Causal Inference: Part I

Sukjin Han

University of Bristol

June 2024

Center for Data Science, Zhejiang University

Causal Inference: Roadmap for Part I

1. Counterfactual framework
2. Relationship to structural models
3. Treatment effects, issue of heterogeneity
4. Treatment parameters of interest
5. Selection problem / sorting problem
6. Overview of possible approaches

Causal Questions

- Examples of questions in causal inference:
  1. Labor economics: University premium, industry wage gap
  2. Public finance: Impacts of health care expenditures and health insurance on health
  3. Education: If school choice, education reform, and school inputs boost learning
  4. Macroeconomics: If expansionary monetary policy revive a troubling economy
  5. Industrial organization: If a monopolist's price increase lower demand
  6. Environmental economics: If firms' green technology adoption reduce local pollution
- In causal inference, we want to know the mechanisms behind

Counterfactual Framework

- D_i: treatment dummy variable for individual i
  - $D_i = 1$ if treated, $= 0$ otherwise
- Y_{1i}: counterfactual outcome for i if treated
  - i.e., what would have been observed if treated
- Y_{0i}: counterfactual outcome for i if not treated
  - i.e., what would have been observed if not treated
- Y_i: observed outcome for i
  - That is,

$$Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i}) = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

- Implicit in the notation:
  - No interaction across units (i.e., no GE effects or peer effects)
- X_i: observed control variables, directly affect Y_{0i} and Y_{1i}

Example: College Premium

- D_i: college education of individual i
  - $D_i = 1$ if received college degree, $= 0$ if not
- Y_{1i}: potential wage of i if worked with college degree
- Y_{0i}: potential wage of i if worked without college degree
- Y_i: observed wage of i
- X_i: characteristics in standard wage equation (e.g., age, gender, location, parental education)

Structural Models

- We now introduce structural models
- Counterfactual notation can be equivalently written with structural notation
  - Example 1: $Y_i = \beta_0 + \beta_1 D_i + X_i\gamma + U_i$

$$Y_{1i} = \beta_0 + \beta_1 + X_i\gamma + U_i$$
$$Y_{0i} = \beta_0 + X_i\gamma + U_i$$

  - Example 2: $Y_i = g(D_i, X_i, U_i)$

$$Y_{1i} = g(1, X_i, U_i)$$
$$Y_{0i} = g(0, X_i, U_i)$$

Counterfactual Framework vs. Structural Models

- There are philosophical differences between counterfactual outcome framework vs. structural models
  - "Effect of causes" (statistical solution) vs. "cause of effects" (scientific solution)

Counterfactual Framework vs. Structural Models

- There are philosophical differences between counterfactual outcome framework vs. structural models
  - "Effect of causes" (statistical solution) vs. "cause of effects" (scientific solution)
- Effect of causes:
  - All in black-box
  - Maybe enough in experimental setting (i.e., with randomization)
  - Hard to extrapolate
- Cause of effects:
  - Want to learn mechanisms behind
  - Use economic theory as guidance
  - Counterfactual analysis: Can forecast effects of treatments that never occurred before

Treatment Effects

- The treatment effect for individual i can be written as

$$Y_{1i} - Y_{0i}$$

- Fundamental challenge of causal inference:
  - Y_{1i} and Y_{0i} are not simultaneously observed
  - e.g., same individual's wages with and without college

Treatment Effects

- The treatment effect for individual i can be written as

$$Y_{1i} - Y_{0i}$$

- Fundamental challenge of causal inference:
  - Y_{1i} and Y_{0i} are not simultaneously observed
  - e.g., same individual's wages with and without college
- One solution:

$$E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$$

  - cf. $Q_\tau(Y_{1i} - Y_{0i}) \neq Q_\tau(Y_{1i}) - Q_\tau(Y_{0i})$
  - cf. Distributional treatment effects

Treatment Effect Heterogeneity

- Let

$$\Delta_i = Y_{1i} - Y_{0i}$$

- Q: How does Δ_i vary with i?

Treatment Effect Heterogeneity

- Let

$$\Delta_i = Y_{1i} - Y_{0i}$$

- Q: How does Δ_i vary with i?

