1. Potential Outcomes

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

1/ Course Overview

Why learn causal inference?

- Social science theories are almost always **causal** in nature.
- We should understand when our methods can have a causal interpretation.
- Charles Manski (Northwestern Econ.): “data + assumptions = conclusions”
  - Causal inference is about making assumptions and conclusions more **transparent!**
- The old way was “kitchen sink” regression + causal weasel words:
  - “associated with”, “leads to”, “the [causal?] effect of”, “[in—decreases]”, “more likely”, “encourages”, “is linked to”, “predicts”
- **Causal (credibility) revolution**: pick 1) a causal estimand and 2) a research design to identify it.

Outline of Topics to be Covered

- Applied Econometrics
    - Regression with Panel Data
    - Regression with a Binary Dependent Variable
    - Instrumental Variables Regression
    - Experiments and Quasi-Experiments
- Assessing Treatment Effects
    - Linear regression
    - Matching
    - Instrumental variables
    - Difference-in-differences
    - Regression discontinuity
- We may also discuss recent advancements
    - Double ML / Meta-Learners
    - Synthetic DiD
    - Causal mechanisms

About Me

Figure 1: Illustration of Counterfactuals in CI vs. XAI

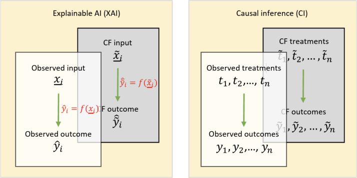


Table 1: Overview of differences in use of counterfactuals

Dimension	**Causal Inference**	**Explainable AI**
Purpose	Estimate causal effect	Explain a predicted outcome
Type of Relationship	Causal	Correlation
Quantity of Interest	Difference in outcomes	Input changes
Aggregation Level	Sample	Unit
Modified Object	Model	Data
Assumptions	On data generating function	On given predictive model
Performance Evaluation	Standard error, conf. interval, significance, robustness tests	Feasibility, sparseness, etc.

Teaching Assistant (TBA)

TBA

Fifi Ding | Justin Kao

Learning Resources

- Lecture: general theoretical and practical issues.
- Round-table discussions and paper presentations.
- Course homepage (GitHub Pages): syllabus, schedule, lecture slides, and readings.
- MS Teams: announcements, Q&A, assignment submissions, and DMs for help/study groups.
- Office Hours: by appointment via Google Calendar (link on homepage).

Textbooks

- Responsibility = material covered in lectures.
- Good books that I'll draw upon:
  - Imbens & Rubin: fairly technical, but covers basics well.
  - Hernan & Robins: slightly less technical, more biostat influence.
  - Angrist & Pischke: universal classic, opinionated, most readable.
  - Morgan & Winship: good combo of potential outcomes and graphs.
- Also check out:
  - The Effect by Nick Huntington-Klein
  - Causal Inference: The Mixtape by Scott Cunningham
  - Mastering Metrics by Angrist & Pischke
  - The Book of Why by Judea Pearl
  - Causal ML by Chernozhukov et al.

Course Information

For the up-to-date syllabus, grading policy, schedule, and readings, please visit the course homepage:

ISS5096 · Spring 2026 | Overview | Schedule | Materials | Evaluation | Policies | ← Back

Experiments and Causal Inference

A graduate seminar on experimental and quasi-experimental designs for causal inference.

ISS5096 | Spring 2026 | TSMC Bldg. R406 | Thursday, 14:20–17:20

ECI — Experiments & Causal Inference, Spring 2026

Syllabus PDF

Course Overview

This course introduces experimental and quasi-experimental methods for causal inference that are widely used in a broad array of domains such as marketing and information systems. The focus is on delivering a breadth of substantive topics and methodological considerations that emerge in utilizing identification-oriented methods.

Throughout the course, we will discuss topics related to methods such as randomized controlled trials (RCT), difference-in-differences (DiD), matching methods such as propensity score matching (PSM) and coarsened exact matching (CEM), and more advanced topics such as regression discontinuity designs (RDD), double-debiased machine learning (DML), synthetic control methods (SCM), and synthetic difference-in-differences.

