Frontier Topics in Empirical Economics: Week 6
IV beyond LATE

Zibin Huang [1]

[1] College of Business, Shanghai University of Finance and Economics

October 21, 2024

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the complier group changes

IV beyond LATE: Limitation of LATE

- We have introduced the 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 single binary treatment and single binary IV
  - It is internally valid, but not externally valid
- Complier group is policy-specific, environment-specific
- When the environment changes, the 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 the 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 $(D \; z \quad 0 \quad 1)$
- Then he/she will definitely choose it when the price is lower $(D \; z \quad 1 \quad 1)$

IV beyond LATE: Choice Model and IV

- You have already used it in the 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 $(D \; z \quad 0 \quad 1)$
- Then he/she will definitely choose it when the price is lower $(D \; z \quad 1 \quad 1)$

IV beyond LATE: Choice Model and IV

- You have already used it in the 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 ($D \; z \quad 0 \quad\quad 1$)
- Then he/she will definitely choose it when the price is lower ($D \; z \quad 1 \quad\quad 1$)

IV beyond LATE: Choice Model and IV

- You have already used it in the 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 $(D \; z \quad 0 \quad 1)$
- Then he/she will definitely choose it when the price is lower $(D \; z \quad 1 \quad 1)$

IV beyond LATE: Choice Model and IV

- You have already used it in the 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 $(D \; z \quad 0 \quad 1)$
- Then he/she will definitely choose it when the price is lower $(D \; z \quad 1 \quad 1)$

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

- We have four equations (final nodes, three types)

$$E[Y|Z=1,D=1] = P(A|Z=1,D=1)\,E[Y_1|Z=1,D=1,A] + P(C|Z=1,D=1)\,E[Y_1|Z=1,D=1,C]$$
$$E[Y|Z=1,D=0] = E[Y_0|Z=1,D=0,N]$$
$$E[Y|Z=0,D=1] = E[Y_1|Z=0,D=1,A]$$
$$E[Y|Z=0,D=0] = P(N|Z=0,D=0)\,E[Y_0|Z=0,D=0,N] + P(C|Z=0,D=0)\,E[Y_0|Z=0,D=0,C]$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal effect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization of
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (nal nodes, three types)

$$E\ Y\ ¶Z\quad 1; D\quad 1\qquad P\ \ A¶Z\quad 1; D\quad 1\ E\ \ Y_1¶Z\quad 1; D\quad 1; A\qquad P\ \ C¶Z\quad 1; D\quad 1\ E\ \ Y_1¶Z\quad 1; D\quad 1; C$$

$$E\ Y\ ¶Z\quad 1; D\quad 0\qquad E\ \ Y_0¶Z\quad 1; D\quad 0; N$$

$$E\ Y\ ¶Z\quad 0; D\quad 1\qquad E\ \ Y_1¶Z\quad 0; D\quad 1; A$$

$$E\ Y\ ¶Z\quad 0; D\quad 0\qquad P\ \ N¶Z\quad 0; D\quad 0\ E\ \ Y_0¶Z\quad 0; D\quad 0; N\qquad P\ \ C¶Z\quad 0; D\quad 0\ E\ \ Y_0¶Z\quad 0; D\quad 0; C$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal e ect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (final nodes, three types)

$$E[Y|Z=1,D=1] = P(A|Z=1,D=1)\,E[Y_1|Z=1,D=1,A] + P(C|Z=1,D=1)\,E[Y_1|Z=1,D=1,C]$$
$$E[Y|Z=1,D=0] = E[Y_0|Z=1,D=0,N]$$
$$E[Y|Z=0,D=1] = E[Y_1|Z=0,D=1,A]$$
$$E[Y|Z=0,D=0] = P(N|Z=0,D=0)\,E[Y_0|Z=0,D=0,N] + P(C|Z=0,D=0)\,E[Y_0|Z=0,D=0,C]$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal effect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (nal nodes, three types)

$$E\ Y\ ¶Z\quad 1; D\quad 1\qquad P\ A¶Z\quad 1; D\quad 1\ E\ Y_1¶Z\quad 1; D\quad 1; A\qquad P\ C¶Z\quad 1; D\quad 1\ E\ Y_1¶Z\quad 1; D\quad 1; C$$

$$E\ Y\ ¶Z\quad 1; D\quad 0\qquad E\ Y_0¶Z\quad 1; D\quad 0; N$$

$$E\ Y\ ¶Z\quad 0; D\quad 1\qquad E\ Y_1¶Z\quad 0; D\quad 1; A$$

$$E\ Y\ ¶Z\quad 0; D\quad 0\qquad P\ N¶Z\quad 0; D\quad 0\ E\ Y_0¶Z\quad 0; D\quad 0; N\qquad P\ C¶Z\quad 0; D\quad 0\ E\ Y_0¶Z\quad 0; D\quad 0; C$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal e ect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization f
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (final nodes, three types)

$$E[Y|Z=1, D=1] = P[A|Z=1, D=1]\,E[Y_1|Z=1, D=1, A] + P[C|Z=1, D=1]\,E[Y_1|Z=1, D=1, C]$$
$$E[Y|Z=1, D=0] = E[Y_0|Z=1, D=0, N]$$
$$E[Y|Z=0, D=1] = E[Y_1|Z=0, D=1, A]$$
$$E[Y|Z=0, D=0] = P[N|Z=0, D=0]\,E[Y_0|Z=0, D=0, N] + P[C|Z=0, D=0]\,E[Y_0|Z=0, D=0, C]$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal effect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization of …
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (final nodes, three types)

$$E[Y|Z=1,D=1] = P(A|Z=1,D=1)\,E[Y_1|Z=1,D=1,A] + P(C|Z=1,D=1)\,E[Y_1|Z=1,D=1,C]$$
$$E[Y|Z=1,D=0] = E[Y_0|Z=1,D=0,N]$$
$$E[Y|Z=0,D=1] = E[Y_1|Z=0,D=1,A]$$
$$E[Y|Z=0,D=0] = P(N|Z=0,D=0)\,E[Y_0|Z=0,D=0,N] + P(C|Z=0,D=0)\,E[Y_0|Z=0,D=0,C]$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal effect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization of Z
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (nal nodes, three types)

$$E[Y|Z=1;D=1] = P[A|Z=1;D=1]\,E[Y_1|Z=1;D=1;A] + P[C|Z=1;D=1]\,E[Y_1|Z=1;D=1;C]$$
$$E[Y|Z=1;D=0] = E[Y_0|Z=1;D=0;N]$$
$$E[Y|Z=0;D=1] = E[Y_1|Z=0;D=1;A]$$
$$E[Y|Z=0;D=0] = P[N|Z=0;D=0]\,E[Y_0|Z=0;D=0;N] + P[C|Z=0;D=0]\,E[Y_0|Z=0;D=0;C]$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal e ect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization of Z
  - LATE can be inverted from expectation functions
  - It can be estimated by the IV estimator