1. Homogeneous treatment effect: $\Delta_i = \Delta$ (doesn't vary with i)
   - Example: $Y_i = \beta_0 + \beta_1 D_i + X_i\gamma + U_i$ then

$$Y_{1i} - Y_{0i} = \beta_1$$

   - Another example: $Y_i = \beta_0 + \beta_1 D_i + g(X_i) + U_i$

Treatment Effect Heterogeneity

2. Homogeneous treatment effect conditional on X_i: $\Delta_i = \Delta(X_i)$

   - That is, if $X_i = X_j$ then $Y_{1i} - Y_{0i} = Y_{1j} - Y_{0j}$ (i.e., individuals with same X have same effect)
   - Example: $Y_i = \beta_0 + \beta_1 D_i X_i + X_i \gamma + U_i$ then

$$Y_{1i} - Y_{0i} = \beta_1 X_i$$

   - Another example: $Y_i = g(D_i, X_i) + U_i$ then

$$Y_{1i} - Y_{0i} = g(1, X_i) - g(0, X_i)$$

Treatment Effect Heterogeneity

3. Heterogeneous treatment effect: Δ_i varies with i, even conditional on X_i
   - Example: $Y_i = \beta_0 + \beta_{1i} D_i + X_i \gamma + U_i$ then
     $$Y_{1i} - Y_{0i} = \beta_{1i}$$
   - Another example: $Y_i = g(D_i, X_i, U_i)$ then
     $$Y_{1i} - Y_{0i} = g(1, X_i, U_i) - g(0, X_i, U_i)$$

Selection Bias and Sorting Gain

- Two subcases of Case 3:
  - (a) $Y_{1i} - Y_{0i}$ is independent of D_i conditional on X_i
  - (b) $Y_{1i} - Y_{0i}$ is correlated with D_i conditional on X_i
- This distinction is different from the one involved in usual selection bias discussion
  - Selection bias: Y_{0i} is correlated of D_i even conditional on X_i
  - No selection bias: Y_{0i} is independent of D_i conditional on X_i
    - e.g., $Y_{0i} = \beta_0 + X_i\gamma + U_i$ with $E[U_i|D_i, X_i] = E[U_i|X_i]$
    - The usual conditional independence condition addresses selection bias
- (a) vs. (b): whether there is "essential heterogeneity" (and sorting on gain) or not
  - Case (b) is when individuals sort themselves based on gain (not only based on baseline outcome Y_0)
  - More later

Objects of Interest

- Homogeneous treatment effects (Cases 1 and 2):
  - Δ, $\Delta(X_i)$, or $E[\Delta(X_i)]$
- Heterogeneous treatment effects (Case 3(a)):
  - $E[\Delta_i]$, $E[\Delta_i|X_i]$ (or more)
- Heterogeneous treatment effects (Case 3(b)):
  - Not clear
  - e.g., local average treatment effect (LATE) (later)

Examples of Mean Treatment Parameters

- Average treatment effect (ATE): $E[Y_{1i} - Y_{0i}]$
- ATE on the treated (TT): $E[Y_{1i} - Y_{0i}|D_i = 1]$
- ATE on the un-treated (TUT): $E[Y_{1i} - Y_{0i}|D_i = 0]$
- ATE conditional on X_i: $E[Y_{1i} - Y_{0i}|X_i]$
- TT conditional on X_i: $E[Y_{1i} - Y_{0i}|D_i = 1, X_i]$
- TUT conditional on X_i: $E[Y_{1i} - Y_{0i}|D_i = 0, X_i]$

Heterogenous Treatment Effects

- Homogeneous treatment effects (Case 1):
  - $ATE = TT = TUT = ATE(X_i) = TT(X_i) = TUT(X_i)$
- Homogeneous treatment effects conditional on X_i (Case 2):
  - $ATE(X_i) = TT(X_i) = TUT(X_i)$ but possible that $ATE \neq TT \neq TUT$
- Heterogeneous treatment effects (Case 3(a)):
  - Same as Case 2
- Heterogeneous treatment effects (Case 3(b)):
  - $ATE \neq TT \neq TUT \neq ATE(X_i) \neq TT(X_i) \neq TUT(X_i)$

Evaluation Problems

- Homogeneous treatment effects (Case 1): Selection bias

$$
\begin{aligned}
& E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
& = E[Y_{1i} - Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
& = \Delta + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]
\end{aligned}
$$

  - $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$ is selection bias
  - e.g., individuals with higher "baseline" tend to attend college
  - Same in Case 2

