

# Frontier Topics in Empirical Economics: Week 6

## IV beyond LATE

Zibin Huang<sup>1</sup>

<sup>1</sup>College of Business, Shanghai University of Finance and Economics

November 30, 2023

# IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlačil, 2007a,b)
  - it relies on binary treatment and binary IV
  - it is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlačil, 2007a,b)
  - It relies on binary treatment and binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlacil, 2007a,b)
  - It relies on binary treatment and binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlacil, 2007a,b)
  - It relies on binary treatment and binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlacil, 2007a,b)
  - It relies on binary treatment and binary IV
    - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlačil, 2007a,b)
  - It relies on binary treatment and binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlačil, 2007a,b)
  - It relies on binary treatment and binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes



## IV beyond LATE: Limitation of LATE

- We have introduced LATE interpretation of IV
- This is the most popular way to think of IV under heterogeneous treatment effect
- It is elegant, policy-relevant, but also limited (Heckman and Vytlačil, 2007a,b)
  - It relies on binary treatment and binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When environment changes, complier group changes

## IV beyond LATE: Limitation of LATE

- In this lecture, we are going to do two things
- First, we relax the assumption of binary treatment, single and binary IV
- To generalize LATE interpretation in its original framework
- Second, we introduce a more general framework with better external validity:  
Marginal Treatment Effect (MTE)
- We are going to see how choice model can be incorporated into IV

## IV beyond LATE: Limitation of LATE

- In this lecture, we are going to do two things
  - First, we relax the assumption of binary treatment, single and binary IV
  - To generalize LATE interpretation in its original framework
  - Second, we introduce a more general framework with better external validity: Marginal Treatment Effect (MTE)
  - We are going to see how choice model can be incorporated into IV

## IV beyond LATE: Limitation of LATE

- In this lecture, we are going to do two things
- First, we relax the assumption of binary treatment, single and binary IV
- To generalize LATE interpretation in its original framework
- Second, we introduce a more general framework with better external validity: Marginal Treatment Effect (MTE)
- We are going to see how choice model can be incorporated into IV

## IV beyond LATE: Limitation of LATE

- In this lecture, we are going to do two things
- First, we relax the assumption of binary treatment, single and binary IV
- To generalize LATE interpretation in its original framework
- Second, we introduce a more general framework with better external validity: Marginal Treatment Effect (MTE)
- We are going to see how choice model can be incorporated into IV

## IV beyond LATE: Limitation of LATE

- In this lecture, we are going to do two things
- First, we relax the assumption of binary treatment, single and binary IV
- To generalize LATE interpretation in its original framework
- Second, we introduce a more general framework with better external validity: Marginal Treatment Effect (MTE)
- We are going to see how choice model can be incorporated into IV

## IV beyond LATE: Limitation of LATE

- In this lecture, we are going to do two things
- First, we relax the assumption of binary treatment, single and binary IV
- To generalize LATE interpretation in its original framework
- Second, we introduce a more general framework with better external validity: Marginal Treatment Effect (MTE)
- We are going to see how choice model can be incorporated into IV

# IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to discuss how to use choice model and economic theory to regularize IV
- An interaction between design-based approach and structural approach



# IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to discuss how to use choice model and economic theory to regularize IV
- An interaction between design-based approach and structural approach

# IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
  - You are thinking about these people's choices under different policy shocks
  - This choice structure is not fully utilized in pure design-based approach
  - It can definitely help you when data is not enough to identify the effect
  - The whole point of this lecture is to discuss how to use choice model and economic theory to regularize IV
  - An interaction between design-based approach and structural approach

## IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to discuss how to use choice model and economic theory to regularize IV
- An interaction between design-based approach and structural approach

## IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to discuss how to use choice model and economic theory to regularize IV
- An interaction between design-based approach and structural approach

## IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to discuss how to use choice model and economic theory to regularize IV
- An interaction between design-based approach and structural approach

# IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to **discuss how to use choice model and economic theory to regularize IV**
- An interaction between design-based approach and structural approach

## IV beyond LATE: Choice Model and IV

- Choice model is intrinsically nested in IV
- When you consider always-taker, complier, never-taker
- You are thinking about these people's choices under different policy shocks
- This choice structure is not fully utilized in pure design-based approach
- It can definitely help you when data is not enough to identify the effect
- The whole point of this lecture is to **discuss how to use choice model and economic theory to regularize IV**
- An interaction between design-based approach and structural approach

## IV beyond LATE: Choice Model and IV

- You have already used it in LATE Theorem: Monotonicity
- The idea of monotonicity comes from assuming treatment is a normal good
- If the agent chooses something when the price is higher ( $D|Z = 0 = 1$ )
- Then he/she will definitely choose it when the price is lower ( $D|Z = 1 = 1$ )



## IV beyond LATE: Choice Model and IV

- You have already used it in LATE Theorem: Monotonicity
- The idea of monotonicity comes from assuming treatment is a normal good
- If the agent chooses something when the price is higher ( $D|Z = 0 \rightarrow 1$ )
- Then he/she will definitely choose it when the price is lower ( $D|Z = 1 \rightarrow 1$ )

## IV beyond LATE: Choice Model and IV

- You have already used it in LATE Theorem: Monotonicity
- The idea of monotonicity comes from assuming treatment is a normal good
- If the agent chooses something when the price is higher ( $D|Z = 0 \rightarrow 1$ )
- Then he/she will definitely choose it when the price is lower ( $D|Z = 1 \rightarrow 1$ )

## IV beyond LATE: Choice Model and IV

- You have already used it in LATE Theorem: Monotonicity
- The idea of monotonicity comes from assuming treatment is a normal good
- If the agent chooses something when the price is higher ( $D|Z = 0 \rightarrow 1$ )
- Then he/she will definitely choose it when the price is lower ( $D|Z = 1 \rightarrow 1$ )

## IV beyond LATE: Choice Model and IV

- You have already used it in LATE Theorem: Monotonicity
- The idea of monotonicity comes from assuming treatment is a normal good
- If the agent chooses something when the price is higher ( $D = 0$ )
- Then he/she will definitely choose it when the price is lower ( $D = 1$ )

# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you  $2 \times 2 = 4$  types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations (final nodes)
- LATE can be inverted from expectation functions from the four final nodes
- It can be identified by the IV estimator

# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you 2  $\times$  2 = 4 types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations ( 4 nodes)
- LATE can be inverted from expectation functions from the four nodes
- It can be identified by the IV estimator

# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you 2  $\times$  2 = 4 types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations ( 4 nodes)
- LATE can be inverted from expectation functions from the four nodes
- It can be identified by the IV estimator

# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you  $2 \times 2 = 4$  types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations ( 4 nodes)
- LATE can be inverted from expectation functions from the four nodes
- It can be identified by the IV estimator



# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you  $2 \times 2 = 4$  types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations ( 4 nodes)
- LATE can be inverted from expectation functions from the four nodes
- It can be identified by the IV estimator

# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you  $2 \times 2 = 4$  types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations ( 4 nodes)
- LATE can be inverted from expectation functions from the four nodes
- It can be identified by the IV estimator

# Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you 2  $\times$  2 = 4 types of people (A,C,N,D)
- By assuming monotonicity, we eliminate D

- We have four equations ( 4 nodes)
- LATE can be inverted from expectation functions from the four nodes
- It can be identified by the IV estimator

# Generalization of LATE: Multiple IV

- What if IV and treatment are multiple or multivalued?
- It will be a complicated weighted average of different TEs for different types (groups)

# Generalization of LATE: Multiple IV

- What if IV and treatment are multiple or multivalued?
- It will be a complicated weighted average of different TEs for different types (groups)

# Generalization of LATE: Multiple IV

- What if IV and treatment are multiple or multivalued?
- It will be a complicated weighted average of different TEs for different types (groups)

# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1, z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$\tau_{2SLS} = \frac{LATE_1}{1} + \frac{LATE_2}{1}$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$

# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1, z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$\tau_{2SLS} = \frac{LATE_1}{1} + \frac{LATE_2}{1}$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$



# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1, z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$\tau_{2SLS} = \frac{LATE_1}{1} = LATE_2$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$

# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1, z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$\tau_{2SLS} = \frac{LATE_1}{1} + \frac{LATE_2}{1}$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$

# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1; z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$\tau_{2SLS} = \frac{LATE_1 - 1}{LATE_2}$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$

# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1, z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$2SLS \quad LATE_1 \quad 1 \quad LATE_2$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$

# Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- This is relatively simple
- We run regressions taking  $z_1, z_2$  as instruments (not  $z$ )
- Assuming monotonicity for both  $z_1$  and  $z_2$
- The corresponding IV estimator can be derived as:

$$\tau_{2SLS} = \frac{LATE_1 - 1}{LATE_2}$$

- $LATE_1; LATE_2$  are LATEs for instrument  $z_1$  and  $z_2$

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $x \in \{0, 1, 2, \dots, s_x\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0,x}, Y_{1,x}, \dots, Y_{s_x,x} \perp D$
  - ACR2 Full rank:  $\text{rank}(E[D \cdot x]) = 1$
  - ACR3 Monotonicity:  $E[D \cdot x] \geq E[D \cdot (x-1)]$
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $s \in \{0, 1, 2, \dots, S\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{Si}; s_{0i}, s_{1i} \perp z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  &  $0 \leq i < n$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $s \in \{0, 1, 2, \dots, S\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{Si}; s_{0i}, s_{1i} \perp z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment



# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $s \in \{0, 1, 2, \dots, S\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{Si}; s_{0i}, s_{1i} \perp Z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $s \in \{0, 1, 2, \dots, S\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{Si} \perp s_{0i}, s_{1i} \mid X_i = z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $s \in \{0, 1, 2, \dots, S\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{Si}; S_{0i}, S_{1i} \perp Z_i$
  - ACR2 First stage existence:  $E[S_{1i} - S_{0i}] > 0$
  - ACR3 Monotonicity:  $S_{1i} \geq S_{0i}$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatments  $s = 0, 1, 2, \dots, s_x$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{s_i}, s_{0i}, s_{1i} \perp z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatments  $s = 0, 1, 2, \dots, s_x$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{s_i}, s_{0i}, s_{1i} \perp z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  &  $0 \leq z_i$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatments  $s \in \{0, 1, 2, \dots, s_x\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{s_i}, s_{0i}, s_{1i} \perp z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  or vice versa
- ACR3 implicitly requires us to have an "ordered" list of values for treatment

# Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment  $s \in \{0, 1, 2, \dots, S\}$
- For example, IV is the implementation of a compulsory education law
- Treatment is the education level, which takes multiple values
- We have the following three assumptions:
  - ACR1 Independence:  $Y_{0i}, Y_{1i}, \dots, Y_{Si}; s_{0i}, s_{1i} \perp z_i$
  - ACR2 First stage existence:  $E[s_{1i} - s_{0i}] > 0$
  - ACR3 Monotonicity:  $s_{1i} \geq s_{0i}$  or vice versa
- ACR3 implicitly requires us to have an **"ordered" list of values for treatment**

# Generalization of LATE: Multivalued Treatment

- Under ACR1-3, IV identifies a weighted average of the unit causal response

When ACR1, ACR2, and ACR3 hold, we have:

$$E(Y_1) - E(Y_0) = \frac{E(Y_1 | Z=1) - E(Y_1 | Z=0)}{E(Z)} - \frac{E(Y_0 | Z=1) - E(Y_0 | Z=0)}{E(Z)}$$

$$= \frac{E(Y_1 | Z=1) - E(Y_0 | Z=1)}{E(Z)} - \frac{E(Y_1 | Z=0) - E(Y_0 | Z=0)}{E(Z)}$$



