



0 0 ECON0021 Microeconometrics Index Splited W1234-10

Table of Content

Table of Contents

1. Table of Content
2. Week 1: Causal Inference, RCTs, QTE
 1. 1 Causal Inference, RCTs, QTE - A
 1. Selections
 2. Treatment Effect and Evaluation Problem
 3. Randomised Control Trails (RCT)
 1. Effects
 2. Limitations and Caveats of RCTs
 4. OLS
 5. Quantile Treatment Effects QTE
 6. Appendix
 3. Week 2: Regressions and Matching
 1. 2 PPT Regression and Matching - A
 1. Matching
 1. Basics
 2. Matching Estimand
 3. Assumptions
 4. Identification
 5. Matching with Discrete Characteristics

- 6. Regression as Pesudo-Matching
 - 7. Choices of Matching Variables
 - 2. Propensity Score Matching
 - 3. Appendix
4. Week 3: Instrumental Variables and Control Functions
- 1. 3 PPT Instrumental Variables and Control Functions - A
 - 1. Selection on Unobservables and Endogeneity
 - 2. IV with Homogeneous Treatment Effect
 - 3. IV With Heterogeneous Treatment Effects
 - 1. 4 Assumptions
 - 2. LATE
 - 4. Control Function Methods
 - 1. Heckman Selection Model
 - 2. (Generalised) Control Functions for Selections on Unobservables and ATE
5. Week 4: Repeated Cross-Sections and Panel Data
- 1. 4 Repeated Cross-Sections and Panel Data - A
 - 1. Difference in Difference
 - 1. Common Trend Assumption
 - 2. DiD Estimand and Identification
 - 3. DiD as Regression
 - 4. Main Fails of Common Trend
 - 5. Policy Shift Level
 - 2. Quantile DiD and Quantile CiC
6. Week 10: Regression Discontinuity Designs
- 1. 10 RDD - A
 - 1. Sharp RD
 - 1. Treatment Assignment
 - 2. Continuity Assumption
 - 3. Sharp RD Estimand and Treatment Effect
 - 4. Implementation
 - 2. Fuzzy RD
 - 1. Setup
 - 2. Estimand
 - 3. Fuzzy RD as IV
 - 3. Practical Advice
7. Others
8. Revisions
-

Week 1: Causal Inference, RCTs, QTE

1 Causal Inference, RCTs, QTE - A

- Neyman-Rubin Potential Outcome Model
 - Treatment Variable:

$$D_i = \begin{cases} 0 & \text{if } i \text{ is treated} \\ 1 & \text{if } i \text{ is not treated} \end{cases}$$

- Potential Outcomes:

$$\begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

- We observe:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$$

Selections

- Selection
- Selection on Levels (\rightsquigarrow Selection bias): individuals choose treatment based on level of outcome in absence of treatment

$$\mathbb{E}[Y_{0i}|D_i = 1] \neq \mathbb{E}[Y_{0i}|D_i = 0]$$

- Selection on Gains (\rightsquigarrow Heterogenous TE): individual choose treatment based on gains from treatment

$$\mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] \neq \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 0]$$

- Example: Roy Model of Occupational Choice

Treatment Effect and Evaluation Problem

- Treatment Effect and Evaluation Problem
 - Treatment Effect for Each Individual:

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

- Fundamental “Evaluation Problem”: either Y_{0i} or Y_{1i} is observed, *never both at the same time* \rightsquigarrow cannot observe treatment effect for any individual β_i directly \rightsquigarrow focus on distribution of β_i instead

- Average Treatment Parameters
 - Average Treatment Effect ATE:

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

- Average Effect of Treatment on the Treated ATT:

$$\mathbb{E}[\beta_i | D_i = 1] = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$$

- Average Effect of Treatment on the Untreated ATU:

$$\mathbb{E}[\beta_i | D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 0]$$

- Comparison of Observed Outcomes in Treatment and Control Group:

$$\begin{aligned} \underbrace{\mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]}_{\text{Observed Difference}} &= \mathbb{E}[Y_{1i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 0] \\ &= \underbrace{\mathbb{E}[Y_{1i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_{0i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 0]}_{\text{Selection Bias}} \end{aligned}$$

- Selection on levels \rightsquigarrow Bias (Observed difference \neq ATT)

Randomised Control Trials (RCT)

Effects

- Randomised Control Trials: RCTs randomly assign individuals into treatment and control group
- Effect of Randomisation: treatment decision is made independent of potential outcomes

$$D_i \perp (Y_{0i}, Y_{1i}) \implies \begin{cases} \mathbb{E}[Y_{1i} | D_i = 1] = \mathbb{E}[Y_{1i} | D_i = 0] = \mathbb{E}[Y_{1i}] \\ \mathbb{E}[Y_{0i} | D_i = 1] = \mathbb{E}[Y_{0i} | D_i = 0] = \mathbb{E}[Y_{0i}] \end{cases}$$

- \implies No selection on levels

$$\mathbb{E}[Y_{0i} | D_i = 1] = \mathbb{E}[Y_{0i} | D_i = 0] = \mathbb{E}[Y_{0i}]$$

- \implies No selection bias (Observed difference = ATT)
- \implies No selection on gains

$$\mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1] = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}]$$

- \implies (ATT=ATE)

- Observed Differences in RCTs = ATE:

$$\begin{aligned} \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0] &= \underbrace{\mathbb{E}[Y_{1i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_{0i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 0]}_{\text{Selection Bias=0}} \\ &= \underbrace{\mathbb{E}[Y_{1i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 1]}_{\text{ATT=ATE}} \\ &= \underbrace{\mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}]}_{\text{ATE}} \end{aligned}$$

Limitations and Caveats of RCTs

- Limitations of RCTs

- Small-scale RCTs:
 - participants not representative
 - possible imbalance across groups (always **test the balance of characteristics**)
 - may miss general equilibrium effects
- Large-scale RCTs:
 - expensive
 - ethical/political economy concerns
- Possible **spillovers**
- RCTs may affect outcomes in a way the actual policy may not, affecting the interpretation and **external validity** (e.g. individuals may choose not to participate when they know they will be subject to randomisation)
- **Potential selection bias** if randomisation is violated and imperfect compliance
- RCTs do not always identify actual policy effects due to **absence of selection on gains** (in reality, people optimise their behaviours and there's likely to be selection on gains)

OLS

- Relationship of Outcome and Treatment Status:

$$\begin{aligned}
 Y_i &= D_i Y_{1i} - (1 - D_i) Y_{0i} \\
 &= Y_{0i} + (Y_{1i} - Y_{0i}) D_i \\
 &= \underbrace{\mathbb{E}[Y_{0i}]}_{\equiv \alpha} + \underbrace{(Y_{1i} - Y_{0i})}_{\equiv \beta_i} D_i + \underbrace{Y_{0i} - \mathbb{E}[Y_{0i}]}_{\equiv u_i}
 \end{aligned}$$