Generalization of LATE: Multiple IV

- We have four equations (final nodes, three types)

$$E[Y|Z=1,D=1] = P(A|Z=1,D=1)E[Y_1|Z=1,D=1,A] + P(C|Z=1,D=1)E[Y_1|Z=1,D=1,C]$$
$$E[Y|Z=1,D=0] = E[Y_0|Z=1,D=0,N]$$
$$E[Y|Z=0,D=1] = E[Y_1|Z=0,D=1,A]$$
$$E[Y|Z=0,D=0] = P(N|Z=0,D=0)E[Y_0|Z=0,D=0,N] + P(C|Z=0,D=0)E[Y_0|Z=0,D=0,C]$$

- We observe four expectations on the LHS, but cannot observe expectations with potential outcomes
- The question boils down to: Can we identify some causal effect using this system
- The answer turns out to be yes:
  - With some other conditions like randomization of
  - LATE can be inverted from expectation functions
  - It can be estimated 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 di erent TEs for di erent types (groups)

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking z_1, z_2 as instruments
- 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
- " $0, 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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
- " 0; 1 , determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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
- " $0; 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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
- " $0; 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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 instruments z_1 and z_2
- " 0; 1 , determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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
- " $0; 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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
- " $0; 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- 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
- " $0; 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- We run regressions taking $z_1; z_2$ as instruments
- Assuming monotonicity for both z_1 and z_2
- The corresponding IV estimator can be derived as:

$$ {}_{2SLS} \qquad LATE_1 \qquad 1 \qquad LATE_2 $$

- $LATE_1; LATE_2$ are LATEs for instrument z_1 and z_2
- " $0; 1$, determined by rst stage relation
- Larger weight is given to IV with larger rst stage power
- See Angrist and Pischke (2009) Chapter 4.5.1

Generalization of LATE: Multivalued Treatment

- Now we consider multivalued treatment and binary IV: Average Causal Response (ACR)
- Assume that we have treatment $s \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}; s_{0i}, s_{1i}\} \perp Z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \; \forall i$ 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 treatment $s \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}; s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \; \forall i$ 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 treatment $s \in \{0, 1, 2, ..., \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, ..., Y_{\bar{s}i}; s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \; \forall i$ 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 treatment $s_i \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}, s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \ \forall i$ 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 treatment $s \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}; s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \; \forall i$ 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 treatment $s \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}, s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \; \forall i$ 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 treatment $s \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}, s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \; \forall i$ 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 treatment $s \in \{0, 1, 2, \dots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \dots, Y_{\bar{s}i}, s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \ \forall i$ 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 treatment $s \in \{0, 1, 2, ..., \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, ..., Y_{\bar{s}i}, s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \ \forall i$ 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 treatment $s \in \{0, 1, 2, \ldots, \bar{s}\}$
- 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: $\{Y_{0i}, Y_{1i}, \ldots, Y_{\bar{s}i}, s_{0i}, s_{1i}\} \perp z_i$
  - ACR2 First stage existence: $E[s_{1i} - s_{0i}] \neq 0$
  - ACR3 Monotonicity: $s_{1i} - s_{0i} \geq 0 \ \forall i$ 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 identi es a weighted average of the unit causal response

When ACR1, ACR2, and ACR3 hold, we have

Generalization of LATE: Multivalued Treatment

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

Theorem 4.5.3 in Angrist and Pischke (2009)

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

$$\frac{E[Y_i|z_i=1]-E[Y_i|z_i=0]}{E[s_i|z_i=1]-E[s_i|z_i=0]} = \sum_{s=1}^{\bar{s}} \omega_s E[Y_{si}-Y_{s-1,i}|s_{1i} \geq s > s_{0i}],$$

$$\text{where } \omega_s = \frac{P[s_{1i} \geq s > s_{0i}]}{\sum_{j=1}^{\bar{s}} P[s_{1i} \geq j > s_{0i}]}$$

Generalization of LATE: Multivalued Treatment

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

Theorem 4.5.3 in Angrist and Pischke (2009)

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

$$\frac{E[Y_i|z_i = 1] - E[Y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]} = \sum_{s=1}^{\bar{s}} \omega_s E[Y_{si} - Y_{s-1,i}|s_{1i} \geq s > s_{0i}],$$

$$\text{where } \omega_s = \frac{P[s_{1i} \geq s > s_{0i}]}{\sum_{j=1}^{\bar{s}} P[s_{1i} \geq j > s_{0i}]}$$

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 $s = 0$ when $z = 0$, but choose $s = 2, 3, \ldots, S$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, \ldots, S$ when $z = 1$
- We then average over all units/steps with a weight ω_s
- ω_s is the proportion of compliers involved in this unit change from $s - 1$ to s
- It is a normalization, with ω_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 choose $s = 0$ when $z = 0$, but choose $s = 2, 3, \ldots, s$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, \ldots, s$ when $z = 1$
- We then average over all units/steps with a weight ω_s
- ω_s is the proportion of compliers involved in this unit change from $s - 1$ to s
- It is a normalization, with ω_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 $s \geq 2$ includes compliers
  - who chooses 0 when $z = 0$, but chooses $2; 3; \ldots; s$ when $z = 1$
  - who chooses 1 when $z = 0$, but chooses $2; 3; \ldots; s$ when $z = 1$
- We then average over all units/steps with a weight ω_s
- ω_s is the proportion of compliers involved in this unit change from $s-1$ to s
- It is a normalization, with ω_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 e ect
- 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 ω_s
- ω_s is the proportion of compliers involved in this unit change from s 1 to s
- It is a normalization, with ω_s summing up to 1 over s

Generalization of LATE: Multivalued Treatment

- $Y_{si} \quad Y_{s \; 1;i}$ 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 $s \quad 1$ to $s \quad 2$ includes compliers
  - who chooses 0 when $z \quad 0$, but chooses $2; 3; \ldots; s$ when $z \quad 1$
  - who chooses 1 when $z \quad 0$, but chooses $2; 3; \ldots; s$ when $z \quad 1$
- We then average over all units/steps with a weight $!_s$
- $!_s$ is the proportion of compliers involved in this unit change from $s \quad 1$ to s
- It is a normalization, with $!_s$ 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 $2, 3, \ldots, s$ when $z = 1$
  - who chooses 1 when $z = 0$, but chooses $2, 3, \ldots, s$ when $z = 1$
- We then average over all units/steps with a weight ω_s
- ω_s is the proportion of compliers involved in this unit change from $s-1$ to s
- It is a normalization, with ω_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 choose $s = 0$ when $z = 0$, but choose $s = 2, 3, \ldots, s$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, \ldots, s$ when $z = 1$
- We then average over all units/steps with a weight ω_s
- ω_s is the proportion of compliers involved in this unit change from $s - 1$ to s
- It is a normalization, with ω_s summing up to 1 over s

Generalization of LATE: Multivalued Treatment

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

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 rst thought of this issue: transit multivalued IV to multiple binary IVs