# Generalization of LATE: Multivalued Treatment

- Under ACR1-3, IV identifies a weighted average of the unit causal response

Theorem 4.5.3 in MHE

When ACR1, ACR2, and ACR3 hold, we have:

$$\frac{E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)}{E(S_i | Z_i = 1) - E(S_i | Z_i = 0)} = \sum_{s=1}^S \pi_s \frac{E(Y_{si} | S_i = s) - E(Y_{si} | S_i = 0)}{P(S_i = s) - P(S_i = 0)}$$

where  $\pi_s = \frac{P(S_i = s) - P(S_i = 0)}{\sum_{j=1}^S P(S_i = j) - P(S_i = 0)}$

# Generalization of LATE: Multivalued Treatment

- Under ACR1-3, IV identifies a weighted average of the unit causal response

## Theorem 4.5.3 in MHE

When ACR1, ACR2, and ACR3 hold, we have:

$$\frac{E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)}{E(S_i | Z_i = 1) - E(S_i | Z_i = 0)} = \sum_{s=1}^S E(Y_{si} - Y_{s-1,i} | S_i = s) \pi_{s0i};$$

where  $\pi_{s0i} = \frac{P(S_i = s | Z_i = 0)}{\sum_{j=1}^S P(S_i = j | Z_i = 0)}$

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1,i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s=2$  includes compliers
  - who choose  $z=0$  when  $z=0$ , but choose  $z=2$  (vs  $z=1$ ) when  $z=1$
  - who choose  $z=1$  when  $z=0$ , but choose  $z=2$  (vs  $z=1$ ) when  $z=1$
- We then average over all units/steps with a weight
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s$  includes compliers
  - who chooses 0 when  $z=0$ , but chooses 2; 3; ...;  $s$  when  $z=1$
  - who chooses 1 when  $z=0$ , but chooses 2; 3; ...;  $s$  when  $z=1$
- We then average over all units/steps with a weight  $\pi_s$
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s$  includes compliers
  - who chooses 0 when  $z=0$ , but chooses 2;3;...;s when  $z=1$
  - who chooses 1 when  $z=0$ , but chooses 2;3;...;s when  $z=1$
- We then average over all units/steps with a weight
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s = 1$  to  $s = 2$  includes compliers
  - who chooses 0 when  $z = 0$ , but chooses 2; 3; ...;  $s$  when  $z = 1$
  - who chooses 1 when  $z = 0$ , but chooses 2; 3; ...;  $s$  when  $z = 1$
- We then average over all units/steps with a weight  $\pi_s$
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s = 1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s$  includes compliers
  - who chooses  $0$  when  $z = 0$ , but chooses  $2; 3; \dots; s$  when  $z = 1$
  - who chooses  $1$  when  $z = 0$ , but chooses  $2; 3; \dots; s$  when  $z = 1$
- We then average over all units/steps with a weight  $\pi_s$
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over  $s$

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s$  includes compliers
  - who chooses  $s-1$  when  $z=0$ , but chooses  $s$  when  $z=1$
  - who chooses  $s-1$  when  $z=0$ , but chooses  $s$  when  $z=1$
- We then average over all units/steps with a weight  $\pi_s$
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over  $s$



# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s = 1$  to  $s = 2$  includes compliers
  - who chooses  $0$  when  $z = 0$ , but chooses  $2; 3; \dots; s$  when  $z = 1$
  - who chooses  $1$  when  $z = 0$ , but chooses  $2; 3; \dots; s$  when  $z = 1$
- We then average over all units/steps with a weight
  - $\pi_s$  is the proportion of compliers involved in this unit change from  $s = 1$  to  $s$
  - It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s$  includes compliers
  - who chooses  $0$  when  $z=0$ , but chooses  $2; 3; \dots; s$  when  $z=1$
  - who chooses  $1$  when  $z=0$ , but chooses  $2; 3; \dots; s$  when  $z=1$
- We then average over all units/steps with a weight
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- $Y_{si} - Y_{s-1;i}$  is the unit response, or stepwise treatment effect
- For each unit/step change, we average over all compliers that cover this unit/step
- For instance, the unit change from  $s-1$  to  $s$  includes compliers
  - who chooses  $s-1$  when  $z=0$ , but chooses  $s$  when  $z=1$
  - who chooses  $s-1$  when  $z=0$ , but chooses  $s$  when  $z=1$
- We then average over all units/steps with a weight
- $\pi_s$  is the proportion of compliers involved in this unit change from  $s-1$  to  $s$
- It is a normalization, with  $\pi_s$  summing up to 1 over

# Generalization of LATE: Multivalued Treatment

- Interpretation for models with multiple instruments and multivalued treatment is the combination of the previous two cases
- A weighted average of the ACR for each instrument
- Of course the weighting scheme could be very complicated
- Weighted average of some weighted averages...

# Generalization of LATE: Multivalued Treatment

- Interpretation for models with multiple instruments and multivalued treatment is the combination of the previous two cases
- A weighted average of the ACR for each instrument
- Of course the weighting scheme could be very complicated
- Weighted average of some weighted averages...

# Generalization of LATE: Multivalued Treatment

- Interpretation for models with multiple instruments and multivalued treatment is the combination of the previous two cases
- A weighted average of the ACR for each instrument
- Of course the weighting scheme could be very complicated
- Weighted average of some weighted averages...

# Generalization of LATE: Multivalued Treatment

- Interpretation for models with multiple instruments and multivalued treatment is the combination of the previous two cases
- A weighted average of the ACR for each instrument
- Of course the weighting scheme could be very complicated
- Weighted average of some weighted averages...

# Generalization of LATE: Multivalued Treatment

- Interpretation for models with multiple instruments and multivalued treatment is the combination of the previous two cases
- A weighted average of the ACR for each instrument
- Of course the weighting scheme could be very complicated
- Weighted average of some weighted averages...



# Generalization of LATE: Multivalued IV

- We have already checked multiple IV and multivalued treatment cases
- We now consider multivalued IV
- This can be a little bit tricky
- There is a simple first thought of this issue: transit multivalued IV to multiple IV

# Generalization of LATE: Multivalued IV

- We have already checked multiple IV and multivalued treatment cases
- We now consider multivalued IV
- This can be a little bit tricky
- There is a simple first thought of this issue: transit multivalued IV to multiple IV

# Generalization of LATE: Multivalued IV

- We have already checked multiple IV and multivalued treatment cases
- We now consider multivalued IV
- This can be a little bit tricky
- There is a simple first thought of this issue: transit multivalued IV to multiple IV

# Generalization of LATE: Multivalued IV

- We have already checked multiple IV and multivalued treatment cases
- We now consider multivalued IV
- This can be a little bit tricky
- There is a simple first thought of this issue: transit multivalued IV to multiple IV

# Generalization of LATE: Multivalued IV

- We have already checked multiple IV and multivalued treatment cases
- We now consider multivalued IV
- This can be a little bit tricky
- There is a simple first thought of this issue: transit multivalued IV to multiple IV

# Generalization of LATE: Multivalued IV

- We can first decompose the multivalued IV to multiple dummies
- Each dummy represents a specific value of IV
- For example, if  $z \in \{0, 1, 2\}$ , we have dummies  $z_1, z_2$  as indicators
- $z_1 = 1$  if  $z = 1$ ;  $z_1 = 0$  if  $z = 0, 2$
- $z_2 = 1$  if  $z = 2$ ;  $z_2 = 0$  if  $z = 0, 1$

# Generalization of LATE: Multivalued IV

- We can first decompose the multivalued IV to multiple dummies
- Each dummy represents a specific value of IV
- For example, if  $z \in \{0, 1, 2\}$ , we have dummies  $z_1, z_2$  as indicators
- $z_1 = 1$  if  $z = 1$ ;  $z_1 = 0$  if  $z \in \{0, 2\}$
- $z_2 = 1$  if  $z = 2$ ;  $z_2 = 0$  if  $z \in \{0, 1\}$

# Generalization of LATE: Multivalued IV

- We can first decompose the multivalued IV to multiple dummies
- Each dummy represents a specific value of IV
- For example, if  $z \in \{0, 1, 2\}$ , we have dummies  $z_1, z_2$  as indicators
- $z_1 = 1$  if  $z = 1$ ;  $z_1 = 0$  if  $z \in \{0, 2\}$
- $z_2 = 1$  if  $z = 2$ ;  $z_2 = 0$  if  $z \in \{0, 1\}$



# Generalization of LATE: Multivalued IV

- We can first decompose the multivalued IV to multiple dummies
- Each dummy represents a specific value of IV
- For example, if  $z \in \{0, 1, 2\}$ , we have dummies  $z_1, z_2$  as indicators
  - $z_1 = 1$  if  $z = 1$ ;  $z_1 = 0$  if  $z \in \{0, 2\}$
  - $z_2 = 1$  if  $z = 2$ ;  $z_2 = 0$  if  $z \in \{0, 1\}$

# Generalization of LATE: Multivalued IV

- We can first decompose the multivalued IV to multiple dummies
- Each dummy represents a specific value of IV
- For example, if  $z \in \{0, 1, 2\}$ , we have dummies  $z_1, z_2$  as indicators
- $z_1 = 1$  if  $z = 1$ ;  $z_1 = 0$  if  $z \in \{0, 2\}$
- $z_2 = 1$  if  $z = 2$ ;  $z_2 = 0$  if  $z \in \{0, 1\}$

# Generalization of LATE: Multivalued IV

- We can first decompose the multivalued IV to multiple dummies
- Each dummy represents a specific value of IV
- For example, if  $z \in \{0, 1, 2\}$ , we have dummies  $z_1, z_2$  as indicators
- $z_1 = 1$  if  $z = 1$ ;  $z_1 = 0$  if  $z \in \{0, 2\}$
- $z_2 = 1$  if  $z = 2$ ;  $z_2 = 0$  if  $z \in \{0, 1\}$

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i(z_1 = 0) < D_i(z_1 = 1) > D_i(z_1 = 2)$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i(z_1 = 0) > D_i(z_1 = 1) > D_i(z_1 = 2)$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i | z_i = 0 \leq D_i | z_i = 1 \leq D_i | z_i = 2$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i | z_i = 0 \leq D_i | z_i = 1 \leq D_i | z_i = 2$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
  - However, it is not true for  $z_1$
  - Because for the group of people with  $z_1 = 1$
  - They can be either  $z = 0$  or  $z = 2$
  - It is possible that  $D_i(z_1 = 0) > D_i(z_1 = 1) > D_i(z_1 = 2)$
  - Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption



# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$ 
  - Because for the group of people with  $z_1 = 1$
  - They can be either  $z = 0$  or  $z = 2$
  - It is possible that  $D_i(z_1 = 0) < D_i(z_1 = 1) < D_i(z_1 = 2)$
  - Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$ 
  - They can be either  $z = 0$  or  $z = 2$
  - It is possible that  $D_i(z_1 = 0) > D_i(z_1 = 1) > D_i(z_1 = 2)$
  - Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i | z_i = 0 \leq D_i | z_i = 1 \leq D_i | z_i = 2$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i | z_i = 0 \leq D_i | z_i = 1 \leq D_i | z_i = 2$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Then we run the regression using the set of dummies  $z_1, z_2$  as instruments
- We can interpret the results as in multiple IV case
- But is this correct?
- An important assumption is monotonicity for each dummy IV
- However, it is not true for  $z_1$
- Because for the group of people with  $z_1 = 1$
- They can be either  $z = 0$  or  $z = 2$
- It is possible that  $D_i(z_1 = 0) > D_i(z_1 = 1) > D_i(z_1 = 2)$
- Then, for  $z_1 = 1$ , some people go to one direction ( $z = 2$ ), some people go to the other, violating the monotonicity assumption

# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

The weak axiom of revealed preference (WARP) states that if  $x$  is chosen in situation  $(w, p)$  and  $y$  is affordable in situation  $(w, p)$ , then  $y$  is not chosen in situation  $(w, p)$ .

$$\text{If } x \in \mathcal{C}(w, p) \text{ and } y \in \mathcal{A}(w, p) \text{ then } y \notin \mathcal{C}(w, p)$$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $\mathcal{B}_A(\cdot)_{R, X_B}$
- This is Weak Axiom of Revealed Preference

# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

## WARP Definition 2.F.1 MWG

The Walrasian demand function  $p; w$  satisfies the weak axiom of revealed preference if the following holds for any two price wealth situations  $p; w$  ;  $p'; w'$  :

If  $p \preceq p'; w'$  &  $w; \text{ and } x \in p'; w' \setminus x \in p; w$  ; then  $p' \preceq p; w \neq w'$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $B_A(\cdot) \in \mathbb{R} \times X_B$
- This is Weak Axiom of Revealed Preference

# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

## WARP Definition 2.F.1 MWG

The Walrasian demand function  $p; w$  satisfies the weak axiom of revealed preference if the following holds for any two price wealth situations  $p; w$  ;  $p'; w'$  :

If  $p \preceq p'; w'$  &  $w; \text{ and } x \in p'; w' \setminus x \in p; w$  ; then  $p' \preceq p; w \neq w'$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $\mathbb{B}_A(\cdot) \in \mathbb{R} \times \mathbb{B}$
- This is Weak Axiom of Revealed Preference



# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

## WARP Definition 2.F.1 MWG

The Walrasian demand function  $x(p; w)$  satisfies the weak axiom of revealed preference if the following holds for any two price wealth situations  $(p; w)$  ;  $(p'; w')$  :

If  $x(p; w) \in X(p'; w')$  &  $x(p'; w') \notin X(p; w)$  ; then  $p' \cdot x(p; w) > p' \cdot w'$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $B_A(\cdot) \in X_B$
- This is Weak Axiom of Revealed Preference

# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

## WARP Definition 2.F.1 MWG

The Walrasian demand function  $x(p; w)$  satisfies the weak axiom of revealed preference if the following holds for any two price wealth situations  $(p; w)$  ;  $(p'; w')$  :

If  $p \succ p'; w'$  &  $w; \text{ and } x(p'; w') \in x(p; w)$  ; then  $p' \succ p; w \succ w'$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $B_A(\cdot) \in \mathbb{R} \times X_B$
- This is Weak Axiom of Revealed Preference

# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

## WARP Definition 2.F.1 MWG

The Walrasian demand function  $x(p; w)$  satisfies the weak axiom of revealed preference if the following holds for any two price wealth situations  $(p; w)$  ;  $(p'; w')$  :

If  $p \preceq p'; w' & w$ ; and  $x(p'; w') \succ x(p; w)$ ; then  $p' \preceq p; w \succ w'$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $B_A(\cdot) \in \mathbb{R}^n \times \mathbb{R}$
- This is Weak Axiom of Revealed Preference

# Generalization of LATE: Multivalued IV

- Thus, monotonicity assumption is not as innocuous as in the 2 case
- We need to go to deep choice structure of this assumption:  
Axiom of Revealed Preference
- In this case, you have to analyze one by one based on your specific context

## WARP Definition 2.F.1 MWG

The Walrasian demand function  $x(p; w)$  satisfies the weak axiom of revealed preference if the following holds for any two price wealth situations  $(p; w)$  ;  $(p'; w')$  :

If  $p \succ p'; w'$  &  $w; \text{ and } x(p'; w') \in x(p; w)$  ; then  $p' \succ p; w \text{ \% } w'$

- If some optimal bundle in situation B is also feasible but not chosen in situation A, then the optimal bundle in situation A is not feasible in situation  $B_A(\cdot) \in \mathbb{R}^n \times \mathbb{R}$
- This is Weak Axiom of Revealed Preference

# Generalization of LATE: Multivalued IV

- A stronger version of WARP is SARP

The market demand function  $p(w)$  satisfies the SARP axiom if revealed preferences  $\succsim$  for any lot of  $(p^1, w^1) \succsim (p^2, w^2)$

with  $x(p^1, w^1) \succ x(p^2, w^2)$  for all  $n \in N$  imply

we have  $x(p^1, w^1) \succ x(p^2, w^2)$  whenever  $(p^1, w^1) \succ (p^2, w^2)$  for all  $n \in N$ .

- SARP adds transitivity to WARP
- If  $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2}; \dots; x_2 \succ_R x_1$ , we have  $x_N \succ_R x_1$
- Let's go to the example of MTO in Pinto (2015)

# Generalization of LATE: Multivalued IV

- A stronger version of WARP is SARP

## SARP Definition 3.J.1 MWG

The market demand function  $x(p; w)$  satisfies the strong axiom of revealed preference if for any list of  $(p^1; w^1); \dots; (p^N; w^N)$

with  $x(p^{n-1}; w^{n-1}) \succ_j x(p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

we have  $(p^N; w^N) \succ (p^1; w^1) \succ (p^N; w^N)$ ; whenever  $(p^N; w^N) \succ (p^{n-1}; w^{n-1})$  &  $(p^{n-1}; w^{n-1}) \succ (p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

- SARP adds transitivity to WARP
- If  $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2}; \dots; x_2 \succ_R x_1$ , we have  $x_N \succ_R x_1$
- Let's go to the example of MTO in Pinto (2015)

# Generalization of LATE: Multivalued IV

- A stronger version of WARP is SARP

## SARP Definition 3.J.1 MWG

The market demand function  $x(p; w)$  satisfies the strong axiom of revealed preference if for any list of  $(p^1; w^1); \dots; (p^N; w^N)$

with  $x(p^{n-1}; w^{n-1}) \succ_j x(p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

we have  $(p^N; w^N) \succ (p^1; w^1)$  whenever  $(p^N; w^N) \succ (p^{n-1}; w^{n-1})$  &  $(p^{n-1}; w^{n-1}) \succ (p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

- SARP adds transitivity to WARP
- If  $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2}; \dots; x_2 \succ_R x_1$ , we have  $x_N \succ_R x_1$
- Let's go to the example of MTO in Pinto (2015)

# Generalization of LATE: Multivalued IV

- A stronger version of WARP is SARP

## SARP Definition 3.J.1 MWG

The market demand function  $x(p; w)$  satisfies the strong axiom of revealed preference if for any list of  $(p^1; w^1); \dots; (p^N; w^N)$

with  $x(p^{n-1}; w^{n-1}) \succ_j x(p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

we have  $(p^N; w^N) \succ (p^1; w^1)$  whenever  $x(p^n; w^n) \succ x(p^{n-1}; w^{n-1})$  &  $w^n$  for all  $n \in \{1, \dots, N\}$

- SARP adds transitivity to WARP
- If  $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2}; \dots; x_2 \succ_R x_1$ , we have  $x_N \succ_R x_1$
- Let's go to the example of MTO in Pinto (2015)



# Generalization of LATE: Multivalued IV

- A stronger version of WARP is SARP

## SARP Definition 3.J.1 MWG

The market demand function  $x(p; w)$  satisfies the strong axiom of revealed preference if for any list of  $(p^1; w^1); \dots; (p^N; w^N)$

with  $x(p^{n-1}; w^{n-1}) \succ_j x(p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

we have  $(p^N; w^N) \succ (p^1; w^1)$  whenever  $(p^N; w^N) \succ (p^{n-1}; w^{n-1})$  &  $(p^{n-1}; w^{n-1}) \succ (p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

- SARP adds transitivity to WARP
- If  $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2}; \dots; x_2 \succ_R x_1$ , we have  $x_N \succ_R x_1$
- Let's go to the example of MTO in Pinto (2015)

# Generalization of LATE: Multivalued IV

- A stronger version of WARP is SARP

## SARP Definition 3.J.1 MWG

The market demand function  $x(p; w)$  satisfies the strong axiom of revealed preference if for any list of  $(p^1; w^1); \dots; (p^N; w^N)$

with  $x(p^{n-1}; w^{n-1}) \succ x(p^n; w^n)$  for all  $n \in \{1, \dots, N\}$

we have  $(p^N; w^N) \succ (p^1; w^1)$  whenever  $(p^N; w^N) \succ x(p^{n-1}; w^{n-1})$  &  $w^n$  for all  $n \in \{1, \dots, N\}$

- SARP adds transitivity to WARP
- If  $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2}; \dots; x_2 \succ_R x_1$ , we have  $x_N \succ_R x_1$
- Let's go to the example of MTO in Pinto (2015)

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood (1)
  - Section 8 group: Vouchers, available for any housing lease arrangement (2)
- There are three choices (three values of treatment)
  - Not relocating (0)
  - Relocating to a low poverty neighborhood (1)
  - Relocating to a high poverty neighborhood (2)

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $t_1$ )
  - Relocating to a low poverty neighborhood ( $t_2$ )
  - Relocating to a high poverty neighborhood ( $t_3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $t_1$ )
  - Relocating to a low poverty neighborhood ( $t_2$ )
  - Relocating to a high poverty neighborhood ( $t_3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $t_1$ )
  - Relocating to a low poverty neighborhood ( $t_2$ )
  - Relocating to a high poverty neighborhood ( $t_3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $Z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $Z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $Z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $T = 1$ )
  - Relocating to a low poverty neighborhood ( $T = 2$ )
  - Relocating to a high poverty neighborhood ( $T = 3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $t_1$ )
  - Relocating to a low poverty neighborhood ( $t_2$ )
  - Relocating to a high poverty neighborhood ( $t_3$ )



# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $t = 1$ )
  - Relocating to a low poverty neighborhood ( $t = 2$ )
  - Relocating to a high poverty neighborhood ( $t = 3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $Z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $Z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $Z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $T = 1$ )
  - Relocating to a low poverty neighborhood ( $T = 2$ )
  - Relocating to a high poverty neighborhood ( $T = 3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $Z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $Z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $Z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $T = 1$ )
  - Relocating to a low poverty neighborhood ( $T = 2$ )
  - Relocating to a high poverty neighborhood ( $T = 3$ )

# Generalization of LATE: Multivalued IV

- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers ( $Z_1$ )
  - Experimental group: Vouchers, available only for housing lease in low poverty neighborhood ( $Z_2$ )
  - Section 8 group: Vouchers, available for any housing lease anywhere ( $Z_3$ )
- There are three choices (three values of treatment)
  - Not relocating ( $T = 1$ )
  - Relocating to a low poverty neighborhood ( $T = 2$ )
  - Relocating to a high poverty neighborhood ( $T = 3$ )

# Generalization of LATE: Multivalued IV

# Generalization of LATE: Multivalued IV

- Thus, we have  $3 \times 3 \times 3 = 27$  types of agents
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

# Generalization of LATE: Multivalued IV

- Thus, we have  $3 \times 3 \times 3 = 27$  types of agents
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

# Generalization of LATE: Multivalued IV

- Thus, we have  $3 \times 3 \times 3 = 27$  types of agents
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP



# Generalization of LATE: Multivalued IV

- Thus, we have  $3 \times 3 \times 3 = 27$  types of agents
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

# Generalization of LATE: Multivalued IV

- Thus, we have  $3 \times 3 \times 3 = 27$  types of agents
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