Evaluation Problems

- Heterogeneous treatment effects (Case 3):

$$
\begin{aligned}
& E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
& = E[Y_{1i} - Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\
& = E[Y_{1i} - Y_{0i}] + E[Y_{1i} - Y_{0i}|D_i = 1] - E[Y_{1i} - Y_{0i}] \\
& + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]
\end{aligned}
$$

  - $E[Y_{1i} - Y_{0i}|D_i = 1] - E[Y_{1i} - Y_{0i}]$ is the sorting gain
  - e.g., individuals with higher college premium tend to attend college
  - Sorting gain is not zero in Case 3(b)

Overview of Possible Approaches

- How to recover some mean treatment parameters?
  1. Randomized experiment
  2. Matching / conditional independence assumption
  3. Difference-in-differences (DD)
  4. Regression discontinuity (RD)
  5. Instrumental variables (IV) methods
- These methods allow heterogeneous treatment effects
  - Which treatment parameter is recovered depends on the method
  - Sometime we use structural models (e.g., linear model) for each method
    - This means we impose more restrictions
    - Treatment effects may even be restricted to be homogeneous

Randomized Experiment

- When D_i is randomized, it satisfies $(Y_{1i}, Y_{0i}) \perp D_i$
  - e.g., random lottery for college (among eligible applicants)
- Then,

$$E[Y_{di}|D_i = d] = E[Y_{di}] \text{ for } d = 1, 0$$

- Random assignment eliminates selection bias and sorting gain:

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &= E[Y_{1i}] - E[Y_{0i}] \\ &= E[Y_{1i} - Y_{0i}] \end{aligned}$$

- Simple difference-in-mean estimator can be used:

$$\frac{\sum_{i=1}^{n} Y_i 1\{D_i = 1\}}{\sum_{i=1}^{n} 1\{D_i = 1\}} - \frac{\sum_{i=1}^{n} Y_i 1\{D_i = 0\}}{\sum_{i=1}^{n} 1\{D_i = 0\}}$$

Matching and Conditional Independence

- First, we consider the approach that imposes conditional independence assumption:

$$(Y_{1i}, Y_{0i}) \perp D_i | X_i$$

  - i.e., conditional on X_i (e.g., demographics, previous educations), we assume D_i (e.g., college) is as if randomized
  - Idea of matching: conditional on X_i, the two groups are balanced
  - This can be weaken to mean independence
- Another assumption needed: For any X_i,

$$0 < \Pr[D_i = 1 | X_i] < 1$$

  - Common support (or overlap) assumption
  - Related to "no multicollinearity" assumption

Matching and Conditional Independence

- Under these assumption,

$$
\begin{aligned}
& E[Y_i|D_i = 1, X_i] - E[Y_i|D_i = 0, X_i] \\
& = E[Y_{1i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 0, X_i] \\
& = E[Y_{1i}|X_i] - E[Y_{0i}|X_i] \\
& = E[Y_{1i} - Y_{0i}|X_i]
\end{aligned}
$$

- X_i can include many covariates, even continuous variables
  - May not be appealing in practice
- Surprising result:

$$(Y_{1i}, Y_{0i}) \perp D_i|X_i \Longleftrightarrow (Y_{1i}, Y_{0i}) \perp D_i|P(X_i)$$

  where $P(X_i) = \Pr[D_i = 1|X_i]$ is the propensity score
  - This is the idea of propensity score matching
  - As long as the propensity of receiving treatment is the same, the two groups are balanced

Matching and Conditional Independence

- That is,

$$E[Y_i|D_i = 1, P(X_i)] - E[Y_i|D_i = 0, P(X_i)]$$
$$= E[Y_{1i} - Y_{0i}|P(X_i)]$$

  - Again, the common support assumption is implicitly used
- Various estimators can be used
  - Regression-based estimator
  - Inverse probability weighting estimator
  - Matching estimator

Matching and Conditional Independence

- Much weaker independence assumption:

$$E[Y_{0i}|D_i = 1, X_i] = E[Y_{0i}|D_i = 0, X_i]$$

- Then,

$$\begin{aligned}
&E[Y_i|D_i = 1, X_i] - E[Y_i|D_i = 0, X_i] \\
&= E[Y_{1i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 0, X_i] \\
&= E[Y_{1i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 1, X_i] \\
&= E[Y_{1i} - Y_{0i}|D_i = 1, X_i]
\end{aligned}$$