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 rst thought of this issue: transit multivalued IV to multiple binary IVs

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 rst thought of this issue: transit multivalued IV to multiple binary IVs

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 rst thought of this issue: transit multivalued IV to multiple binary IVs

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 binary IVs

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 $z = 0; 1; 2$, 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 $z = 0, 2$
- $z_2 = 1$ if $z = 2$; $z_2 = 0$ if $z = 0, 1$

Generalization of LATE: Multivalued IV

- We can rst decompose the multivalued IV to multiple dummies
- Each dummy represents a speci c value of IV
- For example, if z 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 rst decompose the multivalued IV to multiple dummies
- Each dummy represents a speci c value of IV
- For example, if z $0;1;2$, 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 rst decompose the multivalued IV to multiple dummies
- Each dummy represents a speci c value of IV
- For example, if z 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 rst decompose the multivalued IV to multiple dummies
- Each dummy represents a speci c value of IV
- For example, if z $0; 1; 2$, 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

- 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

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

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

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

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

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

Generalization of LATE: Multivalued IV

- Because for the group of people with $z_1 \quad 0$
- They can be either $z \quad 0$ or $z \quad 2$
- If we have $D_i \; z_i \quad 0 \; \$ \; D_i \; z_i \quad 1 \; \$ \; D_i \; z_i \quad 2$
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from $D_i \; z_i \quad 0$ to $D_i \; z_i \quad 1$
  - Some people go from $D_i \; z_i \quad 2$ to $D_i \; z_i \quad 1$
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with z_1 0
- They can be either z 0 or z 2
- If we have D_i z_i 0 \$ D_i z_i 1 \$ D_i z_i 2
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from D_i z_i 0 to D_i z_i 1
  - Some people go from D_i z_i 2 to D_i z_i 1
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with $z_1 \quad 0$
- They can be either $z \quad 0$ or $z \quad 2$
- If we have $D_i \; z_i \quad 0$ \$ $D_i \; z_i \quad 1$ \$ $D_i \; z_i \quad 2$
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from $D_i \; z_i \quad 0$ to $D_i \; z_i \quad 1$
  - Some people go from $D_i \; z_i \quad 2$ to $D_i \; z_i \quad 1$
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with $z_1 \quad 0$
- They can be either $z \quad 0$ or $z \quad 2$
- If we have $D_i \; z_i \quad 0 \; \$ \; D_i \; z_i \quad 1 \; \$ \; D_i \; z_i \quad 2$
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from $D_i \; z_i \quad 0$ to $D_i \; z_i \quad 1$
  - Some people go from $D_i \; z_i \quad 2$ to $D_i \; z_i \quad 1$
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with z_1 0
- They can be either z 0 or z 2
- If we have D_i z_i 0 \$ D_i z_i 1 \$ D_i z_i 2
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from D_i z_i 0 to D_i z_i 1
  - Some people go from D_i z_i 2 to D_i z_i 1
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with z_1 0
- They can be either z 0 or z 2
- If we have D_i z_i 0 \$ D_i z_i 1 \$ D_i z_i 2
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from D_i z_i 0 to D_i z_i 1
  - Some people go from D_i z_i 2 to D_i z_i 1
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with $z_1 \quad 0$
- They can be either $z \quad 0$ or $z \quad 2$
- If we have $D_i \; z_i \quad 0 \; \$ \; D_i \; z_i \quad 1 \; \$ \; D_i \; z_i \quad 2$
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from $D_i \; z_i \quad 0$ to $D_i \; z_i \quad 1$
  - Some people go from $D_i \; z_i \quad 2$ to $D_i \; z_i \quad 1$
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with $z_1 \quad 0$
- They can be either $z \quad 0$ or $z \quad 2$
- If we have $D_i \;\; z_i \quad 0 \;\; \$ \;\; D_i \;\; z_i \quad 1 \;\; \$ \;\; D_i \;\; z_i \quad 2$
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from $D_i \;\; z_i \quad 0$ to $D_i \;\; z_i \quad 1$
  - Some people go from $D_i \;\; z_i \quad 2$ to $D_i \;\; z_i \quad 1$
- Changes of treatment D have opposite directions for these two groups
- It violates the monotonicity assumption

Generalization of LATE: Multivalued IV

- Because for the group of people with $z_1 \quad 0$
- They can be either $z \quad 0$ or $z \quad 2$
- If we have $D_i \; z_i \quad 0 \;\$\; D_i \; z_i \quad 1 \;\$\; D_i \; z_i \quad 2$
- Then, when z_1 changes from 0 to 1, we have two groups of people:
  - Some people go from $D_i \; z_i \quad 0$ to $D_i \; z_i \quad 1$
  - Some people go from $D_i \; z_i \quad 2$ to $D_i \; z_i \quad 1$
- Changes of treatment D have opposite directions for these two groups
- It violates 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 speci c context

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 speci c context

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 speci c context

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 speci c context

Generalization of LATE: Multivalued IV

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'):

$$\text{If } p \cdot x(p',w') \leq w, \text{ and } x(p',w') \neq x(p,w), \text{ then } p' \cdot x(p,w) > w'$$

- 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 $(x_A \succ_R x_B)$
- This is Weak Axiom of Revealed Preference

Generalization of LATE: Multivalued IV

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'):

$$\text{If } p \cdot x(p',w') \leq w; \text{ and } x(p',w') \neq x(p,w); \text{ then } p' \cdot x(p,w) > w'$$

- 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 ($x_A \succ_R x_B$)
- This is Weak Axiom of Revealed Preference

Generalization of LATE: Multivalued IV

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'):

$$\text{If } p \cdot x(p', w') \leq w; \text{ and } x(p', w') \neq x(p, w); \text{ then } p' \cdot x(p, w) > w'$$

- 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 ($x_A \succeq_R x_B$)
- This is Weak Axiom of Revealed Preference

Generalization of LATE: Multivalued IV

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'):

$$\text{If } p \cdot x(p', w') \leq w; \text{ and } x(p', w') \neq x(p, w); \text{ then } p' \cdot x(p, w) > w'$$

- 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. $(x_A \succ_R x_B)$
- This is Weak Axiom of Revealed Preference

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 $(p^1, w^1), \ldots, (p^N, w^N)$ with $x(p^{n+1}, w^{n+1}) \neq x(p^n, w^n)$ for all $n \leq N-1$, we have $p^N \cdot x(p^1, w^1) > w^N$, whenever $p^n \cdot x(p^{n+1}, w^{n+1}) \leq w^n$ for all $n \leq N-1$

- SARP adds transitivity to WARP
- If $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2} \ldots x_2 \succ_R x_1$, 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 $(p^1, w^1), ..., (p^N, w^N)$ with $x(p^{n+1}, w^{n+1}) \neq x(p^n, w^n)$ for all $n \leq N-1$, we have $p^N \cdot x(p^1, w^1) > w^N$, whenever $p^n \cdot x(p^{n+1}, w^{n+1}) \leq w^n$ for all $n \leq N-1$