Students will review relevant research papers on each topic and actively engage in presentations and discussions about the nature of causation and alternative means of

Course Details

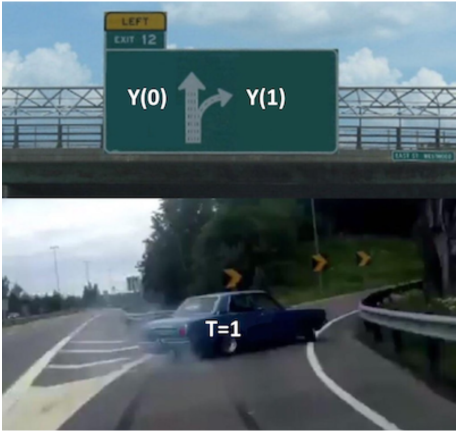
- Feb 23 – Jun 15, 2026
- Thursday
- 14:20–17:20
- TSMC Bldg. R406

Communication

All course communication goes through MS Teams. Check the Teams channel regularly for announcements, discussion threads, and updates. For private matters, email the instructor.

https://j1yoo.github.io/teaching/eci/

2/ Potential Outcomes Framework



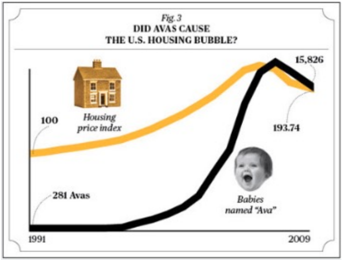
Source: Causal Inference for the Brave and True By Matheus Facure Alves

What is causal inference?

Factual vs. **Counterfactual**

- Q: Does having girls affect a judge's rulings in court?
  - A judge with a daughter gave a pro-choice ruling.
    (p.s., pro-choice: belief that everyone has a right to choose when & whether or NOT to have children $\rightsquigarrow$ pro-abortion)
  - Would they have done that if had a son instead?
- Q: Does having a CEO with a hobby of flying airplanes lead to better innovation outcomes? (Sunder et al. 2017. JFE)
  - Innovation outcomes are higher for companies Led by Pilot CEOs.
  - Would the innovation outcomes have been the same w/o a pilot CEO?
- **Causal inference** is the study of these types of causal questions.

What isn't causal inference?



Source: https://www.bloomberg.com/news/articles/2011-12-01/correlation-or-causation

- Associations: parameters of the joint distribution of **observed data**.
  - Correlations, regression coefficients, odds ratios, etc.
  - Describes the world as it happened.
  - No meaningful "directionality," just a joint distribution.
- But causal questions are about **unobserved data**: counterfactuals!
  - Describes what would happen if we **changed** the world.
  - The backbone of most social science theorizing.

The observational-causal bridge

- Causal inference = missing data problem.
- **Assumptions** connect missing data to observed data.
  - `Present Jaewon` stays up until 3 am prepping for class.
  - How would `Present Jaewon` have felt if he had gone to bed at 10 pm?
  - `Past Jaewon` (w/ a 10pm bedtime) a good substitute? (Assumption!)
- How do we make assumptions crystal clear? $\rightsquigarrow$ `causal notation!`
  - Special notation for counterfactuals and interventions.
  - Precisely state what data helps us learn about counterfactuals.

Motivation: Study of Political Canvassing

- Study of n voters.
  - *Canvassing?*: A systematic initiation of direct contact with individuals, commonly used during political campaigns (think of it as political advertising!)
  - n_1 are canvassed.
  - $n_0 = n - n_1$ are not canvassed.
- For each voter $i \in \{1, 2, \ldots, n\}$ we observe:
  - **Observed outcome** (turnout): Y_i
    - *Turnout?:* The percentage of eligible voters who participated in an election.
  - **Treatment variable**:

$$D_i = \begin{cases} 1 & \text{if treated (canvassed)} \\ 0 & \text{if not treated (not canvassed)} \end{cases}$$

  - **Pretreatment covariates**: X_i
- Causal question of interest: does contact/canvassing affect turnout?