- This looks like a regression except β_i is random rather than a constant

- **OLS Regression Equation:**

$$Y_i = \alpha + \beta D_i + u_i$$

where $D_i = \mathbf{1}[\text{Treated}]$

- We can add controls to improve precision
- **OLS Identification** (derived in appendix)

$$\begin{aligned}
 \beta_{OLS} &= \underbrace{\mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]}_{\text{Observed Difference}} \\
 &= \underbrace{\mathbb{E}[Y_{1i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 1]}_{\text{ATT}} + \underbrace{\mathbb{E}[Y_{0i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 0]}_{\text{Selection Bias}}
 \end{aligned}$$

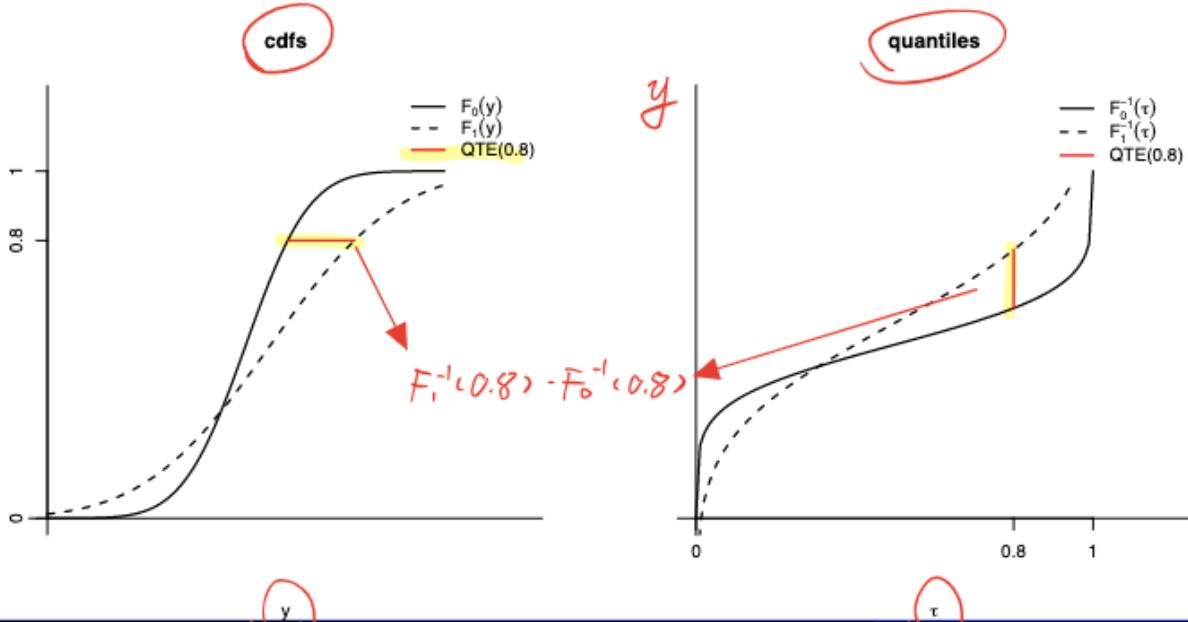
- In RCTs: $D_i \perp (Y_{0i}, Y_{1i}) \implies D_i \perp (u_i, \beta_i) \implies \beta_{OLS} = ATT = ATE$

Quantile Treatment Effects QTE

- **Quantile Treatment Effects**: change in outcomes at a given quantile τ before/after treatment:

$$QTE_\tau = F_1^{-1}(\tau) - F_0^{-1}(\tau)$$

where $F^{-1}(.)$ is the quantile function (inverse of CDF)



- QTE in RCTs:

- RCTs ensure that observed distributions in control/treatment groups are equal to potential outcomes:

$$\begin{cases} F(y|D_i = 0) = F_0(y|D_i = 0) = F_0(y) \\ F(y|D_i = 1) = F_1(y|D_i = 1) = F_1(y) \end{cases}$$

- Thus, QTE equals to observed differences in quantiles:

$$QTE_\tau = \underbrace{F^{-1}(\tau|D_i = 1) - F^{-1}(\tau|D_i = 0)}_{\text{Observed Difference in Quantiles}}$$

- **rank reversals:** QTE_τ is not necessarily the treatment effect β_i for individual whose outcome level is at the τ -quantile
- Example

Appendix

- when ATT = ATE?
- OLS interpretation when regressor is binary

Week 2: Regressions and Matching

2 PPT Regression and Matching - A

Matching

Basics

- Matching Idea
- Assume Selection on Observables: participation decision depends only on observed characteristics X_i
- \implies for people with the same characteristics X_i , the participation decision is random (Conditional Random Assignment / RCT for given characteristics):

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

- With the same observed characteristics, outcome of individual in the control group provides counterfactual for individual in the treatment group

Matching Estimand

- Observed difference in outcomes for individuals with characteristics $X_i = x$ is:

$$\mathbb{E}[Y_i | D_i = 1, X_i = x] - \mathbb{E}[Y_i | D_i = 0, X_i = x]$$

- Then, average all observed difference weighted by their probability of occurring
- Matching Estimand for ATE:

$$\beta_M = \int \left\{ \mathbb{E}[Y_i | D_i = 1, X_i = x] - \mathbb{E}[Y_i | D_i = 0, X_i = x] \right\} dF_x(x)$$

- Matching Estimand for ATT:

$$\beta_M = \int \left\{ \mathbb{E}[Y_i | D_i = 1, X_i = x] - \mathbb{E}[Y_i | D_i = 0, X_i = x] \right\} dF_x(x | D_i = 1)$$

Assumptions

- Matching(1): Conditional Mean Independence / Unconfoundedness: for all characteristics x :

$$\begin{cases} \mathbb{E}[Y_{0i} | X_i = x, D_i = 1] = \mathbb{E}[Y_{0i} | X_i = x, D_i = 0] \\ \mathbb{E}[Y_{1i} | X_i = x, D_i = 1] = \mathbb{E}[Y_{1i} | X_i = x, D_i = 0] \end{cases}$$

- Weaker than full conditional independence / conditional random assignment
- No unobservable variables can jointly determine selection and potential outcomes $\iff X$ includes all variables that simultaneously characterise the participation decision and potential outcomes \iff For treated and non-treated individuals with same characteristics X_i , the participation decision D_i does not depend on potential outcomes
- Rules out selections on unobservables gains: we need

$$\mathbb{E}[Y_{1i} - Y_{0i} | X_i = x, D_i = 1] = \mathbb{E}[Y_{1i} - Y_{0i} | X_i = x, D_i = 0]$$