# Generalization of LATE: Multivalued IV

- Let  $u_i(k; t)$  be the utility function of family  $i$  ( $k$ : consumption,  $t$ : relocation choice)
- Let  $W_i(z; t)$  be the budget set of family  $i$  under relocation decision  $t \in \{1, 2, 3\}$  and MTO voucher  $z \in \{z_1, z_2, z_3\}$
- Let  $S_i = \{C_i(z_1); C_i(z_2); C_i(z_3)\}$  denote the type of family  $i$ , defined by relocation response  $S_i(z)$  given different vouchers

# Generalization of LATE: Multivalued IV

- Let  $u_i(k; t)$  be the utility function of family  $i$  ( $k$  consumption,  $t$  relocation choice)
- Let  $W_i(z; t)$  be the budget set of family  $i$  under relocation decision  $t \in \{1, 2, 3\}$  and MTO voucher  $z \in \{z_1, z_2, z_3\}$
- Let  $S_i = C_i(z_1); C_i(z_2); C_i(z_3)$  denote the type of family  $i$ , defined by relocation response  $C_i(z)$  given different vouchers

# Generalization of LATE: Multivalued IV

- Let  $u_i(k; t)$  be the utility function of family  $i$  ( $k$  consumption,  $t$  relocation choice)
- Let  $W_i(z; t)$  be the budget set of family  $i$  under relocation decision  $t \in \{1, 2, 3\}$  and MTO voucher  $z \in \{z_1, z_2, z_3\}$
- Let  $S_i = C_i(z_1); C_i(z_2); C_i(z_3)$  denote the type of family  $i$ , defined by relocation response  $c_i(z)$  given different vouchers

# Generalization of LATE: Multivalued IV

- Let  $u_i(k; t)$  be the utility function of family  $i$  ( $k$  consumption,  $t$  relocation choice)
- Let  $W_i(z; t)$  be the budget set of family  $i$  under relocation decision  $t \in \{1, 2, 3\}$  and MTO voucher  $z \in \{z_1, z_2, z_3\}$
- Let  $S_i = \{C_i(z_1); C_i(z_2); C_i(z_3)\}$  denote the type of family  $i$ , defined by relocation response  $C_i(z)$  given different vouchers

# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidize nothing
  - Experimental group  $z_2 = 1$  subsidize according to low poverty neighborhood
  - Control B group  $z_3 = 1$  subsidize any neighborhood

According to the feature of MIV, we assume the budget sets satisfy:

$$W_1(z_1=2) \supset W_1(z_2=2) \supset W_1(z_1=2) \quad (1)$$

$$W_1(z_1=3) \supset W_1(z_2=3) \supset W_1(z_1=3) \quad (2)$$

$$W_1(z_1=1) \supset W_1(z_2=2) \supset W_1(z_1=3) \supset W_1(z_2=1) \supset W_1(z_1=3) \supset W_1(z_2=1) \quad (3)$$

- What are the meanings of these three relations?

# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidies nothing
  - Experimental group  $z_2 = 1$  subsidies relocating to low poverty neighborhood
  - Section 8 group  $z_3 = 1$  subsidies any relocation

Assumption A-1, A-2 Pinto (2015)

According to the features of MTO, we assume the budget sets satisfy:

$$W_i(z_1; 2) \succ W_i(z_2; 2) \succ W_i(z_3; 2) \quad (1)$$

$$W_i(z_1; 3) \succ W_i(z_2; 3) \succ W_i(z_3; 3) \quad (2)$$

$$W_i(z_1; 1) \succ W_i(z_1; 2) \succ W_i(z_1; 3) \succ W_i(z_2; 1) \succ W_i(z_2; 3) \succ W_i(z_3; 1) \quad (3)$$

- What are the meanings of these three relations?



# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidies nothing
  - Experimental group  $z_2 = 1$  subsidies relocating to low poverty neighborhood
  - Section 8 group  $z_3 = 1$  subsidies any relocation

Assumption A-1, A-2 Pinto (2015)

According to the features of MTO, we assume the budget sets satisfy:

$$W_i(z_1; 2) \succ W_i(z_2; 2) \succ W_i(z_3; 2) \quad (1)$$

$$W_i(z_1; 3) \succ W_i(z_2; 3) \succ W_i(z_3; 3) \quad (2)$$

$$W_i(z_1; 1) \succ W_i(z_1; 2) \succ W_i(z_1; 3) \succ W_i(z_2; 1) \succ W_i(z_2; 3) \succ W_i(z_3; 1) \quad (3)$$

- What are the meanings of these three relations?

# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidies nothing
  - Experimental group  $z_2 = 1$  subsidies relocating to low poverty neighborhood
  - Section 8 group  $z_3 = 1$  subsidies any relocation

Assumption A-1, A-2 Pinto (2015)

According to the features of MTO, we assume the budget sets satisfy:

$$W_i(z_1; 2) \rightarrow W_i(z_2; 2) \rightarrow W_i(z_3; 2) \quad (1)$$

$$W_i(z_1; 3) \rightarrow W_i(z_2; 3) \rightarrow W_i(z_3; 3) \quad (2)$$

$$W_i(z_1; 1) \rightarrow W_i(z_1; 2) \rightarrow W_i(z_1; 3) \rightarrow W_i(z_2; 1) \rightarrow W_i(z_2; 3) \rightarrow W_i(z_3; 1) \quad (3)$$

- What are the meanings of these three relations?

# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidies nothing
  - Experimental group  $z_2 = 1$  subsidies relocating to low poverty neighborhood
  - Section 8 group  $z_3 = 1$  subsidies any relocation

Assumption A-1, A-2 Pinto (2015)

According to the features of MTO, we assume the budget sets satisfy:

$$W_i(z_1; 2) \supseteq W_i(z_2; 2) \supseteq W_i(z_3; 2) \quad (1)$$

$$W_i(z_1; 3) \supseteq W_i(z_2; 3) \supseteq W_i(z_3; 3) \quad (2)$$

$$W_i(z_1; 1) \supseteq W_i(z_1; 2) \supseteq W_i(z_1; 3) \supseteq W_i(z_2; 1) \supseteq W_i(z_2; 3) \supseteq W_i(z_3; 1) \quad (3)$$

- What are the meanings of these three relations?

# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidies nothing
  - Experimental group  $z_2 = 1$  subsidies relocating to low poverty neighborhood
  - Section 8 group  $z_3 = 1$  subsidies any relocation

## Assumption A-1, A-2 Pinto (2015)

According to the features of MTO, we assume the budget sets satisfy:

$$W_i(z_1; 2) \rightarrow W_i(z_2; 2) \rightarrow W_i(z_3; 2) \quad (1)$$

$$W_i(z_1; 3) \rightarrow W_i(z_2; 3) \rightarrow W_i(z_3; 3) \quad (2)$$

$$W_i(z_1; 1) \rightarrow W_i(z_1; 2) \rightarrow W_i(z_1; 3) \rightarrow W_i(z_2; 1) \rightarrow W_i(z_2; 3) \rightarrow W_i(z_3; 1) \quad (3)$$

- What are the meanings of these three relations?

# Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group  $z_1 = 1$  subsidies nothing
  - Experimental group  $z_2 = 1$  subsidies relocating to low poverty neighborhood
  - Section 8 group  $z_3 = 1$  subsidies any relocation

## Assumption A-1, A-2 Pinto (2015)

According to the features of MTO, we assume the budget sets satisfy:

$$W_i(z_1; 2) \rightarrow W_i(z_2; 2) \rightarrow W_i(z_3; 2) \quad (1)$$

$$W_i(z_1; 3) \rightarrow W_i(z_2; 3) \rightarrow W_i(z_3; 3) \quad (2)$$

$$W_i(z_1; 1) \rightarrow W_i(z_1; 2) \rightarrow W_i(z_1; 3) \rightarrow W_i(z_2; 1) \rightarrow W_i(z_2; 3) \rightarrow W_i(z_3; 1) \quad (3)$$

- What are the meanings of these three relations?

# Generalization of LATE: Multivalued IV

- (1): If you choose to relocate to low poverty neighborhood (2), your consumption would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your consumption would be higher if you are in Section 8 group
- (3): If you choose not to relocate, or relocate to places that is not supported by your MTO group, your budget will not change

# Generalization of LATE: Multivalued IV

- (1): If you choose to relocate to low poverty neighborhood (2), your consumption would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your consumption would be higher if you are in Section 8 group
- (3): If you choose not to relocate, or relocate to places that is not supported by your MTO group, your budget will not change

# Generalization of LATE: Multivalued IV

- (1): If you choose to relocate to low poverty neighborhood (2), your consumption would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your consumption would be higher if you are in Section 8 group
- (3): If you choose not to relocate, or relocate to places that is not supported by your MTO group, your budget will not change



# Generalization of LATE: Multivalued IV

- (1): If you choose to relocate to low poverty neighborhood (2), your consumption would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your consumption would be higher if you are in Section 8 group
- (3): If you choose not to relocate, or relocate to places that is not supported by your MTO group, your budget will not change

# Generalization of LATE: Multivalued IV

- Then we derive the following choice rule

■ If preferences are rational, under Assumption A4 and A5,

$$\begin{aligned} 1) C_1(z_1) &\geq 2) C_1(z_2) & 2) C_1(z_2) &\geq 1) C_1(z_1) \\ 2) C_1(z_1) &\geq 3) C_1(z_2) & 1) C_1(z_1) &\geq 1) C_1(z_2) \\ 3) C_1(z_2) &\geq 1) C_1(z_1) & 1) C_1(z_2) &\geq 2) C_1(z_1) \\ 4) C_1(z_2) &\geq 3) C_1(z_1) & 3) C_1(z_2) &\geq 3) C_1(z_1) \\ 5) C_1(z_2) &\geq 1) C_1(z_1) & 1) C_1(z_2) &\geq 1) C_1(z_1) \\ 6) C_1(z_2) &\geq 2) C_1(z_1) & & \end{aligned}$$

- Test yourself, explain all these six inequalities

# Generalization of LATE: Multivalued IV

- Then we derive the following choice rule

Lemma L-1 Pinto (2015)

If preferences are rational, under Assumption A-1 and A-2:

$$\begin{array}{l} 1: C_1 \succeq z_1 \quad 2 \quad C_1 \succeq z_2 \quad 2; C_1 \succeq z_3 \quad j \quad 1 \\ 2: C_1 \succeq z_1 \quad 3 \quad C_1 \succeq z_2 \quad j \quad 1; C_1 \succeq z_3 \quad j \quad 1 \\ 3: C_1 \succeq z_2 \quad 1 \quad C_1 \succeq z_1 \quad 1; C_1 \succeq z_3 \quad j \quad 2 \\ 4: C_1 \succeq z_2 \quad 3 \quad C_1 \succeq z_1 \quad 3; C_1 \succeq z_3 \quad 3 \\ 5: C_1 \succeq z_3 \quad 1 \quad C_1 \succeq z_1 \quad 1; C_1 \succeq z_2 \quad 1 \\ 6: C_1 \succeq z_3 \quad 2 \quad C_1 \succeq z_2 \quad 2 \end{array}$$