Defining causal effects

- **Potential outcomes** formally encode counterfactuals (Neyman-Rubin)
  - $Y_i(1)$ outcome that unit i would have if treated.
  - $Y_i(0)$ outcome that unit i would have if untreated.
- Connect observed outcomes to potential outcomes (**consistency**).
  - $Y_i = Y_i(D_i)$: we observe the potential outcome of observed treatment.
- **Causal effect** for unit i: $\tau_i = Y_i(1) - Y_i(0)$

Voters	Age	Gender	Contact	Turnout		Casual effect
i	X_{i1}	X_{i2}	D_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	25	M	1	0	???	
2	38	F	0	???	1	
3	67	F	0	???	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
n	43	M	1	1	???	

The fundamental problem of causal inference

- We only observe one potential outcome per unit.
  - $\rightsquigarrow Y_i(1) - Y_i(0)$ is never directly observed.
  - Can learn about the marginal distributions, not joint.
- Generalizes to non-binary treatments:
  - Categorical: $Y_i(d)$ for $d = 0, 1, \ldots, K-1$
  - Continuous (dose-response): $Y_i(d)$ for $d \in \mathbb{R}$
  - Multivariate: $Y_i(d_1, \ldots, d_K)$ for $d_k \in D_K$
- Again, causal inference is **missing data problem!**
  - How do we infer the missing potential outcomes?
    (stick around for the rest of the course)

Key assumptions for defining effects

1. **Causal ordering:** $D_i \to Y_i$
   - No reverse causality or simultaneity.
2. **Consistency:** $Y_i = Y_i(d)$ if $D_i = d$
   - Bridges the **factual** and **counterfactual** worlds: the observed outcome equals the potential outcome under the observed treatment.
   - For this to hold, $Y_i(d)$ must be **well-defined**:
     - No different "versions" of each treatment level (strict; Rubin, 1980).
     - Or that "treatment variance is **irrelevant**" (Vanderweele, 2009): variation in how treatment is delivered is permissible so long as it does not affect the potential outcomes differently.
3. **No interference between units:** $Y_i(D_1, D_2, \ldots, D_N) = Y_i(D_i)$
   - No causal effect of other units' treatment on other units' outcomes.

- Last two combined: **SUTVA** (stable unit-treatment variation assumption)

Manipulation

- $Y_i(d)$ is the value that Y would take under D_i set to d.
  - To be well-defined, D_i should be manipulable at least in principle.
- $\rightsquigarrow$ common motto: **"No causation without manipulation"** Holland (1986)
- Tricky causal problems: immutable characteristics such as race, sex, etc.
  - What is the effect of being a man on my political views?
  - What's the hypothetical manipulation? Very tricky!
- Common alternative: focus on places where we can manipulate these characteristics:
  - Effect of perceived race/gender on legislator replies to constituent mail.
  - Effect of elective female versus male legislators on policy outcomes.
  - Differential effects of treatment by race or gender.

Causal estimands

- Ideal world: estimate unit causal effects $Y_i(1) - Y_i(0)$
  - But... **almost always unidentified** without strong assumptions
- **Sample average treatment effect** (SATE):

$$\text{SATT} = \frac{1}{n}\sum_{i=1}^{n}[Y_i(1) - Y_i(0)] \tag{1}$$

  - Average outcomes if everyone is treated vs. no one.
  - We'll spend a lot time trying to identify this.
- **Sample average treatment effect on the treated** (SATT):

$$\text{SATT} = \frac{1}{n_1}\sum_{i=1}^{n} D_i(Y_i(1) - Y_i(0)) = \frac{1}{n_1}\sum_{i=1}^{n} D_i(Y_i - Y_i(0)) \tag{2}$$

  - We will be looking at this when we have noncompliance issues.

Samples versus Populations

- SATE and SATT are specific to a particular study $i = 1, \ldots, n$.
  - Called **finite-sample** or **finite population** inference.
- What if there is a larger population we would like to target?
  - Assume units are a random sample from a large/infinite population.
  - Called the **superpopulation** or sometimes just **population** inference.
- **Population average treatment effects:**

$$\text{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)] \tag{3}$$

$$\text{PATT} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1] \tag{4}$$

Other causal estimands

- Conditional average treatment effect (CATE):

$$\mathbb{E}[Y_i(1) - Y_i(0)|\mathbf{X_i} = x] \tag{5}$$

  - Useful detecting heterogeneous effects for theory testing or targeting.