- Matching(2): Common Support: for all characteristics x :

$$0 < \underbrace{P(D_i = 1 | X_i = x)}_{\text{Propensity Score}} < 1$$

- Every characteristics $X_i = x$ occurring in the treatment group also occurs in the control group (otherwise $P(D_i = 1|X_i = x) = 1$)
- Every characteristics $X_i = x$ occurring in the control group also occurs in the treatment group (otherwise $P(D_i = 1|X_i = x) = 0$)

Identification

- Matching Est. Identifies ATE:

$$\begin{aligned}
 ATE &= \mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}\left[\mathbb{E}[Y_{1i} - Y_{0i}|X_i]\right] \\
 &= \int \left[\mathbb{E}[Y_{1i}|X_i = x] - \mathbb{E}[Y_{0i}|X_i = x]\right] dF_x(x) \\
 &= \int \left[\mathbb{E}[Y_i|D_i = 1, X_i = x] - \mathbb{E}[Y_i|D_i = 0, X_i = x]\right] dF_x(x) \quad (\text{Conc}) \\
 &= \beta_M
 \end{aligned}$$

Matching with Discrete Characteristics

- Matching With Discrete Characteristics
- If characteristics are discrete (i.e. there are $k = 1, \dots, K$ cells), then matching estimand becomes a weighted sum:

$$ATE = \beta_M = \sum_{k=1}^K \underbrace{\left[\mathbb{E}[Y_i|D_i = 1, X_i = k] - \mathbb{E}[Y_i|D_i = 0, X_i = k] \right]}_{\text{Treatment Effect in Cell } k} \underbrace{P(X_i = k)}_{\text{Cell Pr}}$$

- Compute treatment effect in each cell
- Then, take a linear combination of cell treatment effects weighted by cell probabilities

Regression as Pseudo-Matching

- Regression as Pseudo-Matching
- Setup (no constant):

$$Y_i = \beta_R D_i + \sum_{k=1}^K \alpha_k \underbrace{d_{ik}}_{\text{Cell Dummy}} + \epsilon_i$$

where D_i is the treatment dummy, $d_{ik} = \mathbb{1}[X_i = k]$ is the cell dummy, and β_R is the regression estimand

- No interaction terms
- OLS Identification

$$\beta_R = \sum_{k=1}^K \left\{ \mathbb{E}[Y_i|D_i = 1, X_i = k] - \mathbb{E}[Y_i|D_i = 0, X_i = k] \right\} w_k$$

where

$$w_k = \frac{\overbrace{P(D_i = 1|X_i = k) [1 - P(D_i = 1|X_i = k)]}^{Var(D_i|X_i=k)} P(X_i = k)}{\sum_{k=1}^K \underbrace{P(D_i = 1|X_i = k) [1 - P(D_i = 1|X_i = k)]}_{Var(D_i|X_i=k)} P(X_i = k)}$$

- *Matching vs Regression*

- Regression weights w_k depends on the conditional variance of treatment status (balanced cells weight more)
 - this weighting is different from the matching estimand where $w_k = Pr(X_i = k)$
- Under Cond. Mean Indep., CS, and homogenous TE: $ATE = \beta_M = \beta_R$ (both matching and regression identify ATE)

Choices of Matching Variables

- Choice of Matching Variables
- Appropriate matching variables:
 - Simultaneously characterise participation decision (selection) and outcome in the absence of treatment
 - Pre-treatment variables only
- Trade-off between Conditional Mean Indep. and Common Support (too many cells ↗ more likely for CMI to hold but there may be empty cells ([curse of dimensionality](#)))

Propensity Score Matching

- Curse of Dimensionality and Propensity Score
- Definition of Propensity Score:

$$P(x) \equiv P(D_i = 1|X_i = x)$$

- Balancing Property of the Propensity Score

$$\begin{cases} Y_{0i} \perp D_i | X_i \implies Y_{0i} \perp D_i | P(X_i) \\ Y_{1i} \perp D_i | X_i \implies Y_{1i} \perp D_i | P(X_i) \end{cases}$$

- Disadvantages:
 - This need [full independence](#), not just mean independence
 - $P(X)$ is unknown: use a parametric solution (e.g. logit/probit) to estimate propensity scores (avoiding curse of dimensionality)
- Issue: Propensity score is continuous \implies cannot find two individuals with the same propensity score ↗ need to choose individuals with "similar propensity scores"
 - Nearest neighbour matching (computationally intensive)
 - Kernel matching (computationally intensive)
 - Propensity score re-weighting (simple)

- Propensity Score Re-Weighting
 - Propensity Score Re-Weighting: under Full Conditional Independence ($(Y_{1i}, Y_{0i}) \perp D_i | X_i$):

$$\begin{aligned}
 ATE &= \mathbb{E}[Y_{1i} - Y_{0i}] \\
 &= \underbrace{\mathbb{E}\left[\frac{Y_i D_i}{P(X_i)}\right]}_{\mathbb{E}[Y_{1i}]} - \underbrace{\mathbb{E}\left[\frac{Y_i(1 - D_i)}{1 - P(X_i)}\right]}_{\mathbb{E}[Y_{0i}]} \\
 &= \mathbb{E}\left[\frac{Y_i D_i}{P(X_i)} - \frac{Y_i(1 - D_i)}{1 - P(X_i)}\right]
 \end{aligned}$$

Appendix

- Integral Notation: Let X be a random variable with cdf $F_X(x)$. An integral of the form

$$\int g(x) dF_X(x)$$

means the following:

- If X is continuously distributed with pdf $f_X(x)$:

$$\int g(x) dF_X(x) = \int g(x) f_X(x) dx$$

- If X is discretely distributed with values $\{x_1, \dots, x_k\}$:

$$\int g(x) dF_X(x) = \sum_{k=1}^K g(x_k) P(X = x_k)$$

- Conditional Independence vs. Mean Independence: Independence \implies Conditional Independence; Mean Independence \implies Conditional Mean Independence; all other directions are not true
 - Regression as Pseudo-Matching
 - Deriving the Propensity Score Re-weighting
-

Week 3: Instrumental Variables and Control Functions

3 PPT Instrumental Variables and Control Functions - A

Selection on Unobservables and Endogeneity

- Regression and Selection on Unobservables:

$$\begin{aligned}
Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} \\
&= Y_{0i} + (Y_{1i} - Y_{0i}) D_i \\
&= \underbrace{E[Y_{0i}]}_{\alpha} + \underbrace{(Y_{1i} - Y_{0i})}_{\beta_i} D_i + \underbrace{(Y_{0i} - E[Y_{0i}])}_{u_i}
\end{aligned}$$