- Test yourself, explain all these six inequalities

# Generalization of LATE: Multivalued IV

- Then we derive the following choice rule

## Lemma L-1 Pinto (2015)

If preferences are rational, under Assumption A-1 and A-2:

$$\begin{array}{l} 1: C_1 \succeq z_1 \quad 2 \quad C_1 \succeq z_2 \quad 2; C_1 \succeq z_3 \quad j \quad 1 \\ 2: C_1 \succeq z_1 \quad 3 \quad C_1 \succeq z_2 \quad j \quad 1; C_1 \succeq z_3 \quad j \quad 1 \\ 3: C_1 \succeq z_2 \quad 1 \quad C_1 \succeq z_1 \quad 1; C_1 \succeq z_3 \quad j \quad 2 \\ 4: C_1 \succeq z_2 \quad 3 \quad C_1 \succeq z_1 \quad 3; C_1 \succeq z_3 \quad 3 \\ 5: C_1 \succeq z_3 \quad 1 \quad C_1 \succeq z_1 \quad 1; C_1 \succeq z_2 \quad 1 \\ 6: C_1 \succeq z_3 \quad 2 \quad C_1 \succeq z_2 \quad 2 \end{array}$$

- Test yourself, explain all these six inequalities

# Generalization of LATE: Multivalued IV

- Then we derive the following choice rule

## Lemma L-1 Pinto (2015)

If preferences are rational, under Assumption A-1 and A-2:

$$\begin{array}{llll} 1: C_1 z_1 & 2 & C_1 z_2 & 2; C_1 z_3 \quad j \quad 1 \\ 2: C_1 z_1 & 3 & C_1 z_2 \quad j & 1; C_1 z_3 \quad j \quad 1 \\ 3: C_1 z_2 & 1 & C_1 z_1 & 1; C_1 z_3 \quad j \quad 2 \\ 4: C_1 z_2 & 3 & C_1 z_1 & 3; C_1 z_3 \quad 3 \\ 5: C_1 z_3 & 1 & C_1 z_1 & 1; C_1 z_2 \quad 1 \\ 6: C_1 z_3 & 2 & C_1 z_2 & 2 \end{array}$$

- Test yourself, explain all these six inequalities

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

For each family  $i$ , and for  $z_1, z_2, z_3 \in \mathbb{R}^+$ , if  $C_1 = z_1$  and  $W_1 = z_1$  is a proper subset of  $W_2 = z_2$ , then  $C_2 > C_1$ .

- To eliminate cases like  $C_1 = z_1 = 2; C_1 = z_2 = 2; C_1 = z_3 = 3$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t); C_i(z_2; t); C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

## Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t) \supset C_i(z_2; t); C_i(z_2; t) \supset C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!



# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

## Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t) \supset C_i(z_2; t) \supset C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

## Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t) \supset C_i(z_2; t); C_i(z_2; t) \supset C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

## Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t) \supset C_i(z_2; t); C_i(z_2; t) \supset C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

## Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t) \supset C_i(z_2; t); C_i(z_2; t) \supset C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- **When statistics tools are exhausted, remember you are an economist**
- **Do not think first year Micro and Macro are useless!!!**

# Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

## Assumption A-3 Pinto (2015)

For each family  $i$ , and for  $z; z^{-1} \in \{z_1; z_2; z_3\}$ , if  $C_i(z; t)$  and  $W_i(z; t)$  is a proper subset of  $W_i(z^{-1}; t)$ , then  $C_i(z^{-1}; t)$

- To eliminate cases like  $C_i(z_1; t) \supset C_i(z_2; t); C_i(z_2; t) \supset C_i(z_3; t)$
- Using all above, we can eliminate the number of types from 27 to 7
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

# MTE: Choice Model

- Now we go to the second part, how to improve the external validity
- The reason why LATE is lack of external validity is because it is defined on a policy-specific ex post group
- Not some ex ante group, for example a group of high-skilled workers
- Grouping by post-determined behavior, but not pre-determined characteristics
- This ex post group will change when policy environment changes

- Now we go to the second part, how to improve the external validity
- The reason why LATE is lack of external validity is because it is defined on a policy-specific ex post group
- Not some ex ante group, for example a group of high-skilled workers
- Grouping by post-determined behavior, but not pre-determined characteristics
- This ex post group will change when policy environment changes

# MTE: Choice Model

- Now we go to the second part, how to improve the external validity
- The reason why LATE is lack of external validity is because it is defined on a policy-specific ex post group
- Not some ex ante group, for example a group of high-skilled workers
- Grouping by post-determined behavior, but not pre-determined characteristics
- This ex post group will change when policy environment changes



# MTE: Choice Model

- Now we go to the second part, how to improve the external validity
- The reason why LATE is lack of external validity is because it is defined on a policy-specific ex post group
- Not some ex ante group, for example a group of high-skilled workers
- Grouping by post-determined behavior, but not pre-determined characteristics
- This ex post group will change when policy environment changes

# MTE: Choice Model

- Now we go to the second part, how to improve the external validity
- The reason why LATE is lack of external validity is because it is defined on a policy-specific ex post group
- Not some ex ante group, for example a group of high-skilled workers
- Grouping by post-determined behavior, but not pre-determined characteristics
- This ex post group will change when policy environment changes

# MTE: Choice Model

- Now we go to the second part, how to improve the external validity
- The reason why LATE is lack of external validity is because it is defined on a policy-specific ex post group
- Not some ex ante group, for example a group of high-skilled workers
- Grouping by post-determined behavior, but not pre-determined characteristics
- This ex post group will change when policy environment changes

# MTE: Choice Model

- Now let's explicitly construct a model for agents' compliance behavior
- In this model, we suppress subscript for individuals
- Let  $j = 0, 1$  be the treatment,  $Y_1; Y_0$  be the potential outcomes

$$Y_1 = \beta_1 X + U_1 \quad (4)$$

$$Y_0 = \beta_0 X + U_0 \quad (5)$$

- $X$  is a set of control variables,  $U_j$  is unobserved factor on outcome

# MTE: Choice Model

- Now let's explicitly construct a model for agents' compliance behavior
- In this model, we suppress subscript for individuals
- Let  $j = 0; 1$  be the treatment,  $Y_1; Y_0$  be the potential outcomes

$$Y_1 = \beta_1 X + U_1 \quad (4)$$

$$Y_0 = \beta_0 X + U_0 \quad (5)$$

- $X$  is a set of control variables,  $U_j$  is unobserved factor on outcome

# MTE: Choice Model

- Now let's explicitly construct a model for agents' compliance behavior
- In this model, we suppress subscript for individuals
- Let  $j = 0; 1$  be the treatment,  $Y_1; Y_0$  be the potential outcomes

$$Y_1 = \beta_1 X + U_1 \quad (4)$$

$$Y_0 = \beta_0 X + U_0 \quad (5)$$

- $X$  is a set of control variables,  $U_j$  is unobserved factor on outcome

# MTE: Choice Model

- Now let's explicitly construct a model for agents' compliance behavior
- In this model, we suppress subscript for individuals
- Let  $j = 0, 1$  be the treatment,  $Y_1; Y_0$  be the potential outcomes

$$Y_1 = f_1(X; U_1) \quad (4)$$

$$Y_0 = f_0(X; U_0) \quad (5)$$

- $X$  is a set of control variables,  $U_j$  is unobserved factor on outcome

# MTE: Choice Model

- Now let's explicitly construct a model for agents' compliance behavior
- In this model, we suppress subscript for individuals
- Let  $j \in \{0, 1\}$  be the treatment,  $Y_1; Y_0$  be the potential outcomes

$$Y_1 = \beta_1 X + U_1 \quad (4)$$

$$Y_0 = \beta_0 X + U_0 \quad (5)$$

- $X$  is a set of control variables,  $U_j$  is unobserved factor on outcome



# MTE: Choice Model

- Let  $D$  denote the choice of treatment, determined by a latent index model

$$D^* = \beta Z + V; \quad D = 1 \text{ if } D^* > 0; \quad D = 0 \text{ otherwise} \quad (6)$$

- $Z$  is an instrument that can change individual's choices, is an unobserved factor
- For instance,  $Y$  is wage,  $D$  is college enrollment,  $Z$  is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe  $X$ , but not  $U_0; U_1; V$
- $U_0; U_1; V$  can be correlated with each other

# MTE: Choice Model

- Let  $D$  denote the choice of treatment, determined by a latent index model

$$\tilde{D} = D + Z + V; \quad D = 1 \text{ if } \tilde{D} > 0; \quad D = 0 \text{ otherwise} \quad (6)$$

- $Z$  is an instrument that can change individual's choices,  $V$  is an unobserved factor
- For instance,  $Y$  is wage,  $D$  is college enrollment,  $Z$  is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe  $X$ , but not  $U_0; U_1; V$
- $U_0; U_1; V$  can be correlated with each other

# MTE: Choice Model

- Let  $D$  denote the choice of treatment, determined by a latent index model

$$\tilde{D} = D + Z\gamma + V; \quad D = 1 \text{ if } \tilde{D} > 0; \quad D = 0 \text{ otherwise} \quad (6)$$

- $Z$  is an instrument that can change individual's choices,  $V$  is an unobserved factor
- For instance,  $Y$  is wage,  $D$  is college enrollment,  $Z$  is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe  $X$ , but not  $U_0; U_1; V$
- $U_0; U_1; V$  can be correlated with each other

# MTE: Choice Model

- Let  $D$  denote the choice of treatment, determined by a latent index model

$$\tilde{D} = D + Z + V; \quad D = 1 \text{ if } \tilde{D} > 0; \quad D = 0 \text{ otherwise} \quad (6)$$

- $Z$  is an instrument that can change individual's choices,  $V$  is an unobserved factor
- For instance,  $Y$  is wage,  $D$  is college enrollment,  $Z$  is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe  $X$ , but not  $U_0; U_1; V$
- $U_0; U_1; V$  can be correlated with each other

# MTE: Choice Model

- Let  $D$  denote the choice of treatment, determined by a latent index model

$$D \sim D = Z\gamma + V; \quad D = 1 \text{ if } D \sim > 0; D = 0 \text{ otherwise} \quad (6)$$

- $Z$  is an instrument that can change individual's choices,  $V$  is an unobserved factor
- For instance,  $Y$  is wage,  $D$  is college enrollment,  $Z$  is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe  $Y, X$ , but not  $U_0; U_1; V$
- $U_0; U_1; V$  can be correlated with each other

# MTE: Choice Model

- Let  $D$  denote the choice of treatment, determined by a latent index model

$$D \sim D = \begin{cases} 1 & \text{if } D^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

- $Z$  is an instrument that can change individual's choices,  $V$  is an unobserved factor
- For instance,  $Y$  is wage,  $D$  is college enrollment,  $Z$  is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe  $X$ , but not  $U_0; U_1; V$
- $U_0; U_1; V$  can be correlated with each other

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D = 1$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E(Y_1)$ ;  $E(Y_0)$  are finite
- (A-5)  $0 < \Pr(D = 1) < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E \{Y_1\}; E \{Y_0\}$  are finite
- (A-5)  $0 < \Pr D = 1 \mid X < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$



# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E \{Y_1\}; E \{Y_0\}$  are finite
- (A-5)  $0 < \Pr D = 1 \mid X < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 **$Z$  contain at least one element not in  $X$**
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E \{Y_1\}; E \{Y_0\}$  are finite
- (A-5)  $0 < \Pr D = 1 \mid X < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E \{Y_1\}; E \{Y_0\}$  are finite
- (A-5)  $0 < \Pr D = 1 \mid X < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E \{Y_1\}; E \{Y_0\}$  are finite
- (A-5)  $0 < \Pr D = 1 \mid X < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E[Y_1]; E[Y_0]$  are finite
- (A-5)  $0 < \Pr(D=1|X) < 1$   
Possible to have  $D=1$  or  $D=0$  at any point of  $X$