- Multiple treatments:
  - Controlled direct effect: $\mathbb{E}[Y_i(1, d_2) - Y_i(0, d_2)]$
  - Subtle but important differences from CATE!

- Non-additive effects:
  - **Quantile treatment effects:**
    - Example: $median(Y_i(1)) - median(Y_i(0))$
    - How does treated shift a particular quantile of the outcome distribution?
  - **Odds-ratio:**

$$\frac{\mathbb{P}[Y_i(1) = 1]/\mathbb{P}[Y_i(1) = 0]}{\mathbb{P}[Y_i(0) = 1]/\mathbb{P}[Y_i(0) = 0]} \tag{6}$$

More complicated setup: Truncation by death

- Set up:
  - Units: patients
  - Treatment: new medicine
  - Outcome: cholesterol level
  - Truncation: patient death

- Truncation by "death" problem (Zhang and Rubin 2003, J. Educ. Behav. Stat.):
  - Cholesterol level is undefined for the dead.
    - This is not a missing data problem in the usual sense, the outcome simply does not exist for some units.
  - **Issue**: treatment may also affect survival.
    - $\rightsquigarrow$ If the treatment saves the lives of the people with high cholesterol, it may appear that the treatment increases cholesterol!

Another Truncation Problem

- RQ: effect of a **job training program** D_i on **wages** Y_i
- Truncation by "death" problem:
  - Wages can be observed only for those that are **employed**.
- **Issue?:** program (D_i) might also affect employment status (S_i).
  - If the program helps unemployed people get jobs, the newly employed may have lower wages, making it appear that the program decreases wages!
- This is called post-treatment bias: conditioning on a variable (employment/survival) that is itself affected by the treatment.

Principal stratification (Frangakis and Rubin, 2002. Biometrics)

- Q: How can we think about the causal effect of D_i on Y_i under the truncation by death problem?
  - We only observe Y_i when $S_i = 1$ (i.e., employed).
- Potential variables:
  - Potential employment: $S_i(1)$, $S_i(0)$
  - Potential wages: $Y_i(d, s) \longrightarrow Y_i(1, 0)$; $Y_i(0, 0)$ do not exist.
- Four **principal strata** defined by $(S_i(0), S_i(1))$:
  1. $(1, 1)$: always employed (regardless of program).
  2. $(0, 0)$: never employed (regardless of program).
  3. $(0, 1)$: helped (employed only when treated).
  4. $(1, 0)$: hurt (unemployed only when treated).
- Causal effect is defined only for always employed:

$$\mathbb{E}[Y_i(1, 1) - Y_i(0, 1) | S_i(1) = S_i(0) = 1] \tag{7}$$

- Can't tell which principal stratum each unit belongs to. Why?

Takeaways

1. Causal inference is about comparing **counterfactuals**.
2. Potential outcome (PO) represents these counterfactuals mathematically.
    - $\rightsquigarrow$ Allows us to identify (then estimate) causal estimands of interests!
3. Many, many possible **causal** quantities of interest (any contrast of POs).

3/ In-Class Exercise: Samsung ESG Fund

The Case: Samsung's SME ESG Support Fund

Real program. In September 2024, Samsung Electronics, Samsung Display, and Korea's five major banking groups launched a **1 trillion KRW ($750M)** green-growth fund for SME suppliers.

Key Facts

- **Units:** SME suppliers in Samsung's supply chain.
- **Intervention:** Interest-free loans up to 2B KRW per firm, 3-year term; technical support for ESG compliance (e.g., EU's CSRD).
- **ESG requirements:** Firms must demonstrate contributions to one of six environmental goals and meet human rights, labour, safety, and anti-corruption standards.

The Case: Samsung's SME ESG Support Fund (cont'd)

What happened?

- **One year later:** Only 2% of the fund (18.8B KRW) was disbursed as most SMEs could not meet the eligibility criteria.
- Mostly limited to first-tier suppliers that already had ESG capacity.

Your task: Apply the potential outcomes framework from today's lecture to think critically about how we would evaluate this program.