Before-After Analysis (Event Studies)

- Suppose we observe individuals before/after treatment
  - e.g., before and after job training
- $D_i = 1$ if i receives treatment at given time
- Y_{it}: outcome in period t; Y_{1it} and Y_{0it} are potential outcomes
  - $t = b$ (before) or a (after)
  - Y_{ia}: outcome in period after the treatment ($Y_{ia} = Y_{1ia}$)
  - Y_{ib}: outcome in period before the treatment ($Y_{ib} = Y_{0ib}$)
- Assumption: $E[Y_{0ib}|D_i = 1] = E[Y_{0ia}|D_i = 1]$
- Then,

$$
\begin{aligned}
E[Y_{ia}|D_i = 1] - E[Y_{ib}|D_i = 1] &= E[Y_{1ia}|D_i = 1] - E[Y_{0ib}|D_i = 1] \\
&= E[Y_{1ia}|D_i = 1] - E[Y_{0ia}|D_i = 1] \\
&= E[Y_{1ia} - Y_{0ia}|D_i = 1]
\end{aligned}
$$

  - Treatment effect on the treated (after the treatment)

Difference-in-Differences

- Is the assumption above plausible?
  - e.g., time effects, age effects...
- Suppose we observe treated/untreated individuals, before/after treatment
- Common trend assumption:

$$E[Y_{0ia} - Y_{0ib}|D_i = 1] = E[Y_{0ia} - Y_{0ib}|D_i = 0]$$

  - e.g., "baseline" wage trends are same btw treatment and control groups
  - Let $\Delta Y_{0i} = Y_{0ia} - Y_{0ib}$, then this assumption (conditional on X_i) is conditional indep in terms of ΔY_{0i}

Difference-in-Differences

- Common trend assumption:

$$E[Y_{0ia} - Y_{0ib}|D_i = 1] = E[Y_{0ia} - Y_{0ib}|D_i = 0]$$

- Then,

$$\begin{aligned}
&E[Y_{ia} - Y_{ib}|D_i = 1] - E[Y_{ia} - Y_{ib}|D_i = 0] \\
&= E[Y_{1ia} - Y_{0ib}|D_i = 1] - E[Y_{0ia} - Y_{0ib}|D_i = 0] \\
&= E[Y_{1ia} - Y_{0ia}|D_i = 1] \\
&\quad + E[Y_{0ia} - Y_{0ib}|D_i = 1] - E[Y_{0ia} - Y_{0ib}|D_i = 0] \\
&= E[Y_{1ia} - Y_{0ia}|D_i = 1]
\end{aligned}$$

  - Treatment effect on the treated (after the treatment)

Regression Discontinuity

- Let R_i be the running variable
  - e.g., college test score or eligibility score
- Suppose

$$D_i = \begin{cases} 1 & \text{if } R_i \geq r_0 \\ 0 & \text{if } R_i < r_0 \end{cases}$$

- Comparison:

$$\begin{aligned} & \lim_{\epsilon \downarrow 0} E[Y_i | R_i = r_0 + \epsilon] - \lim_{\epsilon \downarrow 0} E[Y_i | R_i = r_0 - \epsilon] \\ & = \lim_{\epsilon \downarrow 0} E[Y_{1i} | R_i = r_0 + \epsilon] - \lim_{\epsilon \downarrow 0} E[Y_{0i} | R_i = r_0 - \epsilon] \\ & = E[Y_{1i} | R_i = r_0] - E[Y_{0i} | R_i = r_0] \end{aligned}$$

- Local polynomial estimators (with chosen window of R_i)

Instrumental Variables Methods

- Suppose there exists an instrumental variable (IV) that satisfies
  - $cov(D, Z) \neq 0$
  - $Z \perp (Y_0, Y_1)$
    - i.e., Exclusion restriction: The only difference created by IV is in the likelihood of receiving treatment
- e.g., distance to nearest college or density of colleges
- e.g., random lottery for college (but potential non-compliance)