# MTE: Choice Model

- We invoke five assumptions for this model
- (A-1)  $U_0; U_1; V$  are independent of  $Z$  conditional on  $X$   
Independence
- (A-2)  $D \mid Z$  is nondegenerate conditional on  $X$   
 $Z$  contain at least one element not in  $X$
- (A-1) and (A-2) assure the existence of the instrument
- (A-3) The distribution of  $V$  is continuous
- (A-4)  $E \{Y_1\}; E \{Y_0\}$  are finite
- (A-5)  $0 < \Pr D = 1 \mid X < 1$   
Possible to have  $D = 1$  or  $D = 0$  at any point of  $X$

# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y_0$  is working payo , there is relative working cost  $c$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment

# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y$  is working payo, there is relative working cost  $\frac{c_1}{c_0}$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment



# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y$  is working payo, there is relative working cost  $\frac{c_1}{c_0}$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment

# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y$  is working payo, there is relative working cost  $Z_1$   $V_C$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment

# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y$  is working payo, there is relative working cost  $Z_1$   $V_C$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment

# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y$  is working payo, there is relative working cost  $Z_1$   $V_C$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment

# MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors 0 and 1
- $Y$  is working payo, there is relative working cost  $Z_1$   $V_C$  in sector 1,  $Z_1$  is observed and  $V_C$  is unobserved
- Agents choose a sector with higher payo (abstract from cost)
- The unobserved term in treatment function is positively correlated with unobserved treatment return Positive sorting
- People with higher return sort into treatment

# MTE: Choice Model

- Assume additively separable i.i.d.

$$Y_1 = \beta_1 X + U_1$$

$$Y_0 = \beta_0 X + U_0$$

$$D^* = \beta_1 X + U_1 - \beta_0 X - U_0 = Z_1 + V_C; \quad D = 1 \text{ if } D^* > 0; D = 0 \text{ otherwise}$$

- In this case, we have  $E[U_1 | U_0 = V_C]$
- Positive sorting:  $\text{Cov}(U_1, U_0) = E[V_C] > 0$

# MTE: Choice Model

- Assume additively separable utility

$$Y_1 = \beta_1 X + U_1$$

$$Y_0 = \beta_0 X + U_0$$

$$D = \begin{cases} 1 & \text{if } Z_1 + V_C + U_1 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- In this case, we have  $\text{Cov}(U_1, U_0) = -V_C$
- Positive sorting:  $\text{Cov}(U_1, U_0) = -V_C < 0$

# MTE: Choice Model

- Assume additively separable utility

$$Y_1 = \beta_1 X + U_1$$

$$Y_0 = \beta_0 X + U_0$$

$$D = \begin{cases} 1 & \text{if } Z_1 + V_C + U_1 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- In this case, we have  $\text{Cov}(U_1, U_0) = -\text{Cov}(U_1, V_C)$
- Positive sorting:  $\text{Cov}(U_1, U_0) > 0$



# MTE: Choice Model

- Assume additively separable utility

$$Y_1 = \beta_1 X + U_1$$

$$Y_0 = \beta_0 X + U_0$$

$$D = \begin{cases} 1 & \text{if } Z_1 \geq V_C \\ 0 & \text{otherwise} \end{cases}$$

- In this case, we have  $\text{Cov}(U_1, U_0) = 0$
- Positive sorting:  $\text{Cov}(U_1, U_0) = 0$

# MTE: Choice Model

- Let  $P(Z; X) = \Pr(D = 1 | Z; X) = F_{V|X}(D; Z)$   
 $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $X$
- Let  $U_D = F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0; 1)$
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D; Z) = F_{V|X}(V) = U_D$  (larger than this threshold) to get treated
- We have a clear one-to-one mapping between  $U_D$  and  $V$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$

# MTE: Choice Model

- Let  $P(Z|X) = \Pr(D=1|Z;X) = F_{V|X}(D=Z)$   
 $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $X$
- Let  $U_D = F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0; 1)$
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- **Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D=Z) = F_{V|X}(V) = U_D$  (larger than this threshold) to get treated**
- We have a clear one-to-one mapping between  $V$  and  $U_D$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$

# MTE: Choice Model

- Let  $P = Z \mid X$     $\Pr(D = 1 \mid Z; X) = F_{V \mid X}(D; Z)$   
 $F_{V \mid X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $Z$
- Let  $U_D = F_{V \mid X}(V)$ , we have  $U_D \sim \text{Unif}(0; 1)$
- $F_{V \mid X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- **Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V \mid X}(D; Z) = F_{V \mid X}(V) = U_D$  (larger than this threshold) to get treated**
- We have a clear one-to-one mapping between  $V$  and  $U_D$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$

# MTE: Choice Model

- Let  $P = \Pr(D=1|X) = F_{V|X}$   
 $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $X$
- Let  $U_D = F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0, 1)$
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D) = F_{V|X}(V) = U_D$  (larger than this threshold) to get treated
- We have a clear one-to-one mapping between  $V$  and  $U_D$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$

# MTE: Choice Model

- Let  $P = \Pr(D=1|X) = F_{V|X}$   
 $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $X$
- Let  $U_D = F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0, 1)$
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D) = F_{V|X}(V) = U_D$  (larger than this threshold) to get treated
- We have a clear one-to-one mapping between  $V$  and  $U_D$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$

# MTE: Choice Model

- Let  $P \in \mathbb{R}^X$   $\Pr(D = 1 | Z; X) = F_{V|X}(D; Z)$   
 $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $Z$
- Let  $U_D \sim F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0; 1)$
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- **Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D; Z) \geq F_{V|X}(V) = U_D$  (larger than this threshold) to get treated**
- We have a clear one-to-one mapping between  $V$  and  $U_D$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$

# MTE: Choice Model

- Let  $P = \Pr(D=1|X) = F_{V|X}$   
 $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$
- This is the propensity score to get treated for agent with  $X$
- Let  $U_D = F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0, 1)$
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$
- **Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D) = F_{V|X}(V) > U_D$  (larger than this threshold) to get treated**
- We have a clear one-to-one mapping between  $V$  and  $U_D$
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$



# MTE: Choice Model

- Let  $P = \Pr(D=1|X) = F_{V|X}$ .  $F_{V|X}$  denotes the distribution of  $V$  conditional on  $X$ .
- This is the propensity score to get treated for agent with  $X$ .
- Let  $U_D = F_{V|X}(V)$ , we have  $U_D \sim \text{Unif}(0, 1)$ .
- $F_{V|X}(V)$  means the **threshold propensity score** the agent has to pass to get treated when he/she draws  $V$ .
- **Agent has to have an instrument  $Z$  which give him/her a propensity score  $F_{V|X}(D) = F_{V|X}(V) = U_D$  (larger than this threshold) to get treated**
- We have a clear one-to-one mapping between  $V$  and  $U_D$ .
- Thus, for a choice function, an agent can be characterized by  $V$  or  $X; U_D$ .

# MTE: Choice Model

- Vytlacil (2002) proves that (A-1) to (A-5) in this additively separable selection model is equivalent to the LATE model of Imbens and Angrist (1994)
- The intuition is simple:  $V$  could not affect  $D$  or  $Z$
- $D$  and  $Z$  are additively separable for  $Z$  and  $V$
- Thus, given  $z$  and  $z'$ ,  $V$  is either  $D(z) = D(z')$  or  $D(z) < D(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

# MTE: Choice Model

- Vytlacil (2002) proves that (A-1) to (A-5) in this additively separable selection model is equivalent to the LATE model of Imbens and Angrist (1994)
- The intuition is simple:  $V$  could not affect  $D$  or  $Z$
- $D$  or  $Z$  and  $V$  additively separable for  $Z$  and  $V$
- Thus, given  $z$  and  $z'$ ,  $\frac{3}{4}V = D(z) - D(z')$  or  $D(z) & D(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

# MTE: Choice Model

- Vytlacil (2002) proves that (A-1) to (A-5) in this additively separable selection model is equivalent to the LATE model of Imbens and Angrist (1994)
- The intuition is simple:  $V$  could not affect  $D$  or  $Z$
- $D$  and  $Z$  are additively separable for  $Z$  and  $V$
- Thus, given  $z$  and  $z'$ ,  $\frac{3}{4}V = D(z) - D(z')$  or  $D(z) & D(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

# MTE: Choice Model

- Vytlacil (2002) proves that (A-1) to (A-5) in this additively separable selection model is equivalent to the LATE model of Imbens and Angrist (1994)
- The intuition is simple:  $V$  could not affect  $D$  or  $Z$
- $D$  and  $Z$  are additively separable for  $Z$  and  $V$
- Thus, given  $z$  and  $z'$ ,  $\frac{3}{4}V = D(z) - D(z')$  or  $D(z) & D(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

# MTE: Choice Model

- Vytlacil (2002) proves that (A-1) to (A-5) in this additively separable selection model is equivalent to the LATE model of Imbens and Angrist (1994)
- The intuition is simple:  $V$  could not affect  $D$  or  $Z$
- $D$  and  $Z$  are additively separable for  $Z$  and  $V$
- Thus, given  $z$  and  $z'$ ,  $\frac{3}{4}V = D(z) - D(z')$  or  $D(z) & D(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

# MTE: Choice Model

- Vytlacil (2002) proves that (A-1) to (A-5) in this additively separable selection model is equivalent to the LATE model of Imbens and Angrist (1994)
- The intuition is simple:  $V$  could not affect  $D$  or  $Z$
- $D$  and  $Z$  are additively separable for  $Z$  and  $V$
- Thus, given  $z$  and  $z'$ ,  $\frac{3}{4}V = D(z) - D(z')$  or  $D(z) & D(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

# MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let  $Y_1 = Y_0 + \tau X + U_1 - U_0$
- ATE is de ned as usual:  $ATE(x) = E[\tau X | x]$
- MTE is de ned as the mean effect of treatment on those for who  $\tau_0(x) = U_0 = U_1 - \tau(x)$

The Marginal Treatment Effect is de ned as:

$$MTE(x) = E[\tau(x) | U_0 = U_1 - \tau(x)]$$



# MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let  $Y_1 = Y_0 + \tau X + U_D + V$
- ATE is de ned as usual:  $ATE(x) = E(\tau X | x)$
- MTE is de ned as the mean effect of treatment on those for whom  $X = x$  and  $U_D = u_D$ :  $MTE(x; u_D) = E(\tau X | x; U_D = u_D)$

## Definition of the MTE

The Marginal Treatment Effect is de ned as:

$$MTE(x; u_D) = E(\tau X | x; U_D = u_D)$$

# MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let  $Y_1 = Y_0 + \tau X + U_D + V$
- ATE is defined as usual:  $ATE(x) = E[\tau X | x]$
- MTE is defined as the mean effect of treatment on those for whom  $X = x$  and  $U_D = u_D$ :  $MTE(x; u_D) = E[\tau X | x; U_D = u_D]$

## Definition of the MTE

The Marginal Treatment Effect is defined as:

$$MTE(x; u_D) = E[\tau X | x; U_D = u_D]$$

# MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let  $Y_1 = Y_0 + \tau X$
- ATE is de ned as usual:  $ATE(x) = E[\tau X | x]$
- MTE is de ned as the mean effect of treatment on those for whom  $U_D = u_D$  and  $V = v$

## Definition of the MTE

The Marginal Treatment Effect is de ned as:

$$MTE(x; u_D) = E[\tau X | x; U_D = u_D]$$

# MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let  $Y_1 = Y_0 + \tau X$
- ATE is de ned as usual:  $ATE(x) = E[\tau X | x]$
- MTE is de ned as the mean effect of treatment on those for whom  $X = x$  and  $U_D = u_D$ :  $MTE(x; u_D) = E[\tau X | x; U_D = u_D]$