- SARP adds transitivity to WARP
- If $x_N \succsim_R x_{N-1}, x_{N-1} \succsim_R x_{N-2} ... x_2 \succsim_R x_1$, we have $x_N \succsim_R x_1$
- 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 $x(p^{n+1},w^{n+1}) \neq x(p^n,w^n)$ for all $n \le N-1$, we have $p^N \cdot x(p^1,w^1) > w^N$, whenever $p^n \cdot x(p^{n+1},w^{n+1}) \le w^n$ for all $n \le N-1$

- SARP adds transitivity to WARP
- If $x_N \succsim_R x_{N-1}; x_{N-1} \succsim_R x_{N-2} \dots x_2 \succsim_R x_1$, we have $x_N \succsim_R x_1$
- 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 $(p^1, w^1), \ldots, (p^N, w^N)$ with $x(p^{n+1}, w^{n+1}) \neq x(p^n, w^n)$ for all $n \leq N-1$, we have $p^N \cdot x(p^1, w^1) > w^N$, whenever $p^n \cdot x(p^{n+1}, w^{n+1}) \leq w^n$ for all $n \leq N-1$

- SARP adds transitivity to WARP
- If $x_N \succ_R x_{N-1}, x_{N-1} \succ_R x_{N-2} \ldots x_2 \succ_R x_1$, 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 $(p^1, w^1), \ldots, (p^N, w^N)$ with $x(p^{n+1}, w^{n+1}) \neq x(p^n, w^n)$ for all $n \leq N-1$, we have $p^N \cdot x(p^1, w^1) > w^N$, whenever $p^n \cdot x(p^{n+1}, w^{n+1}) \leq w^n$ for all $n \leq N-1$

- SARP adds transitivity to WARP
- If $x_N \succsim_R x_{N-1}, x_{N-1} \succsim_R x_{N-2}, \ldots, x_2 \succsim_R x_1$, we have $x_N \succsim_R x_1$
- 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 $(p^1,w^1),\ldots,(p^N,w^N)$ with $x(p^{n+1},w^{n+1}) \neq x(p^n,w^n)$ for all $n \leq N-1$, we have $p^N \cdot x(p^1,w^1) > w^N$, whenever $p^n \cdot x(p^{n+1},w^{n+1}) \leq w^n$ for all $n \leq N-1$

- SARP adds transitivity to WARP
- If $x_N \succ_R x_{N-1}; x_{N-1} \succ_R x_{N-2} \ldots x_2 \succ_R x_1$, 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([illegible])
  - Experimental group: Vouchers, available only for housing lease in low-poverty neighborhood([illegible])
  - Section 8 group: Vouchers, available for any housing lease ([illegible])
- There are three choices (three values of treatment)
  - Not relocating ([illegible])
  - Relocating to a low poverty neighborhood ([illegible])
  - Relocating to a high poverty neighborhood ([illegible])

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 ($d = 1$)
  - Relocating to a low poverty neighborhood ($d = 2$)
  - Relocating to a high poverty neighborhood ($d = 3$)

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 ($d = 1$)
  - Relocating to a low poverty neighborhood ($d = 2$)
  - Relocating to a high poverty neighborhood ($d = 3$)

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 (1)
  - Relocating to a low poverty neighborhood (2)
  - Relocating to a high poverty neighborhood (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 (1)
  - Relocating to a low poverty neighborhood (2)
  - Relocating to a high poverty neighborhood (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 ($d = 1$)
  - Relocating to a low poverty neighborhood ($d = 2$)
  - Relocating to a high poverty neighborhood ($d = 3$)

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 27 types of agents in total:
  3^3 27, 3 moving decisions under 3 possible vouchers
- In general, behavior types of agent grow exponentially when you have more IV values
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal e ect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- Thus, we have 27 types of agents in total:
  3^3 27, 3 moving decisions under 3 possible vouchers
- In general, behavior types of agent grow exponentially when you have more IV values
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal e ect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- Thus, we have 27 types of agents in total:
  $3^3 \quad 27$, 3 moving decisions under 3 possible vouchers
- In general, behavior types of agent grow exponentially when you have more IV values
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal e ect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- Thus, we have 27 types of agents in total:
  3^3 27, 3 moving decisions under 3 possible vouchers
- In general, behavior types of agent grow exponentially when you have more IV values
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal e ect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- Thus, we have 27 types of agents in total:
  3^3 27, 3 moving decisions under 3 possible vouchers
- In general, behavior types of agent grow exponentially when you have more IV values
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal e ect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- Thus, we have 27 types of agents in total:
  3^3 27, 3 moving decisions under 3 possible vouchers
- In general, behavior types of agent grow exponentially when you have more IV values
- Only 12 available equations for observed expectations
- It is impossible to invert a linear system of 9 equations to identify any causal e ect with 27 behavior types
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- Let $u_\omega(k, t)$ be the utility function of family ω (k consumption, t relocation choice)
- Let $W_\omega(z, t)$ be the budget set of family ω under relocation decision $t \in \{1, 2, 3\}$ and MTO voucher $z \in \{z_1, z_2, z_3\}$
- Let $S_\omega = [C_\omega(z_1), C_\omega(z_2), C_\omega(z_3)]$ denote the type of family ω, defined by relocation responses $C_\omega(z)$ given different vouchers

Generalization of LATE: Multivalued IV

- Let $u_\omega(k; t)$ be the utility function of family ω (k consumption, t relocation choice)
- Let $W_\omega(z; t)$ be the budget set of family ω under relocation decision $t \in \{1; 2; 3\}$ and MTO voucher $z \in \{z_1; z_2; z_3\}$
- Let $S_\omega = (C_\omega(z_1); C_\omega(z_2); C_\omega(z_3))$ denote the type of family ω, defined by relocation responses $C_\omega(z)$ given different vouchers

Generalization of LATE: Multivalued IV

- Let $u_\omega(k; t)$ be the utility function of family ω (k consumption, t relocation choice)
- Let $W_\omega(z; t)$ be the budget set of family ω under relocation decision $t \in \{1; 2; 3\}$ and MTO voucher $z \in \{z_1; z_2; z_3\}$
- Let $S_\omega = (C_\omega(z_1); C_\omega(z_2); C_\omega(z_3))$ denote the type of family ω, defined by relocation responses $C_\omega(z)$ given different vouchers

Generalization of LATE: Multivalued IV

- Let $u_\omega(k, t)$ be the utility function of family ω (k consumption, t relocation choice)
- Let $W_\omega(z, t)$ be the budget set of family ω under relocation decision $t \in \{1, 2, 3\}$ and MTO voucher $z \in \{z_1, z_2, z_3\}$
- Let $S_\omega = [C_\omega(z_1), C_\omega(z_2), C_\omega(z_3)]$ denote the type of family ω, defined by relocation responses $C_\omega(z)$ given different vouchers

Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

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

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

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

$$W_! \quad z_1; 2 \; \text{à} \, W_! \quad z_2; 2 \qquad W_! \quad z_3; 2 \tag{1}$$

$$W_! \quad z_1; 3 \qquad W_! \quad z_2; 3 \; \text{à} \; W_! \quad z_3; 3 \tag{2}$$

$$W_! \quad z_1; 1 \qquad W_! \quad z_1; 2 \qquad W_! \quad z_1; 3 \qquad W_! \quad z_2; 1 \qquad W_! \quad z_2; 3 \qquad W_! \quad z_3; 1 \tag{3}$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

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

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

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

$$W_! \quad z_1; 2 \;\; \text{à} W_! \quad z_2; 2 \qquad W_! \quad z_3; 2 \tag{1}$$

$$W_! \quad z_1; 3 \qquad W_! \quad z_2; 3 \;\; \text{à} \; W_! \quad z_3; 3 \tag{2}$$

$$W_! \quad z_1; 1 \qquad W_! \quad z_1; 2 \qquad W_! \quad z_1; 3 \qquad W_! \quad z_2; 1 \qquad W_! \quad z_2; 3 \qquad W_! \quad z_3; 1 \tag{3}$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

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

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

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

$$W_! \quad z_1; 2 \text{ à} W_! \quad z_2; 2 \qquad W_! \quad z_3; 2 \tag{1}$$

$$W_! \quad z_1; 3 \qquad W_! \quad z_2; 3 \text{ à } W_! \quad z_3; 3 \tag{2}$$

$$W_! \quad z_1; 1 \qquad W_! \quad z_1; 2 \qquad W_! \quad z_1; 3 \qquad W_! \quad z_2; 1 \qquad W_! \quad z_2; 3 \qquad W_! \quad z_3; 1 \tag{3}$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

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

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

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

$$W_! \quad z_1; 2 \;\; \text{à} W_! \quad z_2; 2 \qquad W_! \quad z_3; 2 \tag{1}$$

$$W_! \quad z_1; 3 \qquad W_! \quad z_2; 3 \;\; \text{à} \; W_! \quad z_3; 3 \tag{2}$$

$$W_! \quad z_1; 1 \qquad W_! \quad z_1; 2 \qquad W_! \quad z_1; 3 \qquad W_! \quad z_2; 1 \qquad W_! \quad z_2; 3 \qquad W_! \quad z_3; 1 \tag{3}$$

- 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 z_1 subsidies nothing
  - Experimental group z z_2 subsidies relocating to low poverty neighborhood
  - Section 8 group z z_3 subsidies any relocation

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

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

$$W_{!} \quad z_1; 2 \quad \text{à} W_{!} \quad z_2; 2 \qquad W_{!} \quad z_3; 2 \tag{1}$$

$$W_{!} \quad z_1; 3 \qquad W_{!} \quad z_2; 3 \quad \text{à} \; W_{!} \quad z_3; 3 \tag{2}$$

$$W_{!} \quad z_1; 1 \qquad W_{!} \quad z_1; 2 \qquad W_{!} \quad z_1; 3 \qquad W_{!} \quad z_2; 1 \qquad W_{!} \quad z_2; 3 \qquad W_{!} \quad z_3; 1 \tag{3}$$

- 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 z_1 subsidies nothing
  - Experimental group z z_2 subsidies relocating to low poverty neighborhood
  - Section 8 group z z_3 subsidies any relocation

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

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

$$W_! \quad z_1;2 \quad \grave{a} W_! \quad z_2;2 \qquad W_! \quad z_3;2 \tag{1}$$

$$W_! \quad z_1;3 \qquad W_! \quad z_2;3 \quad \grave{a} \; W_! \quad z_3;3 \tag{2}$$

$$W_! \quad z_1;1 \qquad W_! \quad z_1;2 \qquad W_! \quad z_1;3 \qquad W_! \quad z_2;1 \qquad W_! \quad z_2;3 \qquad W_! \quad z_3;1 \tag{3}$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

- (1): If you choose to relocate to low poverty neighborhodd (2), your budget would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhodd (3), your budget 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 budget would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your budget 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 budget would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your budget 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 budget would be higher if you are in Experimental or Section 8 groups
- (2): If you choose to relocate to high poverty neighborhood (3), your budget 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

Lemma L-1 Pinto (2015)

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

$1{:}C \quad z_1 \quad 2 \quad C \quad z_2 \quad 2{:}C \quad z_3 \quad 1$

$2{:}C \quad z_1 \quad 3 \quad C \quad z_2 \quad 1{:}C \quad z_3 \quad 1$

$3{:}C \quad z_2 \quad 1 \quad C \quad z_1 \quad 1{:}C \quad z_3 \quad 2$

$4{:}C \quad z_2 \quad 3 \quad C \quad z_1 \quad 3{:}C \quad z_3 \quad 3$

$5{:}C \quad z_3 \quad 1 \quad C \quad z_1 \quad 1{:}C \quad z_2 \quad 1$

$6{:}C \quad z_3 \quad 2 \quad C \quad z_2 \quad 2$

- 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:

$$1{:}C_! \quad z_1 \quad 2 \quad C_! \quad z_2 \quad 2; C_! \quad z_3 \quad j \quad 1$$
$$2{:}C_! \quad z_1 \quad 3 \quad C_! \quad z_2 \quad j \quad 1; C_! \quad z_3 \quad j \quad 1$$
$$3{:}C_! \quad z_2 \quad 1 \quad C_! \quad z_1 \quad 1; C_! \quad z_3 \quad j \quad 2$$
$$4{:}C_! \quad z_2 \quad 3 \quad C_! \quad z_1 \quad 3; C_! \quad z_3 \quad 3$$
$$5{:}C_! \quad z_3 \quad 1 \quad C_! \quad z_1 \quad 1; C_! \quad z_2 \quad 1$$
$$6{:}C_! \quad z_3 \quad 2 \quad C_! \quad z_2 \quad 2$$

- 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:

$$
\begin{array}{llllllllll}
1{:}C_{!} & z_1 & 2 & C_{!} & z_2 & & 2;\, C_{!} & z_3 & j & 1 \\
2{:}C_{!} & z_1 & 3 & C_{!} & z_2 & j & 1;\, C_{!} & z_3 & j & 1 \\
3{:}C_{!} & z_2 & 1 & C_{!} & z_1 & & 1;\, C_{!} & z_3 & j & 2 \\
4{:}C_{!} & z_2 & 3 & C_{!} & z_1 & & 3;\, C_{!} & z_3 & & 3 \\
5{:}C_{!} & z_3 & 1 & C_{!} & z_1 & & 1;\, C_{!} & z_2 & & 1 \\
6{:}C_{!} & z_3 & 2 & C_{!} & z_2 & & 2 & & &
\end{array}
$$