Sources: "Samsung, top banks to create W2tr green growth fund" (*Korea Herald*, 2024); "Samsung launches 1T KRW ESG fund" (*JoongAng Daily*, 2024).

Question 1: Defining Potential Outcomes

Suppose you are a VP at Samsung and need to evaluate this fund. Using today's notation, formally define the **treatment** (D), a suitable **outcome** (Y), and the **potential outcomes** for this case.

Guide:

- $D_i = 1$ if SME i receives the fund; $D_i = 0$ if SME i does not.
- $Y_i(1)$: outcome if SME i **gets** support.
- $Y_i(0)$: outcome if SME i **does not** get support.
- Individual causal effect: $\tau_i = Y_i(1) - Y_i(0)$.

Discussion: If you had to pick one outcome measure for Y—ESG score, carbon emissions, or something else—which would you choose and why?

Hint: Think about measurability, timing, and what Samsung would actually care about.

Question 2: SUTVA Violations

Read the two scenarios below. For each one, decide: which component of SUTVA does it violate, and in which **direction** would it bias our estimate of the fund's effect?

Scenario A — Supply-chain spillover:

- A funded SME A starts requiring its own sub-contractors (unfunded SME B) to meet higher ESG standards. B's ESG outcome improves even though B never received the fund.

Scenario B — Loans vs. consulting:

- The fund offers interest-free loans to some SMEs and technical consulting to others. A firm borrowing 2B KRW at ≈4% saves ≈7M KRW/month; a firm receiving consulting gets a very different "dose."

Question 2: SUTVA Violations (cont'd)

Discussion:

- Can you think of a **third** scenario, either another interference channel or another hidden treatment version?
- If we ignore these violations and estimate the ATE anyway, do we **over-** or **underestimate** the true effect? Why?

Question 3: The Selection Problem

Samsung did **not** randomly assign the fund—firms must meet strict ESG criteria to qualify. After one year, only **2%** was disbursed, mostly to first-tier suppliers already ESG-capable. Why can't we simply compare funded vs. unfunded firms?

Guide:

- Firms with high baseline ESG capability self-select in $\Rightarrow$ **selection bias**.
- Formally: $\mathbb{E}[Y_i(0) \mid D_i = 1] \neq \mathbb{E}[Y_i(0) \mid D_i = 0]$.
- The 2% uptake rate means the "treated" group was already ESG-ready before the fund.

Preview: If Samsung uses an internal ESG score with a cutoff for eligibility, we could compare firms just below vs. just above the threshold $\rightsquigarrow$ **Regression Discontinuity Design (RDD)**, coming later this semester.

Question 3 (cont'd): Can Firms Game the Cutoff?

A quick thought experiment on the RDD idea:

> If firms know the ESG-score cutoff exists, do they have an incentive to inflate their scores just above it? If so, what happens to our RDD strategy?

Discussion:

- Which types of firms are most likely to game—small firms near the margin, or large first-tier suppliers?
- Does gaming invalidate RDD entirely, or can we still salvage something?

We won't answer this formally today. When we reach the RDD chapter, we'll learn how to test for this (McCrary density test) and why it matters.

Question 4: External Validity (SATE vs. PATE)

Suppose we cleanly estimate the causal effect of Samsung's ESG fund on its own suppliers. Can we generalize this result to SME suppliers of other conglomerates (Hyundai, SK, LG, ...)?

Discussion:

- Are Samsung's suppliers "representative" of all Korean SMEs? Why or why not?
- What would need to be true for the effect to **generalize**?
- How does this relate to the **SATE** vs. **PATE** distinction from today's lecture?

Key takeaway: Identification (internal validity) and generalization (external validity) are separate problems.

Discussion Format

1. **Q1**: define D, Y, potential outcomes; discuss outcome choice.
2. **Q2**: identify SUTVA violations and bias direction; brainstorm a third scenario.
3. **Q3–Q4**: selection bias, cutoff gaming, and whether results generalize beyond Samsung.

Grading: This is a participation-based exercise. The goal is to practice applying the framework to a real world setting, not to arrive at a single correct answer.

Have a Great Weekend! :)

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo.github.io/