- Selection on unobservables: Y_{0i}, D_i have common unobserved factors $\implies Y_{0i} \not\perp D_i \implies u_i \not\perp D_i$
 - D_i is endogenous: $Cov(D_i, u_i) \neq 0$
 - OLS estimator of β is biased and inconsistent

IV with Homogeneous Treatment Effect

- IV: Basic Idea

- IV Assumptions

- Exclusion: $Cov(Z_i, u_i) = 0$ (untestable)

- Relevance: $Cov(Z_i, D_i) \neq 0$

- In 1st stage regression ($D_i = \pi_0 + \pi_1 Z_i + v_i$), $\pi_1 = \frac{Cov(Z_i, D_i)}{Var(Z_i)} \neq 0$, we can test this

- Identification

- First-stage regression (D_i on Z_i):

$$D_i = \pi_0 + \pi_1 Z_i + v_i$$

- Second-stage regression (Y_i on D_i):

$$Y_i = \alpha + \beta D_i + u_i$$

- Reduced-form* regression (Y_i on Z_i):

$$Y_i = \alpha + \beta \pi_0 + \beta \pi_1 Z_i + u_i + \beta v_i$$

- $\implies Cov(Y_i, Z_i) = \beta Cov(D_i, Z_i) + \underbrace{Cov(u_i, Z_i)}_{=0 \text{ by exclusion}}$

- Identification of ATE and ATT (by relevance):

$$\beta_{IV} = \frac{\frac{Cov(Y_i, Z_i)}{Var(Z_i)}}{\underbrace{\frac{Cov(D_i, Z_i)}{Var(Z_i)}}_{\text{Reduced-Fm / 1st Stage}}} = \frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} \underbrace{= \beta}_{\text{Under Homogeneous TE}} = ATE = ATT$$

- Binary IV: Wald Estimand: when IV is *binary*:

$$\beta_{Wald} = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]} = \beta$$

IV With Heterogeneous Treatment Effects

- Heterogeneous treatment effects: β_i varies across i
- Potential Treatment Indicators and Potential Outcomes
 - Treatment: $D_{Z_i,i}$
 - Potential outcomes: $Y_{D_i,Z_i,i}$

4 Assumptions

- Independence (*random assignment of IV*):

$$(Y_{11i}, Y_{10i}, Y_{01i}, Y_{00i}, D_{1i}, D_{0i}) \perp Z_i$$

- Implication: we can estimate causal effects of Z_i on D_i and Z_i on Y_i as in RCTs

- Exclusion (*IV does not directly influence potential outcomes*):

$$\begin{cases} Y_{11i} = Y_{10i} = Y_{1i} \\ Y_{01i} = Y_{00i} = Y_{0i} \end{cases}$$

- Implication: we have the usual expression for observed outcomes:

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i} = \underbrace{\mathbb{E}[Y_{0i}]}_{\alpha} + \underbrace{(Y_{1i} - Y_{0i})}_{\beta} D_i + \underbrace{(Y_{0i} - \mathbb{E}[Y_{0i}])}_{u_i}$$

- Monotonicity (*No defiers*): increasing IV may induce some untreated individuals to take treatment, but cannot induce any treated individuals to leave treatment:

$$D_{1i} \geq D_{0i} \quad \forall i$$

- Relevance (*IV induces variation in the treatment*): IV needs to induce variations in treatment indicator:

$$\mathbb{E}[D_{1i} - D_{0i}] \neq 0$$

- This is satisfied if individuals with $Z_i = 1$ are more likely to participate:

$$P(D_{1i} = 1) = \mathbb{E}[D_{1i}] > \mathbb{E}[D_{0i}] = P(D_{0i} = 1)$$

LATE

- Reactive Subpopulations
 - Always takers $D_{1i} = D_{0i} = 1$
 - Never takers $D_{1i} = D_{0i} = 0$
 - Compliers $D_{1i} = 1, D_{0i} = 0$
 - Defiers $D_{1i} = 0, D_{0i} = 1$
- Wald Estimator identifies Local Average Treatment Effect (LATE)
 - Reduced Form:

$$\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}] P(D_{1i} > D_{0i})$$

- First Stage:

$$\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] = P(D_{1i} > D_{0i})$$

- Reduced Form Divided by First Stage
- Local Average Treatment Effect (LATE) Identification:

$$\begin{aligned}\beta_{Wald} &= \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} \\ &= \frac{\mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i}]P(D_{1i} > D_{0i})}{P(D_{1i} > D_{0i})} \\ &= \underbrace{\mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]}_{\text{LATE}}\end{aligned}$$

- Interpretation of LATE:

$$LATE \equiv \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

is the average treatment effect for those who are induced to take the treatment when the IV is increased ([treatment effect for the compliers only](#))

- *Under independence + exclusion + monotonicity + relevance, Wald Estimand identifies LATE:*

$$\beta_{Wald} = LATE \begin{cases} = ATE = ATT & \text{if Homogeneous TE} \\ \neq ATE \text{ or } ATT & \text{if Heterogenous TE} \end{cases}$$

- LATE is a [local effect](#) (only in complier group)
- different instruments identify different LATE (we need policy relevant instrument)
- LATE With Variable Treatment Intensity

- If treatment indicators D_{0i}, D_{1i} can take on multiple values $\{0, 1, \dots, \bar{d}\}$
- Under independence + exclusion + monotonicity + relevance, Wald est. identifies a [weighted average of causal effects](#)

$$\beta_{Wald} = \sum_{d=1}^{\bar{d}} w_d \mathbb{E}[Y_{di} - Y_{(d-1)i}|D_{1i} \geq d > D_{0i}]$$

where the weights are

$$w_d = \frac{P(D_{1i} \geq d > D_{0i})}{\sum_{j=1}^{\bar{d}} P(D_{1i} \geq j > D_{0i})}$$

- LATE Interpretation
- Practical Advice for Using LATE
 - Motivate IV
 - Motivate IV validity (relevance + exclusion)
 - Placebo test on outcomes that should not be affected by IV

- Control variables (exclusion more likely to valid; LATE interpretation more complex)
- Who are the compilers? (is the instrument policy relevant?)
- Check IV
 - Always report 1st-stage est. (ideally $F > 104.7$)
 - Inspect reduced form estimates
- PS1:1.d when LATE=ATE?

