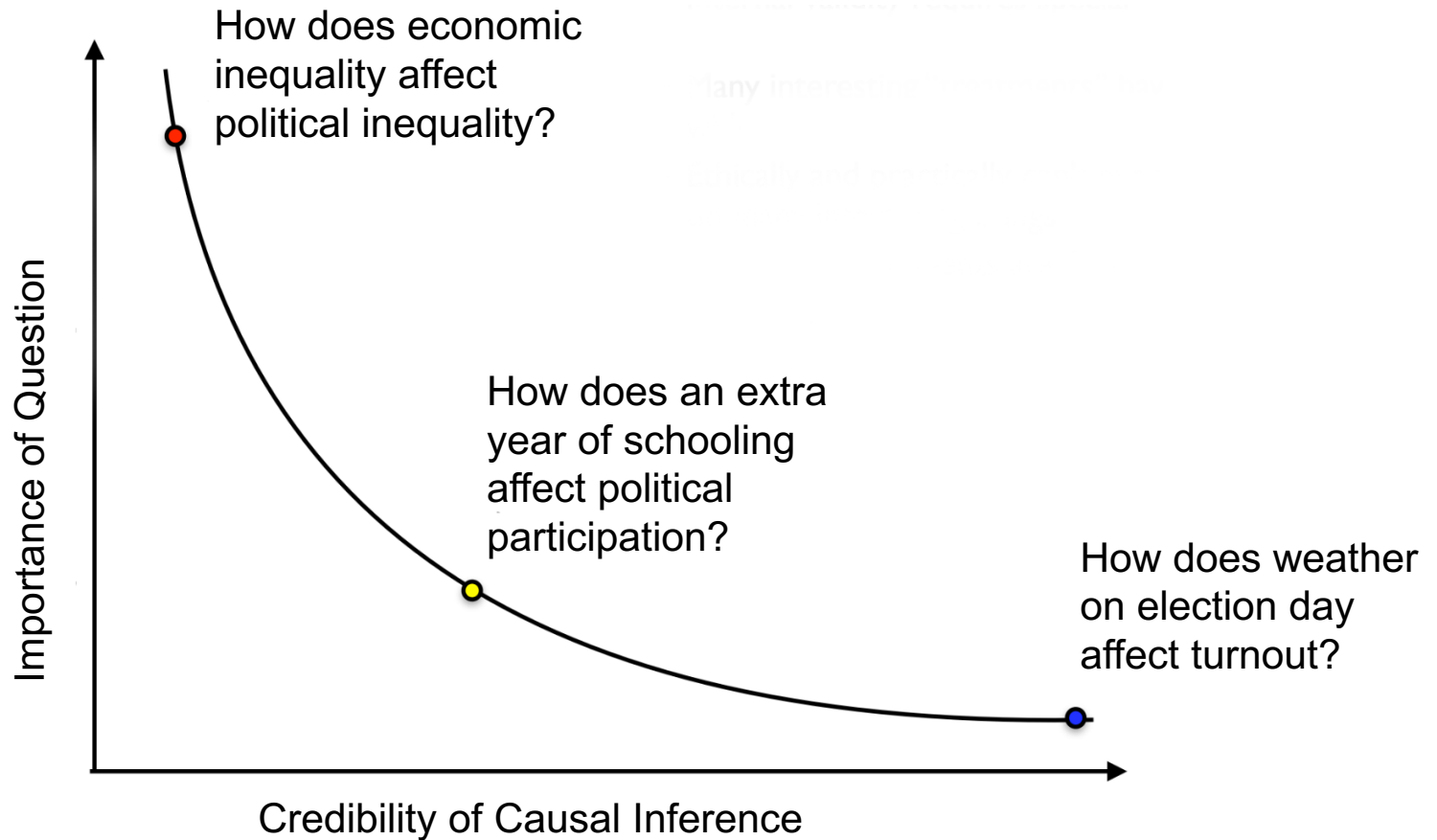
过去：问题大，多元回归



现在：问题细，实验或自然实验



# 可信度与重要性的取舍



# 下一节课：潜在结果框架



Jerzy Neyman



Donald Rubin





# 谢谢!



上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY

国际与公共事务学院  
SCHOOL OF INTERNATIONAL AND PUBLIC AFFAIRS