## De nition of the MTE

The Marginal Treatment Effect is de ned as:

$$MTE(x; u_D) = E[\tau X | x; U_D = u_D]$$

# MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let  $Y_1 = Y_0 + \tau X + U_D + V$
- ATE is defined as usual:  $ATE(x) = E[\tau X | x]$
- MTE is defined as the mean effect of treatment on those for whom  $X = x$  and  $U_D = u_D$ :  $MTE(x; u_D) = E[\tau X | x; U_D = u_D]$

## Definition of the MTE

The Marginal Treatment Effect is defined as:

$$MTE(x; u_D) = E[\tau X | x; U_D = u_D]$$

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics and unobserved taste on treatment
- People with observed characteristics who would be indifferent between treatment or not if they were randomly assigned a value  $z$  of  $z$  such that  $P_z = u_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $u_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics and unobserved taste on treatment
- People with observed characteristics who would be indifferent between treatment or not if they were randomly assigned a value  $z$  of  $z$  such that  $P_z = u_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $u_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics and unobserved taste on treatment
- People with observed characteristics who would be indifferent between treatment or not if they were randomly assigned a value  $z$  of  $z$  such that  $P_z = u_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $u_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid



# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics  $X$  and unobserved taste on treatment  $U_D$
- People with observed characteristics  $X$  who would be indifferent between treatment or not if they were randomly assigned a value  $Z$  of  $z$  such that  $P_z = U_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $U_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics and unobserved taste on treatment
- People with observed characteristics who would be indifferent between treatment or not if they were randomly assigned a value  $z$  of  $z$  such that  $P_z = u_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics  $X$  and unobserved taste on treatment  $U_D$
- People with observed characteristics  $X$  who would be indifferent between treatment or not if they were randomly assigned a value  $Z$  of  $z$  such that  $P_z = U_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $U_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics  $X$  and unobserved taste on treatment  $U_D$
- People with observed characteristics  $X$  who would be indifferent between treatment or not if they were randomly assigned a value  $Z$  of  $z$  such that  $P_z = U_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $U_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics  $X$  and unobserved taste on treatment  $U_D$
- People with observed characteristics  $X$  who would be indifferent between treatment or not if they were randomly assigned a value  $Z$  of  $z$  such that  $P_z = U_D$
- That is why it is called "marginal"  
Marginal people who have just the threshold  $U_D$
- Different from LATE, it is not defined by any instrument in an ex post way
- This is a deep structural parameter that does not change when IV is changed
- Thus, it is externally valid

# MTE: De ning MTE

- Selection on MTE in a positive sorting Roy model

# MTE: De ning MTE

- Selection on MTE in a positive sorting Roy model

# MTE: MTE as a Framework

- We can prove that MTE is a general framework with various causal parameters as its special cases
- LATE can be written as a weighted average of MTE:

$$\begin{aligned} \text{LATE} &= E[Y_1 - Y_0 | X = x; D = z = 1; D = z^* = 0] \\ &= E[Y_1 - Y_0 | X = x; u_D^* \leq U_D & u_D < u_D^*] \\ &= E^{u_D} \text{MTE}(x; u) du \end{aligned}$$

- Here  $u_D^* = \Pr(D = z = 1)$ ;  $u_D = \Pr(D = z^* = 1)$  are the threshold propensity scores for instrument  $Z = z$  and  $Z = z^*$
- We can interpret LATE as the average TE for people whose threshold is below but above  $z^*$



# MTE: MTE as a Framework

- We can prove that MTE is a general framework with various causal parameters as its special cases
- LATE can be written as a weighted average of MTE:

$$\text{LATE} = E[Y_1 - Y_0 | X = x; D = z = 1; D = z^* = 0] \\ = E[Y_1 - Y_0 | X = x; u_D^* \leq U_D & u_D < u_D^*] \\ = E^{u_D^*}_{u_D} \text{MTE}(x; u) du$$

- Here  $u_D^* = \Pr(D = z = 1 | X = x)$  and  $u_D = \Pr(D = z^* = 1 | X = x)$  are the threshold propensity scores for instrument  $Z = z$  and  $Z = z^*$
- We can interpret LATE as the average TE for people whose threshold is below  $u_D^*$  but above  $u_D$

# MTE: MTE as a Framework

- We can prove that MTE is a general framework with various causal parameters as its special cases
- LATE can be written as a weighted average of MTE:

$$\begin{aligned} \text{LATE} &= E[Y_1 - Y_0 | X = x; D = z = 1; D = z^* = 0] \\ &= E[Y_1 - Y_0 | X = x; u_D^* \leq U_D \leq u_D] \\ &= E^{u_D^*}_{u_D} \text{MTE}(x; u) du \end{aligned}$$

- Here  $u_D^* = \Pr(D = z = 1 | X = x)$  and  $u_D = \Pr(D = z^* = 1 | X = x)$  are the threshold propensity scores for instrument  $Z = z$  and  $Z = z^*$
- We can interpret LATE as the average TE for people whose threshold is below  $u_D^*$  but above  $u_D$

# MTE: MTE as a Framework

- We can prove that MTE is a general framework with various causal parameters as its special cases
- LATE can be written as a weighted average of MTE:

$$\begin{aligned} \text{LATE} &= E[Y_1 - Y_0 | X = x; D = z = 1; D = z = 0] \\ &= E[Y_1 - Y_0 | X = x; u_D^+ \leq U_D \leq u_D^-] \\ &= E^{u_D^+}_{u_D^-} \text{MTE}(x; u) du \end{aligned}$$

- Here  $u_D^+ = \Pr(D = z = 1 | u)$  and  $u_D^- = \Pr(D = z = 0 | u)$  are the threshold propensity scores for instrument  $Z = z$  and  $Z = z^c$
- We can interpret LATE as the average TE for people whose threshold is below  $u_D^+$  but above  $u_D^-$

# MTE: MTE as a Framework

- We can prove that MTE is a general framework with various causal parameters as its special cases
- LATE can be written as a weighted average of MTE:

$$\begin{aligned} \text{LATE} &= E[Y_1 - Y_0 | X = x; D = z = 1; D = z^* = 0] \\ &= E[Y_1 - Y_0 | X = x; u_D^* \leq U_D & U_D < u_D] \\ &= E^{u_D^*}_{u_D} \text{MTE}(x; u) du \end{aligned}$$

- Here  $u_D^* = \Pr(D = z = 1 | X = x)$  and  $u_D = \Pr(D = z^* = 1 | X = x)$  are the threshold propensity scores for instrument  $Z = z$  and  $Z = z^*$
- We can interpret LATE as the average TE for people whose threshold is below  $u_D^*$  but above  $u_D$

# MTE: MTE as a Framework

- In general, we can express treatment parameter by MTE as:

$$TE_j = E_0^1 \int_{MTE_j(x; u_D)} MTE(x; u_D) \lambda_j(x; u_D) du_D$$

- $\lambda_j$  is the weight for  $j$

# MTE: MTE as a Framework

- In general, we can express treatment parameters by MTE as:

$$TE_j = E \int_0^1 MTE(x; u_D) \lambda_j(x; u_D) du_D$$

- $\lambda_j$  is the weight for  $j$

# MTE: MTE as a Framework

- In general, we can express treatment parameters by MTE as:

$$TE_j = E \int_0^1 MTE(x; u_D) \lambda_j(x; u_D) du_D$$

- $\lambda_j$  is the weight for  $j$

# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P$   $Z$   $p$ :

$$\text{LIV}_p = \frac{\partial E(Y|P, Z, p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in  $Z$
- A policy shock changes  $Z$  changes propensity score  $Z$  changes outcome  $Y$



# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on  $\mathbf{X}$
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P(Z = p)$ :

$$LIV_p = \frac{\partial E(Y | P(Z = p))}{\partial p}$$

- LIV is the mean response to a policy change embodied in changing  $P(Z = p)$
- A policy shock changes  $Z$  changes propensity score  $P(Z = p)$  changes outcome  $Y$

# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on  $\mathbf{X}$
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P(Z = p)$ :

$$\text{LIV}_p = \frac{\partial E(Y | P(Z = p))}{\partial p}$$

- LIV is the mean response to a policy change embodied in changing  $P(Z = p)$
- A policy shock changes  $Z$  changes propensity score  $P(Z = p)$  changes outcome  $Y$

# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on  $\mathbf{X}$
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P$   $Z$   $p$ :

$$\text{LIV}_p = \frac{\partial E(Y|P, Z, p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in  $Z$
- A policy shock changes  $Z$  changes propensity score  $P$   $Z$  changes outcome  $Y$

# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on  $\mathbf{X}$
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P$   $Z$   $p$ :

$$LIV_p = \frac{\partial E(Y|P, Z, p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in  $Z$
- A policy shock changes  $Z$  changes propensity score  $P$   $Z$  changes outcome  $Y$

# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on  $\mathbf{X}$
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P$   $Z$   $p$ :

$$LIV_p = \frac{\partial E(Y|P, Z, p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in  $Z$
- A policy shock changes  $Z$  changes propensity score  $P$  changes outcome  $Y$

# MTE: Estimate MTE Using LIV

- Now we have defined MTE and shown that it is a general framework
- We suppress notation of conditional on  $\mathbf{x}$
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expectation of  $Y$  w.r.t  $P(Z)$ :

$$LIV_p = \frac{\partial E(Y|P(Z)=p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in  $Z$
- A policy shock changes  $Z$  changes propensity score  $P(Z)$  changes outcome  $Y$

# MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\text{MTE}_{\rho} = \text{LIV}_{\rho} \frac{\partial E(Y|Z)_{\rho}}{\partial Z_{\rho}}$$

- For MTE at any propensity threshold  $\rho$ , we can use LIV at this point to identify it
- What is the intuition?
- MTE at a threshold means the causal effect on marginal people who would just change their treatment at this point  $\rho$
- LIV is the changes of outcome at this marginal point  $\rho$  driven by an exogenous variation on instrument  $Z$

# MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$MTE_p = LIV_p = \frac{\partial E(Y|P=Z) \partial P}{\partial Z} \Big|_{P=Z=p}$$

- For MTE at any propensity threshold  $p$ , we can use LIV at this point to identify it
- What is the intuition?
- MTE at a threshold means the causal effect on marginal people who would just change their treatment at this point of  $P=Z=p$
- LIV is the changes of outcome at this marginal point of  $P=Z=p$  driven by an exogenous variation on instrument  $Z$



# MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\text{MTE}_p = \text{LIV}_p = \frac{\partial E(Y|P=Z=p)}{\partial p}$$

- For MTE at any propensity threshold  $p$ , we can use LIV at this point to identify it
- What is the intuition?
- MTE at a threshold means the causal effect on marginal people who would just change their treatment at this point of  $P=Z=p$
- LIV is the changes of outcome at this marginal point of  $P=Z=p$  driven by an exogenous variation on instrument  $Z$

# MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\text{MTE}_p = \text{LIV}_p = \frac{\partial E(Y|P=Z=p)}{\partial p}$$

- For MTE at any propensity threshold  $p$ , we can use LIV at this point to identify it
- What is the intuition?
  - MTE at a threshold means the causal effect on marginal people who would just change their treatment at this point of  $P=Z=p$
  - LIV is the changes of outcome at this marginal point of  $P=Z=p$  driven by an exogenous variation on instrument  $Z$

# MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\text{MTE}_p = \text{LIV}_p = \frac{\partial E(Y|P=Z)_{p^*}}{\partial p}$$