Control Function Methods

- Idea: specify the exact nature of dependence between observables and unobservables (explicitly model and control for potential selection bias)

Heckman Selection Model

- Outcome:

$$Y_i = \begin{cases} \beta^\top X_i + u_i, & \text{if } D_i = 1 \\ \text{unobserved}, & \text{if } D_i = 0 \end{cases}$$

with $E[u_i|X_i, Z_i] = 0$

- Selection: $D_i = 1[\gamma^\top Z_i + v_i \geq 0]$
- Further assumptions:
 - X_i is a strict sub-vector of Z_i
 - $Z_i \perp (u_i, v_i)$
- Additional distributional assumption:
 - $(u_i, v_i) \sim \mathcal{I}\left(0, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}\right)$
- Under all those assumptions:

$$E[Y_i|Z_i, D_i = 1] = \beta^\top X_i + \rho E[v|D_i = 1] = \beta^\top X_i + \rho \lambda(\gamma^\top Z_i)$$

where ρ is the correlation between u_i, v_i and $\lambda(\gamma^\top Z_i)$ is the Inverse Mill's Ratio

- Estimation
 - Step 1 (Logit/Probit on the whole sample)

$$Pr(D_i = 1|Z_i) = \Phi(\gamma^\top Z_i)$$

- Step 2 (Main Regression)

$$E[Y_i|Z_i, D_i = 1] = \beta^\top X_i + \rho \underbrace{\lambda(\hat{\gamma}^\top Z_i)}_{\text{Inverse Mill's Ratio}}$$

- Caveats
 - SE needs to be adjusted to account for the 1st step estimation of γ

- Strictly speaking, we can have $X_i = Z_i$ because $\lambda(\cdot)$ is a non-linear function, but it's very similar to a linear function, so having more variables in Z_i prevents such issue

(Generalised) Control Functions for Selections on Unobservables and ATE

- More flexible setup: *no distributional assumption*
- Model:
 - Outcome:

$$Y_i = \underbrace{\alpha}_{E[Y_{0i}]} + \underbrace{\beta_i}_{Y_{1i}-Y_{0i}} D_i + \underbrace{u_i}_{Y_{0i}-E[Y_{0i}]}$$

- Selection: $D_i = \mathbb{1}[\gamma^\top Z_i + v_{Di} \geq 0]$
- Further assumption:
 - **Exogeneity**: $(Y_{0i}, Y_{1i}, v_{Di}) \perp Z_i$
 - This allows $(Y_{0i}, Y_{1i}) \not\perp v_{Di} \implies (Y_{0i}, Y_{1i}) \not\perp D_i$ even condition on Z_i , so there is endogeneity and we cannot use matching
 - **Relevance**: Z_i induces changes in D_i
 - X_i is a strict sub-vector of Z_i
- This implies:

$$\begin{cases} \mathbb{E}[Y_i | Z_i, D_i = 1] = \alpha + ATE + K_1(P(Z_i)) \\ \mathbb{E}[Y_i | Z_i, D_i = 0] = \alpha + K_0(P(Z_i)) \end{cases}$$

where $K_1(P(Z_i)) = \text{const.} \times \frac{\phi(\Phi^{-1}(1-P(Z_i)))}{1-\Phi(\Phi^{-1}(1-P(Z_i)))}$ and $K_0(P(Z_i))$ can be constructed similarly, and we can retrieve ATE accordingly

Week 4: Repeated Cross-Sections and Panel Data

4 Repeated Cross-Sections and Panel Data - A

- Before/After Analysis
 - **Natural Experiment**: a policy shift that affects one part of the population, but not another
- Potential Outcomes
 - Notation: $Y_{D_{it},it}$
 - In period 1: $Y_{0,i1}$
 - In period 2: $\begin{cases} Y_{0,i2}, & \text{if } D_{i2} = 0 \\ Y_{1,i2}, & \text{if } D_{i2} = 1 \end{cases}$
 - Observed outcome in period t

$$Y_{it} = D_{it} Y_{1,it} + (1 - D_{it}) Y_{0,it} = D_{it} (Y_{1,it} - Y_{0,it}) + Y_{0,it}$$

- Object of interest: $ATT = \mathbb{E}[Y_{1,i2} - Y_{0,i2} | D_{i2} = 1]$

- No-Trends Assumption

- If

$$\mathbb{E}[Y_{0,i2}|D_{i2} = 1] = \mathbb{E}[Y_{0,i1}|D_{i2} = 1]$$

then FD identifies ATT: $\mathbb{E}[Y_{i2} - Y_{i1}|D_{i2} = 1] = \underbrace{\mathbb{E}[Y_{1,i2} - Y_{0,i2}|D_{i2} = 1]}_{ATT}$

- First Differences

Difference in Difference

Common Trend Assumption

- Common Trend Assumption: in absence of treatment, the changes in outcomes in control/treatment groups are the same:

$$\mathbb{E}[Y_{0,i2} - Y_{0,i1}|D_{i2} = 1] = \mathbb{E}[Y_{0,i2} - Y_{0,i1}|D_{i2} = 0]$$

- Allows for
 - Selection on non-treatment levels: $\mathbb{E}[Y_{0,it}|D_{i2} = 1] \neq \mathbb{E}[Y_{0,it}|D_{i2} = 0]$, for $t = 1, 2$
 - Selection on gains (our objective of interest is ATT anyway):
 $\mathbb{E}[Y_{1,i2} - Y_{0,i2}|D_{i2} = 1] \neq \mathbb{E}[Y_{1,i2} - Y_{0,i2}|D_{i2} = 0]$
 - Individual-specific and time-invariant unobservables

DiD Estimand and Identification

- Difference-in-Differences
- Difference in the treatment group:

$$\begin{aligned}\mathbb{E}[Y_{i2} - Y_{i1}|D_{i2} = 1] &= \mathbb{E}[D_{i2}(Y_{1,i2} - Y_{0,i2}) + Y_{0,i2} - Y_{0,i1}|D_{i2} = 1] \\ &= \underbrace{\mathbb{E}[Y_{1,i2} - Y_{0,i2}|D_{i2} = 1]}_{ATT} + \underbrace{\mathbb{E}[Y_{0,i2} - Y_{0,i1}|D_{i2} = 1]}_{\text{Trend for the Treated}}\end{aligned}$$

- Difference in the control group:

$$\begin{aligned}\mathbb{E}[Y_{i2} - Y_{i1}|D_{i2} = 1] &= \mathbb{E}[D_{i2}(Y_{1,i2} - Y_{0,i2}) + Y_{0,i2} - Y_{0,i1}|D_{i2} = 0] \\ &= \underbrace{\mathbb{E}[Y_{0,i2} - Y_{0,i1}|D_{i2} = 0]}_{\text{Trend for the Control}}\end{aligned}$$