- 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:

$$
\begin{array}{lllllllll}
1{:}C_! & z_1 & 2 & C_! & z_2 & & 2;\, C_! & z_3 & j \;\; 1 \\
2{:}C_! & z_1 & 3 & C_! & z_2 & j & 1;\, C_! & z_3 & j \;\; 1 \\
3{:}C_! & z_2 & 1 & C_! & z_1 & & 1;\, C_! & z_3 & j \;\; 2 \\
4{:}C_! & z_2 & 3 & C_! & z_1 & & 3;\, C_! & z_3 & \;\;\;\; 3 \\
5{:}C_! & z_3 & 1 & C_! & z_1 & & 1;\, C_! & z_2 & \;\;\;\; 1 \\
6{:}C_! & z_3 & 2 & C_! & z_2 & & 2 & &
\end{array}
$$

- Test yourself, explain all these six inequalities

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different vouchers: SARP (equation (37) of the paper)
    Preference on one moving choice to the other (“under vouchers” can be carried over to another voucher situation”, if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different voucher z by SARP [equation (37) of the paper]
    Preference on one moving choice t to the other t' under voucher z can be carried over to another voucher situation z', if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different voucher z by SARP [equation (37) of the paper]

    Preference on one moving choice t to the other t' under voucher z can be carried over to another voucher situation z', if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different voucher z by SARP [equation (37) of the paper]
    Preference on one moving choice t to the other $\bar{t}$ under voucher z can be carried over to another voucher situation $\bar{z}$, if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different vouchers z by SARP [equation (37) of the paper]
    Preference on one moving choice t to the other $\bar{t}$ under voucher z can be carried over to another voucher situation $\bar{z}$, if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different voucher z by SARP [equation (37) of the paper]
    Preference on one moving choice t to the other $\bar{t}$ under voucher z can be carried over to another voucher situation $\bar{z}$, if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- I will not show the proof of this Lemma L-1
- Please refer to Appendix A of Pinto (2015)
- The basic idea is as follows:
  - First, combine moving choice and consumption into one choice bundle
  - Second, derive the condition of preference transitivity under different voucher z by SARP [equation (37) of the paper]
    Preference on one moving choice t to the other $\bar{t}$ under voucher z can be carried over to another voucher situation $\bar{z}$, if the budget of the preferred choice is weakly extended and the unpreferred is weakly shrunk.
  - Repeatedly apply the rule derived from the second step

Generalization of LATE: Multivalued IV

- Let's take 1 as an example:

  C_i z_1 2 means that under z_1, we choose t 2 over t 1; 3 when no action is nanced.

  Then under z_2, moving to rich t 2 is nanced, extending the budget. But no moving t 1 or moving to poor t 3 are not. Thus, t 2 is still preferred than t 1; 3 in this case.

  Similarly, under z_3, moving to rich and poor t 2; 3 are both nanced, extending the budget (but may be in di erent magnitude, cannot compare). But no moving t 1 is not. Thus, t 2 is still preferred than t 1, but not necessarily t 3 in this case.

Generalization of LATE: Multivalued IV

- Let's take 1 as an example:

  $C_{!}$ z_1 2 means that under z_1, we choose t 2 over t 1; 3 when no action is nanced.

  Then under z_2, moving to rich t 2 is nanced, extending the budget. But no moving t 1 or moving to poor t 3 are not. Thus, t 2 is still preferred than t 1; 3 in this case.

  Similarly, under z_3, moving to rich and poor t 2; 3 are both nanced, extending the budget (but may be in di erent magnitude, cannot compare). But no moving t 1 is not. Thus, t 2 is still preferred than t 1, but not necessarily t 3 in this case.

Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

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

- To eliminate cases like $C_i(z_1) \ldots 2; C_i(z_2) \ldots 2; C_i(z_3) \ldots 3$
- Using all above, we can eliminate the number of types from 27 to 7

Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

Assumption A-3 Pinto (2015)

For each family ω, and for $z, z' \in \{z_1, z_2, z_3\}$, if $C_\omega(z) = t$ and $W_\omega(z, t)$ is a proper subset of $W_\omega(z', t)$, then $C_\omega(z') = t$

- To eliminate cases like $C_\omega(z_1) = 2; C_\omega(z_2) = 2; C_\omega(z_3) = 3$
- Using all above, we can eliminate the number of types from 27 to 7

Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

Assumption A-3 Pinto (2015)

For each family ω, and for $z, z' \in \{z_1, z_2, z_3\}$, if $C_\omega(z) = t$ and $W_\omega(z, t)$ is a proper subset of $W_\omega(z', t)$, then $C_\omega(z') = t$

- To eliminate cases like $C_\omega(z_1) = 2; C_\omega(z_2) = 2; C_\omega(z_3) = 3$
- Using all above, we can eliminate the number of types from 27 to 7

Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

Assumption A-3 Pinto (2015)

For each family ω, and for $z, z' \in \{z_1, z_2, z_3\}$, if $C_\omega(z) = t$ and $W_\omega(z, t)$ is a proper subset of $W_\omega(z', t)$, then $C_\omega(z') = t$

- To eliminate cases like $C_\omega(z_1) = 2; C_\omega(z_2) = 2; C_\omega(z_3) = 3$
- Using all above, we can eliminate the number of types from 27 to 7

Generalization of LATE: Multivalued IV

- We further assume that neighborhood is a normal good

Assumption A-3 Pinto (2015)

For each family ω, and for $z, z' \in \{z_1, z_2, z_3\}$, if $C_\omega(z) = t$ and $W_\omega(z, t)$ is a proper subset of $W_\omega(z', t)$, then $C_\omega(z') = t$

- To eliminate cases like $C_\omega(z_1) = 2; C_\omega(z_2) = 2; C_\omega(z_3) = 3$
- Using all above, we can eliminate the number of types from 27 to 7

Generalization of LATE: Multivalued IV

- Only after this elimination, can we do something for causal identi cation
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identi cation
- When statistics tools are exhausted, remember you are an economist
- Do not think rst year Micro and Macro are useless!!!

Generalization of LATE: Multivalued IV

- Only after this elimination, can we do something for causal identi cation
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identi cation
- When statistics tools are exhausted, remember you are an economist
- Do not think rst year Micro and Macro are useless!!!

Generalization of LATE: Multivalued IV

- Only after this elimination, can we do something for causal identi cation
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identi cation
- When statistics tools are exhausted, remember you are an economist
- Do not think rst year Micro and Macro are useless!!!

Generalization of LATE: Multivalued IV

- Only after this elimination, can we do something for causal identi cation
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identi cation
- When statistics tools are exhausted, remember you are an economist
- Do not think rst year Micro and Macro are useless!!!

Generalization of LATE: Multivalued IV

- Only after this elimination, can we do something for causal identi cation
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identi cation
- When statistics tools are exhausted, remember you are an economist
- Do not think rst year Micro and Macro are useless!!!