Challenges with Essential Heterogeneity

- Consider

$$\begin{aligned} Y &= Y_0 + D(Y_1 - Y_0) \\ &= E[Y_0] + DE[Y_1 - Y_0] + (\varepsilon + \eta D) \end{aligned}$$

where $\varepsilon = Y_0 - E[Y_0]$ and $\eta = (Y_1 - Y_0) - E[Y_1 - Y_0]$

- Q: Does linear IV recover a parameter of interest?
  - If Δ const, classical IV results hold and IV recovers treatment effects
  - If Δ hetero and if essential hetero, classical IV results not hold and IV not recover interpretable parameters
  - If Δ hetero and if essential hetero, and if impose selection model (i.e., LATE monotonicity), IV recovers interpretable parameters (may/may not be of interest)

Challenges with Essential Heterogeneity

- Case 1: Δ const (i.e., $\eta = 0$)
  - Then, $cov(Z, Y_0) = 0$ implies $cov(Z, \varepsilon) = 0$
  - Then,

$$\frac{cov(Y, Z)}{cov(D, Z)} = E[Y_1 - Y_0]$$

  - If there is another IV, it identifies the same parameter
- Case 3: Δ varies even conditional on X
  - In general, we cannot identify $E[Y_1 - Y_0]$
  - We need $E[\varepsilon + \eta D | Z] = 0$
  - $E[\varepsilon | Z] = 0$, but

$$E[\eta D | Z] = E[\eta | D = 1, Z] P[D = 1 | Z]$$

    and even if $E[\eta | Z] = 0$, $E[\eta | D = 1, Z] \neq 0$ (i.e., essential hetero)

Challenges with Essential Heterogeneity

- Three approaches:
  1. LATE and MTE approaches (selection model approach)
     - May focus on different parameters
  2. Nonparametric IV approach (may be restrictive)
     - May be restrictive to allow for essential heterogeneity
  3. Nonparametric control function approach

Local Average Treatment Effect (LATE)

- Suppose Z_i is binary
  - e.g., close to college ($Z_i = 1$) or distant to college ($Z_i = 0$)
- We cannot recover ATE $E[Y_{1i} - Y_{0i}]$ in general
- Define counterfactual treatment: D_{1i} and D_{0i}
  - e.g., $D_{1i} = 1$ (or 0): i would have attended (or not attend) college, had i lived close to college
- "Monotonicity" assumption: $D_{1i} \geq D_{0i}$ for all i or $D_{1i} \leq D_{0i}$ for all i
  - e.g., no individual who would have attended college if living far from college but have not attended if living close to college
  - i.e., no defiers $\{D_{1i} = 0, D_{0i} = 1\}$

Local Average Treatment Effect (LATE)

- Then,

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} = 1, D_{0i} = 0]$$

  - $\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$ is the Wald estimand (or TSLS estimand)
  - $E[Y_{1i} - Y_{0i}|D_{1i} = 1, D_{0i} = 0]$ is called LATE
  - Individuals who behave like $\{D_{1i} = 1, D_{0i} = 0\}$ are called "compliers"
  - e.g., individuals who would have attended college if living close to college but have not attended if living far
- Need to understand which parameter you are estimating!

Marginal Treatment Effects (MTE)

- Suppose

$$D_i = 1[h(Z_i) \geq V_i]$$

  - The structure can be motivated by agent's optimizing behavior
    - e.g., attend college when net utility is positive
  - This model is equivalent to "monotonicity" assumption above!

- Assume Z_i is continuous, and define MTE as

$$E[Y_{1i} - Y_{0i} | V_i = v]$$

  - ATE for those who are indifferent (i.e., those on the "margin")

Marginal Treatment Effects (MTE)

- MTE:

$$E[Y_{1i} - Y_{0i} | V_i = v]$$

- Note that

$$\begin{aligned} E[Y_{1i} - Y_{0i} | D_{z'i} = 1, D_{zi} = 0] &= E[Y_{1i} - Y_{0i} | h(z') \geq V_i, h(z) < V_i] \\ &= E[Y_{1i} - Y_{0i} | h(z) < V_i \leq h(z')] \end{aligned}$$

therefore

$$E[Y_{1i} - Y_{0i} | V_i = h(z)] = \lim_{h(z') \to h(z)} E[Y_{1i} - Y_{0i} | h(z) < V_i \leq h(z')]$$

Marginal Treatment Effects (MTE)