- DiD Estimand identifies ATT under common trend assumption:

$$\beta_{DiD} = \underbrace{\mathbb{E}[Y_{i2} - Y_{i1}|D_{i2} = 1]}_{\text{Diff. in Treatment Group}} - \underbrace{\mathbb{E}[Y_{i2} - Y_{i1}|D_{i2} = 0]}_{\text{Diff. in Control Group}} = \underbrace{\mathbb{E}[Y_{1,i2} - Y_{0,i2}|D_{i2} = 1]}_{ATT}$$

DiD as Regression

- DiD As Regression:

$$Y_{it} = \beta_0 + \beta_1 treat_i + \beta_2 after_t + \beta_3 treat_i \times after_t + u_{it}$$

where β_3 is the DiD estimator

	treatment	control	difference
before	$\beta_0 + \beta_1$	β_0	β_1
after	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_2$	$\beta_1 + \beta_3$
difference	$\beta_2 + \beta_3$	β_2	β_3

- *Advantage:*

- easy to compute SE (allows for serial correlations in U_{it} or clustered SE)
- easy to add more covariates (to control for confounding trends / increase precision)

- *Disadvantage:*

- this identifies ATT only when ATT is homogenous, otherwise (e.g. when the policy implementation is staggered) it identifies a complex weighted average

Main Fails of Common Trend

- Main Fails of Common Trends Assumption
 - Differential Macroeconomic Trends
 - always check pre-treatment trends
 - Composition Effects: treatment may change the composition of the groups
 - We can redefine groups such that compositions are not affected by treatment, but the TE we identify will be Intention to Treat (ITT) instead of ATT
 - Idiosyncratic temporary shocks (e.g. Ashenfelter's Dip): enrolment is more likely when a temporary dip in earnings occurs just before the programme starts -- faster earnings growth expected even in absence of the programme

Policy Shift Level

- Aggregate Policy Shift
 - If policy shift happens on *aggregate level*, we only need repeated cross-sections:

$$Y_{ist} = \beta_0 + \beta_1 treat_s + \beta_2 after_t + \beta_3 treat_s \times after_t + u_{ist}$$
 - If policy shift happens on *individual level*, we will need panel data:

$$Y_{it} = \beta_0 + \beta_1 treat_i + \beta_2 after_t + \beta_3 treat_i \times after_t + u_{it}$$

Quantile DiD and Quantile CiC

- DiD and Quantile Treatment Effects
- Use non-linear DiD to estimate QTE

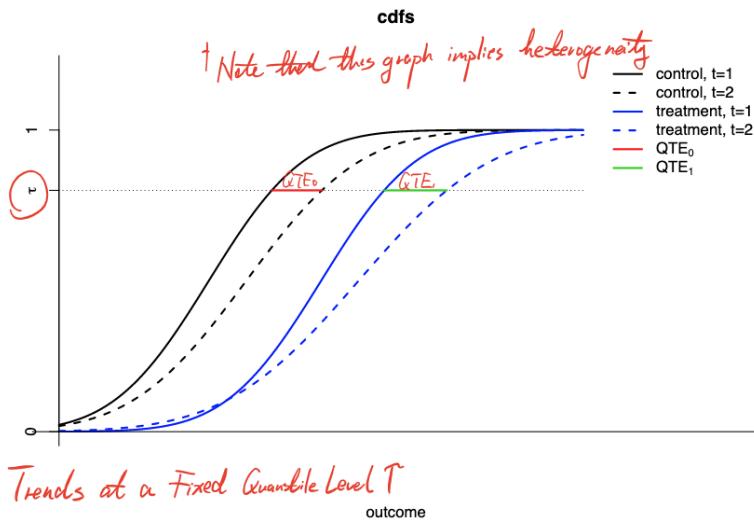
- Quantile DiD

$$\text{Quantile DiD} = \underbrace{[F_2^{-1}(\tau) - F_1^{-1}(\tau)|D_{i2} = 1]}_{QTE_1} - \underbrace{[F_2^{-1}(\tau) - F_1^{-1}(\tau)|D_{i2} = 0]}_{QTE_0}$$

- Steps:

- Fix quantile level τ
- Compute QTEs at τ for the treatment (QTE_1) and control (QTE_0)
- Take difference: $QTE_1 - QTE_0$

- This requires common trend at the fixed quantile level τ



- Quantile CiC

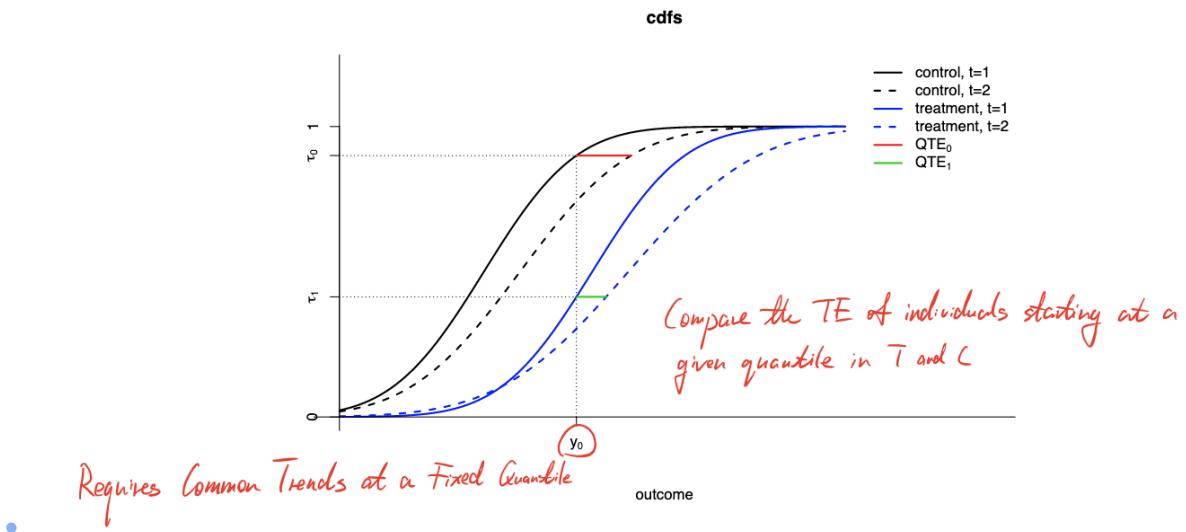
$$\text{Quantile CiC} = \underbrace{[F_2^{-1}(\tau_1) - F_1^{-1}(\tau_1)|D_{i2} = 1]}_{QTE_1} - \underbrace{[F_2^{-1}(\tau_0) - F_1^{-1}(\tau_0)|D_{i2} = 0]}_{QTE_0}$$

where τ_1, τ_0 is the quantile corresponding to a given y_0 in the treatment/control group in $t=0$ respectively