Generalization of LATE: Multivalued IV

- Only after this elimination, can we do something for causal identi cation
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identi cation
- When statistics tools are exhausted, remember you are an economist
- Do not think rst 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 de ned on a policy-speci c 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 de ned on a policy-speci c 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 de ned on a policy-speci c 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 de ned on a policy-speci c 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 de ned on a policy-speci c 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 de ned on a policy-speci c 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 $j \quad 0; 1$ be the treatment, $Y_1; Y_0$ be the potential outcomes

$$Y_1 \qquad {}_1\ X ; U_1 \tag{4}$$
$$Y_0 \qquad {}_0\ X ; U_0 \tag{5}$$

- X is a set of control variables, U 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

$$Y_1 \qquad {}_1\, X; U_1 \tag{4}$$

$$Y_0 \qquad {}_0\, X; U_0 \tag{5}$$

- X is a set of control variables, U 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

$$Y_1 \qquad {}_1\; X; U_1 \tag{4}$$

$$Y_0 \qquad {}_0\; X; U_0 \tag{5}$$

- X is a set of control variables, U 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

$$Y_1 \qquad {}_1\, X; U_1 \tag{4}$$

$$Y_0 \qquad {}_0\, X; U_0 \tag{5}$$

- X is a set of control variables, U 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 $j \quad 0; 1$ be the treatment, $Y_1; Y_0$ be the potential outcomes

$$Y_1 \qquad {}_1\, X; U_1 \tag{4}$$

$$Y_0 \qquad {}_0\, X; U_0 \tag{5}$$

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

MTE: Choice Model

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

$$\tilde{D} = \mu_D(Z) - V; \quad D = 1 \text{ if } \tilde{D} \geq 0; \ D = 0 \text{ otherwise} \tag{6}$$

- Here, μ_0 is normalized to zero
- 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 (Z, 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

$$\tilde{D} \qquad {}_D Z \qquad V; \quad D \quad 1 \text{ if } \tilde{D} \;'\; 0; \; D \quad 0 \text{ otherwise} \qquad (6)$$

- Here, ${}_0$ is normalized to zero
- 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 Z, 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

$$\tilde{D} \quad {}_D\, Z \quad V; \quad D \quad 1 \text{ if } \tilde{D} \; ' \; 0; \; D \quad 0 \text{ otherwise} \tag{6}$$

- Here, ${}_0$ is normalized to zero
- Z is an instrument that can change individual's choice, 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 Z, 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

$$\tilde{D} \qquad {}_D \; Z \qquad V; \quad D \quad 1 \text{ if } \tilde{D} \; ' \; 0; \; D \quad 0 \text{ otherwise} \tag{6}$$

- Here, ${}_0$ is normalized to zero
- 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 Z, 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

$$\tilde{D} \quad _D Z \quad V; \quad D \quad 1 \text{ if } \tilde{D} \ ' \ 0; \ D \quad 0 \text{ otherwise} \tag{6}$$

- Here, $_0$ is normalized to zero
- 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 Z, 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

$$\tilde{D} \qquad {}_D Z \qquad V; \quad D \quad 1 \text{ if } \tilde{D} \; ' \; 0; \; D \quad 0 \text{ otherwise} \tag{6}$$

- Here, ${}_0$ is normalized to zero
- 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 Z, 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

$$\tilde{D} \qquad {}_D\; Z \qquad V; \quad D \quad 1 \text{ if } \tilde{D} \;'\; 0; \; D \quad 0 \text{ otherwise} \tag{6}$$

- Here, $_0$ is normalized to zero
- 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 Z, 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 of the instrument
- (A-2) $\mu_D(Z)$ 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) $E|Y_1|, E|Y_0| < \infty$ 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 ve assumptions for this model
- (A-1) $U_0; U_1; V$ are independent of Z conditional on X
  Independence of the instrument
- (A-2) $_D$ 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 nite
- (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 ve assumptions for this model
- (A-1) $U_0; U_1; V$ are independent of Z conditional on X
  Independence of the instrument
- (A-2) $_D$ 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 nite
- (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 of the instrument
- (A-2) $\mu_D(Z)$ 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 ve assumptions for this model
- (A-1) $U_0; U_1; V$ are independent of Z conditional on X
  Independence of the instrument
- (A-2) $_D Z$ 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 nite
- (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 ve assumptions for this model
- (A-1) $U_0; U_1; V$ are independent of Z conditional on X
  Independence of the instrument
- (A-2) $_D$ 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 nite
- (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 ve assumptions for this model
- (A-1) $U_0; U_1; V$ are independent of Z conditional on X
  Independence of the instrument
- (A-2) $_D$ 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 nite
- (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 ve assumptions for this model
- (A-1) $U_0; U_1; V$ are independent of Z conditional on X
  Independence of the instrument
- (A-2) $_D \; Z$ 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 nite
- (A-5) $0 \;\$\; \Pr\; D \quad 1¶X \;\; \$\; 1$
  Possible to have $D \quad 1$ or $D \quad 0$ at any point of X

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost C Z_1 V_C in modern 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 (go to modern sector)

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost C Z_1 V_C in modern 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 (go to modern sector)

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost C Z_1 V_C in modern 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 (go to modern sector)

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost C Z_1 V_C in modern 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 (go to modern sector)

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost C Z_1 V_C in modern 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 (go to modern sector)

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost $C \quad Z_1 \quad V_C$ in modern 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 (go to modern sector)

MTE: Choice Model

- An example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payo , there is relative working cost C Z_1 V_C in modern 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 (go to modern sector)

MTE: Choice Model

- Assume additively separable in U

$$Y_1 \quad {}_1 X \quad U_1$$
$$Y_0 \quad {}_0 X \quad U_0$$
$$\tilde{D} \quad {}_1 X \quad U_1 \quad {}_0 X \quad U_0 \quad Z_1 \quad V_C; \quad D \quad 1 \text{ if } \tilde{D} \quad 0; \; D \quad 0 \text{ otherwise}$$

- We can transform to have:

$$\tilde{D} \quad \underbrace{{}_1 X \quad {}_0 X \quad Z_1}_{D} \quad \underbrace{U_1 \quad U_0 \quad V_C}_{V}$$

- In this case, we have $V \quad U_1 \quad U_0 \quad V_C$
- Positive sorting: $Cov \; U_1 \quad U_0; U_1 \quad U_0 \quad V_C \quad \%0$

MTE: Choice Model

- Assume additively separable in U

$$Y_1 \quad {}_1 X \quad U_1$$
$$Y_0 \quad {}_0 X \quad U_0$$
$$\tilde{D} \quad {}_1 X \quad U_1 \quad {}_0 X \quad U_0 \quad Z_1 \quad V_C; \quad D \quad 1 \text{ if } \tilde{D} \; ' \; 0; \; D \quad 0 \text{ otherwise}$$

- We can transform to have:

$$\tilde{D} \quad \underbrace{{}_1 X \quad {}_0 X \quad Z_1}_{D} \quad \underbrace{U_1 \quad U_0 \quad V_C}_{V}$$