- For MTE at any propensity threshold  $p$ , we can use LIV at this point to identify it
- What is the intuition?
- MTE at a threshold means the causal effect on marginal people who would just change their treatment at this point of  $P=Z=p$
- LIV is the changes of outcome at this marginal point of  $P=Z=p$  driven by an exogenous variation on instrument  $Z$

# MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\text{MTE}_p = \text{LIV}_p = \frac{E(Y|P=z_p) - E(Y|P=z_{p-})}{\Delta P}$$

- For MTE at any propensity threshold  $p$ , we can use LIV at this point to identify it
- What is the intuition?
- MTE at a threshold means the causal effect on marginal people who would just change their treatment at this point of  $P = z_p$
- LIV is the changes of outcome at this marginal point of  $P = z_p$  driven by an exogenous variation on instrument  $Z$

# MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- First, assume a treatment choice function (Probit or logit), and propensity score function  $p(z)$
- Second, estimate outcome  $y$  given control  $X$  and propensity score function  $p(z)$ . Using non/semi-parametric methods such as local linear regression or partial linear regression
- Then estimate derivatives by small perturbation  
Or it would be just the regression coefficient if you assume a linear model
- Or we can estimate the whole model in a fully parametric way (Kline and Walters, 2016)

# MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- First, assume a treatment choice function (Probit or logit), and propensity score function  $p(z)$
- Second, estimate outcome  $y$  given control  $X$  and propensity score function  $p(z)$ . Using non/semi-parametric methods such as local linear regression or partial linear regression
- Then estimate derivatives by small perturbation. Or it would be just the regression coefficient if you assume a linear model for
- Or we can estimate the whole model in a fully parametric way (Kline and Walters, 2016)

# MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- First, assume a treatment choice function (Probit or logit), and propensity score function  $p(z)$
- Second, estimate outcome  $y$  given control  $X$  and propensity score function  $p(z)$   
Using non/semi-parametric methods such as local linear regression or partial linear regression
- Then estimate derivatives by small perturbation  
Or it would be just the regression coefficient if you assume a linear model for
- Or we can estimate the whole model in a fully parametric way (Kline and Walters, 2016)

# MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- First, assume a treatment choice function (Probit or logit), and propensity score function  $p(z)$
- Second, estimate outcome  $y$  given control  $X$  and propensity score function  $p(z)$   
Using non/semi-parametric methods such as local linear regression or partial linear regression
- Then estimate derivatives by small perturbation  
Or it would be just the regression coefficient if you assume a linear model for
- Or we can estimate the whole model in a fully parametric way (Kline and Walters, 2016)



## MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- First, assume a treatment choice function (Probit or logit), and propensity score function  $p(z)$
- Second, estimate outcome  $y$  given control  $X$  and propensity score function  $p(z)$   
Using non/semi-parametric methods such as local linear regression or partial linear regression
- Then estimate derivatives by small perturbation  
Or it would be just the regression coefficient if you assume a linear model for  $y$
- Or we can estimate the whole model in a fully parametric way (Kline and Walters, 2016)

## MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- First, assume a treatment choice function (Probit or logit), and propensity score function  $p(z)$
- Second, estimate outcome  $y$  given control  $X$  and propensity score function  $p(z)$   
Using non/semi-parametric methods such as local linear regression or partial linear regression
- Then estimate derivatives by small perturbation  
Or it would be just the regression coefficient if you assume a linear model for  $y$
- Or we can estimate the whole model in a fully parametric way (Kline and Walters, 2016)

# MTE: Estimate MTE Using LIV

- Implementation: Stata packagenteefe
- This package can give you estimations of various causal parameters
- And a full distribution of treatment effect
- There is a slides of user guidance by Martin Andresen

# MTE: Estimate MTE Using LIV

- **Implementation: Stata package `mtefe`**
- This package can give you estimations of various causal parameters
- And a full distribution of treatment effect
- There is a slides of user guidance by Martin Andresen

# MTE: Estimate MTE Using LIV

- Implementation: Stata package `mtefe`
- This package can give you estimations of various causal parameters
  - And a full distribution of treatment effect
  - There is a slides of user guidance by Martin Andresen

# MTE: Estimate MTE Using LIV

- Implementation: Stata package `mtefe`
- This package can give you estimations of various causal parameters
- And a full distribution of treatment effect
- There is a slides of user guidance by Martin Andresen

# MTE: Estimate MTE Using LIV

- Implementation: Stata package `mtefe`
- This package can give you estimations of various causal parameters
- And a full distribution of treatment effect
- There is a slides of user guidance by Martin Andresen

# MTE: Estimate MTE Using LIV



# MTE: Estimate MTE Using LIV

# MTE: Estimate MTE Using LIV

# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

# MTE: Conclusion

- **LATE is internally valid but not externally valid**
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package



# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

# MTE: Conclusion

- LATE is internally valid but not externally valid
- We can combine choice model with IV to have a new framework: MTE
- MTE measures the treatment effect for people with specific characteristics and some unobserved treatment taste (or treatment threshold)
- It is externally valid and not IV-specific
- Various causal parameters are special cases of weighted MTEs
- We can estimate it using LIV with non/semi/parametric methods
- It can be implemented using a Stata package

## Application: Kline and Walters (2016)

- We illustrate the method we learn today by reading Kline and Walters (2016)
- This paper is so interesting and insightful
- Reading one paper like this carefully, is much better than reading 100 reg monkey papers (for these, you can just read the abstracts)
- It investigates the cost-benefit analysis for social programs when some close substitutes exist

# Application: Kline and Walters (2016)

- We illustrate the method we learn today by reading Kline and Walters (2016)
- This paper is so interesting and insightful
- Reading one paper like this carefully, is much better than reading 100 reg monkey papers (for these, you can just read the abstracts)
- It investigates the cost-benefit analysis for social programs when some close substitutes exist

## Application: Kline and Walters (2016)

- We illustrate the method we learn today by reading Kline and Walters (2016)
- This paper is so interesting and insightful
- Reading one paper like this carefully, is much better than reading 100 reg monkey papers (for these, you can just read the abstracts)
- It investigates the cost-benefit analysis for social programs when some close substitutes exist

## Application: Kline and Walters (2016)

- We illustrate the method we learn today by reading Kline and Walters (2016)
- This paper is so interesting and insightful
- Reading one paper like this carefully, is much better than reading 100 reg monkey papers (for these, you can just read the abstracts)
- It investigates the cost-benefit analysis for social programs when some close substitutes exist

## Application: Kline and Walters (2016)

- We illustrate the method we learn today by reading Kline and Walters (2016)
- This paper is so interesting and insightful
- Reading one paper like this carefully, is much better than reading 100 reg monkey papers (for these, you can just read the abstracts)
- It investigates the cost-benefit analysis for social programs when some close substitutes exist

## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people enroll in HS and people do not enroll in any program
- Meanwhile, RCTs compare people enroll in HS and people do not enroll in HS. But many other programs exist
- People can actively sort into other programs



## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people enroll in HS and people do not enroll in any program
- Meanwhile, RCTs compare people enroll in HS and people do not enroll in HS. But many other programs exist
- People can actively sort into other programs

## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people enroll in HS and people do not enroll in any program
- Meanwhile, RCTs compare people enroll in HS and people do not enroll in HS. But many other programs exist
- People can actively sort into other programs

## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people who enroll in HS and people who do not enroll in any program
- Meanwhile, RCTs compare people who enroll in HS and people who do not enroll in HS. But many other programs exist
- People can actively sort into other programs

## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people enroll in HS and people do not enroll in any program
- Meanwhile, RCTs compare people enroll in HS and people do not enroll in HS  
But many other programs exist
- People can actively sort into other programs

## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people enroll in HS and people do not enroll in any program
- Meanwhile, RCTs compare people enroll in HS and people do not enroll in HS. But many other programs exist
- People can actively sort into other programs

## Application: Kline and Walters (2016)

- Head Start (HS) is an early childhood education program provided for poor families in the U.S.
- People find large impact from observational studies, but small effect from RCT. Does it mean that this HS is ineffective?
- Kline and Walters (2016) claim that it is not because observational studies are not well-designed
- Rather, it is because observational studies compare people enroll in HS and people do not enroll in any program
- Meanwhile, RCTs compare people enroll in HS and people do not enroll in HS. But many other programs exist
- People can actively sort into other programs

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - $ATE$  and  $LATE$  are naturally valid when the complier of always-takers are
  - $LATE$  naturally valid when the complier of never-takers are
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS



## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

## Application: Kline and Walters (2016)

- The treatment actually has three values: no program, other program, HS program
- Kline and Walters (2016) first categorize people to all behavior types and use ARP to eliminate some of them
- Then they verify various causal parameters needed for different evaluation targets
  - ITT and LATE: not externally valid when the composition of compliers changes
  - MTE: externally valid when the composition of compliers changes
- Using these causal estimates, they implement new cost-benefit analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not extremely valid when compliance group diverges
- To fix these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach

# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not externally valid when complier group changes
- To x these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach



# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not externally valid when complier group changes
- To x these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach

# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not externally valid when complier group changes
- To x these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach

# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not externally valid when complier group changes
- To x these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach

# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not externally valid when complier group changes
- To x these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach

# Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV too many types
  - Not externally valid when complier group changes
- To x these two issues, we need to go deep into the compliance (treatment selection) problem
- Treatment selection is intrinsically a part of IV, but not fully explored by pure design-based approach

# Conclusion

- First, we use ARP and other reasonable economic assumptions to simplify the identification in complicated multivalued IV cases
- Second, we introduce MTE framework to deal with external validity issues
- MTE is the treatment effect of a small group of people with specific value of characteristics  $X$  and treatment taste  $V$  (or treatment threshold  $U_D$ )
- It can be identified and estimated using LIV

# Conclusion

- First, we use ARP and other reasonable economic assumptions to simplify the identification in complicated multivalued IV cases
- Second, we introduce MTE framework to deal with external validity issues
- MTE is the treatment effect of a small group of people with specific value of characteristics  $X$  and treatment taste  $V$  (or treatment threshold  $U_D$ )
- It can be identified and estimated using LIV

# Conclusion

- First, we use ARP and other reasonable economic assumptions to simplify the identification in complicated multivalued IV cases
- Second, we introduce MTE framework to deal with external validity issues
- MTE is the treatment effect of a small group of people with specific value of characteristics  $X$  and treatment taste  $V$  (or treatment threshold  $U_D$ )
- It can be identified and estimated using LIV



# Conclusion

- First, we use ARP and other reasonable economic assumptions to simplify the identification in complicated multivalued IV cases
- Second, we introduce MTE framework to deal with external validity issues
- MTE is the treatment effect of a small group of people with specific value of characteristics  $X$  and treatment taste  $V$  (or treatment threshold  $U_D$ )
- It can be identified and estimated using LIV

# Conclusion

- First, we use ARP and other reasonable economic assumptions to simplify the identification in complicated multivalued IV cases
- Second, we introduce MTE framework to deal with external validity issues
- MTE is the treatment effect of a small group of people with specific value of characteristics  $X$  and treatment taste  $V$  (or treatment threshold  $U_D$ )
- It can be identified and estimated using LIV

# References

- Heckman, James J and Edward J Vytlacil. 2007a. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." *Handbook of Econometrics* 6:4779–4874.
- . 2007b. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." *Handbook of Econometrics* 6:4875–5143.
- Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475.
- Kline, Patrick and Christopher R Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *The Quarterly Journal of Economics* 131 (4):1795–1848.
- Pinto, Rodrigo. 2015. "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity." *Unpublished Ph. D. Thesis, University of Chicago, Department of Economics* .
- Vytlacil, Edward. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica* 70 (1):331–341.