- Steps:

- Fix y at y_0 , find its corresponding quantile in treatment (τ_1) and control (τ_0)
- Compute QTE at τ_1 for the treatment (QTE_1) and QTE at τ_0 for the control (QTE_0)
- Take difference: $QTE_1 - QTE_0$

- This requires common trend at a fixed initial y_0



Week 10: Regression Discontinuity Designs

10 RDD - A

- Regression Discontinuity (RD) Design: D_i is assigned based on whether some "running variable" A_i is above/below a cutoff c

Sharp RD

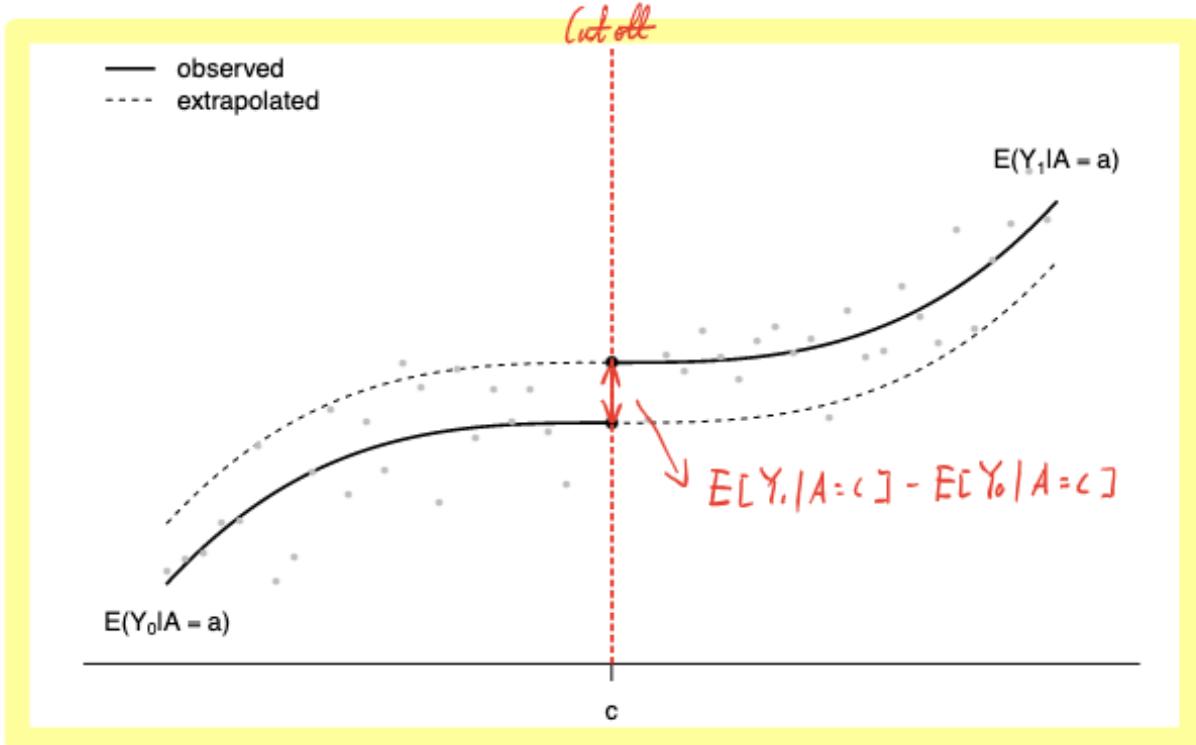
Treatment Assignment

- Sharp RD: Treatment Assignment: Treatment D_i is a *deterministic* function of A_i :

$$D_i = \begin{cases} 1, & \text{if } A_i \geq c \\ 0, & \text{if } A_i < c \end{cases}$$

- Extrapolation: we cannot perform matching on A_i because there is *no common support*: only Y_{1i} is observed if $A_i \geq c$ and only Y_{0i} is observed if $A_i < c$. We need to *extrapolate*

potential outcomes' dependence on A_i outside of group



Continuity Assumption

- **Continuity Assumption:** $\mathbb{E}[Y_1|A_i = a]$ and $\mathbb{E}[Y_0|A_i = a]$ are **continuous** functions of a at cutoff c .
- This **requires**:
 - D_i jumps at $A_i = c$
 - All factors determining Y_{1i}, Y_{0i} change continuously around c
 - **No manipulation** of A_i near the cutoff (individuals cannot precisely control A_i near the cutoff to intentionally get treated or not) \implies the distribution of A_i should be continuous at c
- This **implies**:
 - Variation in treatment D_i near the cut-off is random (**Conditional Random Assignment / Local RCT**)
 - Predetermined characteristics X_i should have the same distribution just to the left/right of the cutoff
 - We can use outcomes of individuals just to the left of c as counter-factual for individuals just to the right of c
- Check by **McCrory Test**:
 - Check **continuity of the density of the running variable at the cutoff** to examine manipulation
 - Check **continuity of other characteristics at the cutoff** to exclude the possibility that the change in outcome is caused by them

- It is not a formal test for the continuity assumption: even if those observed characteristics are continuous at the cutoff, they do not imply the continuity assumption is satisfied (there could be unobserved variables that affect Y_{1i}, Y_{0i} and jump at the cutoff)

Sharp RD Estimand and Treatment Effect

- Sharp RD Estimand:

$$\beta_{SD} = \lim_{\epsilon \downarrow 0} \left\{ \mathbb{E}[Y_i | A_i = c + \epsilon] - \mathbb{E}[Y_i | A_i = c - \epsilon] \right\}$$

- Under continuity assumption, Sharp RD estimand identifies **TE as the cutoff c** :

$$\beta_{SD} = \mathbb{E}[Y_{1i} - Y_{0i} | A_i = c]$$

- If TE is homogenous, then $\beta_{SD} = ATE = ATT$

Implementation

- Implementation
- Steps:

1. Specify 2 regression models for either side of the cutoff:

$$\begin{cases} Y_i = \alpha_l + f_l(c - A_i) + \epsilon_{li}, & \text{if } A_i < c \\ Y_i = \alpha_r + f_r(A_i - c) + \epsilon_{ri}, & \text{if } A_i \geq c \end{cases}$$

where $f_r(\cdot), f_l(\cdot)$ are some functions with $f_r(0) = 0, f_l(0) = 0$

2. Estimate α_l, α_r
3. RD Estimand: $\beta_{RD} = \alpha_r - \alpha_l$

- *Key choice*:

- Specification of $f_r(\cdot), f_l(\cdot)$
- Choice of **bandwidth**: how far away from the cutoff do we use data?