- MTE can be viewed as a building block to generate various treatment parameters:

$$\tau_k = \int \omega_k(v,z) E[Y_{1i} - Y_{0i} | V_i = v] dv$$

  - $\omega_k(z,v)$ is known weight specific to the parameter of interest

- For example,

$$ATE = E[Y_{1i} - Y_{0i}] = \int_0^1 E[Y_{1i} - Y_{0i} | V_i = v] dv$$

$$LATE = E[Y_{1i} - Y_{0i} | D_{zi} = 1, D_{z'i} = 0]$$

$$= \int_{P(z')}^{P(z)} \frac{E[Y_{1i} - Y_{0i} | V_i = v]}{P(z) - P(z')} dv$$

$$ATT = E[Y_{1i} - Y_{0i} | D_i = 1] = \int_0^{P(z)} \frac{E[Y_{1i} - Y_{0i} | V_i = v]}{P[D = 1]} dv$$

Marginal Treatment Effects (MTE)

- Moreover, MTE can be recovered by

$$E[Y_{1i} - Y_{0i}|V_i = p] = \frac{\partial E[Y_i|P(Z_i) = p]}{\partial p}$$

where $P(X_i) = \Pr[D_i = 1|X_i]$

  - Continuity of $P(Z_i)$ and thus continuity of Z_i is important
    - e.g., Z_i is actual distance to nearest college
  - Support of $P(Z_i)$ and thus support of Z_i can be important, depending on parameters
    - e.g., for ATE, $P(Z_i) \to 1, 0$, which means $Z_i \to +\infty, -\infty$
- MTE itself can be a parameter of interest
  - Non-constant MTE reflects heterogeneity
- MTE can be estimated nonparametrically, but typically after imposing more structure

Nonparametric IV Approach

- Let

$$Y_i = g(D_i, U_i)$$

  - Want to know g because $Y_{1i} = g(1, U_i)$ and $Y_{0i} = g(0, U_i)$
- Let Z_i be an IV that satisfies $E[U_i|Z_i] = 0$
- Assume U_i is scalar and $g(D_i, \cdot)$ is strictly monotonic
  - e.g., $Y_i = g(D_i) + U_i$
  - If U_i is continuous, Y_i should be continuous

Nonparametric IV Approach

- Then

$$0 = E[U_i|Z_i] = E[g^{-1}(D_i, Y_i)|Z_i]$$

  - e.g., $0 = E[U_i|Z_i] = E[Y_i - g(D_i)|Z_i]$

- If we additionally impose completeness condition (i.e., Z_i is relevant for D_i in "nonparametric sense"), then g can be recovered from

$$E[Y_i|Z_i] = E[g(D_i)|Z_i]$$

- Estimation is more challenging due to the ill-posed inverse problem
  - $E[\cdot]$ is smooth, so its inverse is non-smooth
  - Related to "small denominator" problem
  - Regularization is needed

Nonparametric Control Function Approach

- Assume

$$D_i = h(Z_i, V_i)$$

where V_i is scalar and $h(Z_i, \cdot)$ is strictly monotonic
  - e.g., $D_i = h(Z_i) + V_i$
  - If V_i is continuous, D_i should be continuous (e.g., years of education)

- Then, construct a CF:

$$V_i = h^{-1}(Z_i, D_i)$$

  - e.g., $V_i = D_i - h(Z_i)$

Nonparametric Control Function Approach

- Assume $E[U_i|V_i, Z_i] = E[U_i|V_i]$
- Let $Y_i = g(D_i) + U_i$ for simplicity
- Then

$$E[Y_i|D_i, Z_i] = g(D_i) + E[U_i|D_i, Z_i] = g(D_i) + E[U_i|V_i, Z_i]$$
$$= g(D_i) + E[U_i|V_i] = g(D_i) + \lambda(V_i)$$

- Equivalently

$$Y_i = g(D_i) + \lambda(V_i) + \eta_i$$

where $E[\eta_i|D_i, Z_i] = 0$

- Nonparametrically estimate g and λ after estimating V_i

References

- Holland, P. W. (1986). Statistics and Causal Inference. Journal of the American Statistical Association, 81(396):945-960.
- Heckman, J. J. (2008). Econometric causality. International statistical review, 76(1), 1-27.
- Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.
- Heckman, J. J., & Vytlacil, E. J. (2007). Econometric evaluation of social programs, part I. Handbook of econometrics, 6, 4779-4874.
- Heckman, J. J., & Vytlacil, E. J. (2007). Econometric evaluation of social programs, part II. Handbook of econometrics, 6, 4875-5143.

Thank You! ☺