Recitation 3: Understanding Marginals

Jon Cohen

September 31, 2021
Recitation Takeaways

1. Marginal treatment effects as framework for policy extrapolation (applied to MH, AS, and selection on MH)
   - Accessible review article: Cornellison et al. (2016) Labour Economics

2. Alternative ways to characterize marginal compliers
   - Derive gap between marginal and average characteristics using regression equation Gruber, Levine, and Staiger (1999) QJE
   - Derive any functional of IV complier characteristics or potential outcomes Abadie (2003) JoE
Outline

Marginal Treatment Effects

Characteristics of the Marginals
MTE vs. LATE

- **Rough intuition**: MTE is the continuous version of the LATE

- **Usefulness**: Various treatment effects of interest—ATE, ATT, ATUT, LATE, etc.—can be expressed as averages of MTEs
  - Selection on gains for different IV’s deliver internally valid LATE that may not be useful for extrapolating to ATE

- **Notation**: Outcome $Y$, (binary endogenous) treatment $D$, instrument $Z$
  - E.g. $Y \equiv$ healthcare utilization, $D \equiv$ health insurance coverage, $Z \equiv$ (randomly assigned) insurance premium

- See Cornellison et al. (2016) Labour Economics for more details
Visual IV for LATE (Binary $Z$)

$E[Y|Z]$

$Z = 0$

Reduced form

Slope is IV Wald estimate (LATE)

$Z = 1$

1st stage

$E[D|Z]$
Visual IV for LATE (Non-binary $Z$)

See Figure 1 of Angrist (1990) AER on Vietnam draft lottery
Recasting Instrument as Revealing Unobservables

- Translate x-axis to propensity score: \( E[D|Z] = P(D = 1|Z) \equiv P(Z) \in [0, 1] \)
- \( Z \) traces out unobserved willingness to select into treatment
- Slope at a given point reveals *marginal* treatment effect at a given quantile of the willingness to select into treatment distribution
Visual IV for MTE (Non-binary $Z$)

Slopes are local IV estimates (MTE)
Understanding MTE’s Using Potential Outcomes

- Previous graphs showed outcomes of both $D = 1$ and $D = 0$ at each $Z$
- Instead, we can separately show (potential) outcomes for $D = 1$ and $D = 0$ by $Z$
  - **EFC**: $Y_1$ as utilization w/ insurance, $Y_0$ as utilization w/o insurance, $Z$ as randomly assigned price
  - See Brinch, Mogstad, Wiswall (2017) JPE for more details
No Selection and No Causal Effects

\[ E[Y|Z] \]

\[ E[D|Z] \equiv P(Z) \]

- "No MH or AS"
- \( ATE = ATT = ATUT = E[Y|D = 1] - E[Y|D = 0] = 0 \)
Causal Effects but No Selection on Levels or Slopes

\[ E[Y|Z] \]

\[ E[D|Z] \equiv P(Z) \]

- "MH but not AS"
- \[ ATE = ATT = ATUT = LATE = E[Y|D = 1] - E[Y|D = 0] \neq 0 \]
No Causal Effects but Selection on Levels

$E[Y|Z]$

$E[D|Z] \equiv P(Z)$

- “AS but not MH”
- $ATE = ATT = ATUT = LATE = 0 \neq E[Y|D = 1] - E[Y|D = 0]$
Causal Effect with Selection on Levels and Slopes

- "Selection on MH"
- $ATE \neq ATT, ATUT, LATE$ varies by $Z$

\[ E[Y|Z] \]

\[ E[D|Z] \equiv P(Z) \]
Aside: Estimating MTE’s with Binary $Z$

- Previous graphs suggest that you can implement MTE’s with a binary $Z$ assuming linearity of potential outcomes
  - Test with linearity assumption for $LATE \neq ATE$ is testing for unequal slopes by $D$
- More variation in $Z$ allows you to relax assumptions
Aside: Selection Bias Formula for ATE

• Formula for ATT should be familiar:

\[ E[Y_1|D = 1] - E[Y_0|D = 0] = E[Y_1 - Y_0|D = 1] + E[Y_0|D = 1] - E[Y_0|D = 0] \]

Observed diff. in outcomes \( \text{ATT} \) Selection bias

• ATE decomposition has additional term of treatment effect heterogeneity:

\[ E[Y_1|D = 1] - E[Y_0|D = 0] = E[Y_1 - Y_0] + \]

Observed diff. in outcomes \( \text{ATE} \)

\[ E[Y_0|D = 1] - E[Y_0|D = 0] + \]