- Regression Specifications

- Regression Specifications

- Ideal local randomised experiment: $Y_i \perp A_i$ on either side of the cutoff:

$$\begin{cases} Y_i = \alpha_l + \epsilon_{li}, & \text{if } A_i < c \\ Y_i = \alpha_r + \epsilon_{ri}, & \text{if } A_i \geq c \end{cases}$$

- Smoothly contaminated local randomised experiment: Y_i depends on A_i at either side of the cutoff
- We then need to choose a linear slope, quadratic slope,..., or a Non-parametric function

Fuzzy RD

Setup

- Fuzzy RD Design
- Treatment D_i is still random given a value of A_i , but the probability of treatment given A_i jumps at cutoff c :

$$\lim_{\epsilon \downarrow 0} P(D_i = 1 | A_i = c + \epsilon) \neq \lim_{\epsilon \downarrow 0} P(D_i = 1 | A_i = c - \epsilon)$$

- Sharp RD is a special case of Fuzzy RD where the probability jumps from 0 to 1

Estimand

- Fuzzy RD Estimand:

$$\beta_{FD} = \frac{\lim_{\epsilon \downarrow 0} \left\{ \mathbb{E}[Y_i | A_i = c + \epsilon] - \mathbb{E}[Y_i | A_i = c - \epsilon] \right\}}{\lim_{\epsilon \downarrow 0} \left\{ P[D_i = 1 | A_i = c + \epsilon] - P[D_i = 1 | A_i = c - \epsilon] \right\}}$$

Fuzzy RD as IV

- Fuzzy RD as IV: Fuzzy RD can be interpreted as IV near the cutoff
- Instrument:

$$Z_i = \begin{cases} 1, & \text{if } A_i \geq c \\ 0, & \text{if } A_i < c \end{cases}$$

- Potential treatment:

$$D_i = \begin{cases} D_{1i}, & \text{if } Z_i = 1 \\ D_{0i}, & \text{if } Z_i = 0 \end{cases}$$

- For small $\epsilon > 0$, RD estimand can be *understood as a Wald estimand*, and identifies LATE:

$$\begin{aligned} \beta_{FD} &\approx \frac{\mathbb{E}[Y_i | A_i = c + \epsilon] - \mathbb{E}[Y_i | A_i = c - \epsilon]}{P[D_i = 1 | A_i = c + \epsilon] - P[D_i = 1 | A_i = c - \epsilon]} \\ &= \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]} \\ &= \mathbb{E}[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}] \end{aligned}$$

- Interpretation as LATE:

- Local: ATE for people who are
 - *At the cutoff* ($A_i = c$)
 - *Switch to treatment at the cutoff (compliers)*
 - Here, compliers are individuals who would not participate if c is just above A_i , but would participate if c is just below A_i

Practical Advice

- Motivate the validity of design -- why individuals cannot manipulate assignment variable A_i ?
 - Test validity of design -- check continuity of outcomes, covariates, and density of A_i around the cutoff
 - Show robustness of RD estimates with respect to specification of $f_r(\cdot)$, $f_l(\cdot)$ and choice of bandwidth
-

Others

- ECON0021 Week 7 Question
- Estimand is a population concept; Estimator is its finite sample analogue
- Inverse of a 2×2 matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

- Statistical independence:

$$\begin{aligned} A \perp B &\Leftrightarrow F_{A|B}(a|b) = F_A(a) \\ &\Leftrightarrow Pr(A < a|b) = Pr(A < a) \\ &\Leftrightarrow P(A = a, B = b) = P(A = a)P(B = b) \\ &\Leftrightarrow f_{AB}(a, b) = f_a(a)f_b(b) \\ &\Leftrightarrow f_{AB}(a, b) = f_A(a)f_B(b) \end{aligned}$$

- Conditional Expectation:

$$E[A|B] = \int_{-\infty}^{\infty} af_{A|B}(a|b) da$$

and $f_{A|B}(a|b) = \frac{f_{A,B}(a,b)}{f_B(b)}$

- Central Limit Theorem: Let $X_i, i = 1, \dots, n$ are random samples independently drawn from a distribution with mean μ and variance σ^2 . Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ be the sample average, then as $n \rightarrow \infty$:

$$\sqrt{n}(\bar{X}_n - \mu) \sim^d \mathcal{N}(0, \sigma^2)$$

- Confidence Interval (95%):

$$\left[-1.96 \frac{\sigma}{\sqrt{n}} + \bar{X}_n < \mu < 1.96 \frac{\sigma}{\sqrt{n}} + \bar{X}_n \right]$$

- OLS Estimand in Matrices:

$$\beta = E[X'X]^{-1}E[X'Y]$$

- OLS Estimator in Matrices:

$$\hat{\beta} = \left(\sum_{i=1}^n X_i X'_i \right)^{-1} \left(\sum_{i=1}^n X_i Y_i \right)$$

- pdf of $\mathcal{N}(0, \sigma^2)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

- Chain Rule for Vectors

$$\frac{\partial}{\partial \beta} F(x^\top \beta) = x F'(x^\top \beta); \frac{\partial}{\partial \beta^T} F(x^\top \beta) = x^\top F'(x^\top \beta)$$

- Jensen Inequality

- For $f(x) = x^2$, we have a nice proof:

$$Var(X) = E[X^2] - (E[X])^2 > 0 \implies E[X^2] > (E[X])^2$$

Revisions

- pdf of continuous distributed variable:

$$f_x(x) = Pr(x \in (x - \epsilon, x + \epsilon))$$

- conditional to unconditional pdf

$$Pr(X|Y) = \frac{Pr(X, Y)}{Pr(Y)} \rightsquigarrow (\text{apply to cdf and pdf})$$

- expectation as integrals

$$\mathbb{E}[X] = \begin{cases} \sum_k k \cdot Pr(X = k) & \text{for discrete X} \\ \rightarrow \int k \cdot f_X(x = k) dx & \text{as } k \rightarrow \infty \text{ for continuous X} \end{cases}$$

- conditional expectations

$$\mathbb{E}[X|Y] = \int k f_X(x = k) dk$$

- LIE in integrals

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \int \int x f_{X|Y}(k|y) dx f_Y dy$$

- *make brief explanations of the steps*
- *make a cheatsheet of the full MLE workflow of possible distributions*
- MLE: sample likelihood -> take log -> FOC with respect to parameters -> score -> Hessian -> asympt Var

- Asymptotic var:

$$AsympVar(X) = \sqrt{n} \cdot Var(x)$$