Selection bias

\[ (1 - P(D = 1))(E[Y_1 - Y_0|D = 1] - E[Y_1 - Y_0|D = 0]) \]

Share untreated \( \text{ATT} \) \( \text{ATUT} \)

• See here for full derivation
Taking Stock

- Growing recognition of treatment effect heterogeneity
- MTEs provide a formal framework for:
  1. Aggregating heterogeneous treatment effects to policy-relevant parameters
  2. Considering how treatment effect heterogeneity interacts with selection into treatment
Outline

Marginal Treatment Effects

Characteristics of the Marginals
• **Specification**: \( c_i = \gamma + \delta p_i + u_i \)

• **Sample**: \( i \) who select into coverage at price \( p_i \) (of measure \( D(p) \))

• **Variation**: \( p_i \) randomly assigned

• **Intuition**: \( p_i \) has no causal effect on \( c_i \) so \( \delta \neq 0 \) is due to sample selection

• **Translating to marginal outcome**: Use chain rule to express marginal (costs at \( p \)) in terms of average (costs at \( p \)) and total number (\( D(p) \))
Gruber, Levine, and Staiger (1999): Gap Between Marginal and Average

- **Research question**: What is the impact of abortion on average living standards due to selection?

- **Specification**: \( O_{st}/B_{st} = \alpha \ln(B_{st}) + \text{controls} \)
  
  - E.g. \( O/B \equiv \% \) infants under FPL

  \[
  \alpha = \frac{\partial O/B}{\partial \ln(B)} 
  \]

  \[
  = B \frac{\partial O/B}{\partial B} 
  \]

  \[
  = \frac{\partial O}{\partial B} - \frac{O}{B}
  \]

- **Variation**: Instrument for state births \( B_{st} \) using abortion law repeal

- **Intuition**: Same as EFC
Abadie (2003): Extending LATE Theorem Logic

• Binary instrument $Z$, binary treatment $D$, outcome $Y$
• Potential outcomes $Y_{zd}$ and $D_z$ for $d \in \{0, 1\}$, $z \in \{0, 1\}$
• Standard IV assumptions:
  ■ Independence: $(Y_{00}, Y_{01}, Y_{10}, Y_{11} \perp Z)$
  ■ Exclusion: $P(Y_{1d} = Y_{0d}) = 1$ for $d \in \{0, 1\}$
  ■ 1st stage: $0 < P(Z = 1) < 1$ and $P(D_1) > P(D_0)$
  ■ Monotonicity: $P(D_1 \geq D_0) = 1$
Abadie’s $\kappa$ in words

- Split population into compliers ($C$), always-takers ($AT$), and never-takers ($NT$)
- Use law of total probability to decompose any observable into those for $C$, $AT$, $NT$
- Observables for $AT$ revealed by $(D, Z) = (1, 0)$ and $NT$ by $(D, Z) = (0, 1)$
- “Subtract off” $AT$ and $NT$ by reweighting based on realized $(D, Z)$
  - Applicable for any function $^1 g(\cdot)$ (e.g. quantile) applied to any observable (i.e. outcome $Y$, treatment $D$, or covariate $X$)
  - Applicable for complier $Y(1)$ [$Y(0)$] by subtracting off $AT$ [$NT$] from treated [untreated] outcomes
Abadie’s $\kappa$ in math

**Complier observables**

- Define $\kappa = 1 - \frac{D(1 - Z)}{P(Z = 0)} - \frac{(1 - D)Z}{P(Z = 1)}$
  
  \begin{align*}
  \text{Subtract off } AT & \quad \text{Subtract off } NT
  
  & \text{complier observables }
  \end{align*}

- $E[g(Y, D, X)|D_1 > D_0] = \frac{1}{P(D_1 > D_0)} E[\kappa g(Y, D, X)]$
  
  \begin{align*}
  \text{scale by size of } C & \quad \text{weight each observation }
  
  & \text{Complier Treated Potential Outcomes}
  
  \end{align*}

- Define $\kappa(1) = 1 - \frac{D}{P(Z = 0)} - \frac{(Z - P(Z = 1)}{P(Z = 0)P(Z = 1)}$
  
  \begin{align*}
  \text{NT get 0 weight } & \quad \text{AT get weight < 0 NT}
  
  & \text{Analogous for Complier Untreated Potential Outcomes}
  \end{align*}
Comparing Approaches

- **EFC**: Clear mapping to visual plots of potential outcomes
- **Gruber et al.**: Derivation from regression specification
- **Abadie**: Derivation from LATE theorem logic
  - Powerful to be able to estimate any function of potential outcomes (and therefore treatment effects on those functions)