- In this case, we have $V \quad U_1 \quad U_0 \quad V_C$
- Positive sorting: $Cov \; U_1 \quad U_0; U_1 \quad U_0 \quad V_C \; \%0$

MTE: Choice Model

- Assume additively separable in U

$$Y_1 \quad {}_1 X \quad U_1$$
$$Y_0 \quad {}_0 X \quad U_0$$
$$\tilde{D} \quad {}_1 X \quad U_1 \quad {}_0 X \quad U_0 \quad Z_1 \quad V_C; \quad D \quad 1 \text{ if } \tilde{D} \; ' \; 0; \; D \quad 0 \text{ otherwise}$$

- We can transform to have:

$$\tilde{D} \quad \underbrace{{}_1 X \quad {}_0 X \quad Z_1}_{D} \quad \underbrace{U_1 \quad U_0 \quad V_C}_{V}$$

- In this case, we have $V \quad U_1 \quad U_0 \quad V_C$
- Positive sorting: $Cov \; U_1 \quad U_0; U_1 \quad U_0 \quad V_C \quad \%0$

MTE: Choice Model

- Assume additively separable in U

$$Y_1 \qquad {}_1 X \qquad U_1$$
$$Y_0 \qquad {}_0 X \qquad U_0$$
$$\tilde{D} \qquad {}_1 X \qquad U_1 \qquad {}_0 X \qquad U_0 \qquad Z_1 \qquad V_C; \quad D \qquad 1 \text{ if } \tilde{D} \ ' \ 0; \ D \qquad 0 \text{ otherwise}$$

- We can transform to have:

$$\tilde{D} \qquad \underbrace{{}_1 X \qquad {}_0 X \qquad Z_1}_{D} \qquad \underbrace{U_1 \qquad U_0 \qquad V_C}_{V}$$

- In this case, we have $V \qquad U_1 \qquad U_0 \qquad V_C$
- Positive sorting: $Cov\ U_1 \qquad U_0; U_1 \qquad U_0 \qquad V_C \quad \%0$

MTE: Choice Model

- Assume additively separable in U

$$Y_1 \quad {}_1 X \quad U_1$$
$$Y_0 \quad {}_0 X \quad U_0$$
$$\tilde{D} \quad {}_1 X \quad U_1 \quad {}_0 X \quad U_0 \quad Z_1 \quad V_C; \quad D \quad 1 \text{ if } \tilde{D} \; ' \; 0; \; D \quad 0 \text{ otherwise}$$

- We can transform to have:

$$\tilde{D} \quad \underbrace{{}_1 X \quad {}_0 X \quad Z_1}_{D} \quad \underbrace{U_1 \quad U_0 \quad V_C}_{V}$$

- In this case, we have $V \quad U_1 \quad U_0 \quad V_C$
- Positive sorting: $Cov \; U_1 \quad U_0; U_1 \quad U_0 \quad V_C \quad \%0$

MTE: Choice Model

- Let $P(Z) = \Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
  $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0, 1]$
- $F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) = \Pr(D = 1 \mid Z, X) = F_{V|X}(\mu_D(Z))$. $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) \equiv \Pr(D = 1 | Z, X) = F_{V|X}(\mu_D(Z))$. $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0, 1]$
- $F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) \equiv \Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$, $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0, 1]$
- $F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) \equiv \Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
  $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) = \Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
  $F_{V|X}$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, we have $U_D \sim \mathrm{Unif}(0,1)$
- $F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) = \Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
  $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, V) or (X, U_D)

MTE: Choice Model

- Let $P(Z) \equiv \Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
  $F_{V|X}(\cdot)$ denotes the distribution of V conditional on X
- This is the propensity score to get treated for agent with Z
- Let $U_D = F_{V|X}(V)$, 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 $F_{V|X}(\mu_D(Z)) \geq F_{V|X}(V) = U_D$ (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 (X, 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 $\mu_D(Z)$
- $\tilde{D}$ $\mu_D(Z)$ V additively separable for Z and V
- Thus, given z and z', $\forall V$ $\tilde{D}(z) \geq \tilde{D}(z')$ or $\tilde{D}(z) \leq \tilde{D}(z')$
- 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 $\mu_D(Z)$
- $\tilde{D}$ $\mu_D(Z)$ V additively separable for Z and V
- Thus, given z and z', $\forall V$ $\tilde{D}(z) \geq \tilde{D}(z')$ or $\tilde{D}(z) \leq \tilde{D}(z')$
- 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 a ect $_D$ Z
- D˜ $_D$ Z V additively separable for Z and V
- Thus, given z and z¬, ¾V 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 a ect D Z
- D˜ D Z V additively separable for Z and V
- Thus, given z and z', ¾V 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 a ect $_D$ Z
- $\tilde{D}$ $_D$ Z V additively separable for Z and V
- Thus, given z and z′, ¾V $\tilde{D}$ z ' $\tilde{D}$ z′ or $\tilde{D}$ z & $\tilde{D}$ 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 a ect $_D$ Z
- D ˜ $_D$ Z V additively separable fo Z and V
- Thus, given z and z¬, ¾V 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
- Conditional ATE is de ned as usual: ^{ATE} x E ¶X x
- MTE is de ned as the mean e ect of treatment on those for whom X x and U_D u_D V v

De nition of the MTE

The Marginal Treatment E ect is de ned as:

$$^{MTE} x; u_D \quad E \quad ¶X \quad x; U_D \quad u_D$$

MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let Y_1 Y_0
- Conditional ATE is de ned as usual: ^{ATE} x E ¶X x
- MTE is de ned as the mean e ect of treatment on those for whom X x and U_D u_D V v

De nition of the MTE

The Marginal Treatment E ect is de ned as:

$$^{MTE} \quad x; u_D \quad E \quad ¶X \quad x; U_D \quad u_D$$

MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let Y_1 Y_0
- Conditional ATE is de ned as usual: ^{ATE} x E ¶X x
- MTE is de ned as the mean e ect of treatment on those for whom X x and U_D u_D V v

De nition of the MTE

The Marginal Treatment E ect is de ned as:

$$^{MTE} \; x; u_D \quad E \quad ¶X \quad x; U_D \quad u_D$$

MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let Y_1 Y_0
- Conditional ATE is de ned as usual: ^{ATE} x E ¶X x
- MTE is de ned as the mean e ect of treatment on those for whom X x and U_D u_D V v

De nition of the MTE

The Marginal Treatment E ect is de ned as:

$$^{MTE} \quad x; u_D \quad E \quad ¶X \quad x; U_D \quad u_D$$

MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let Y_1 Y_0
- Conditional ATE is de ned as usual: ^{ATE} x E ¶X x
- MTE is de ned as the mean e ect of treatment on those for whom X x and U_D u_D V v

De nition of the MTE

The Marginal Treatment E ect is de ned as:

$$^{MTE} x; u_D \quad E \quad ¶X \quad x; U_D \quad u_D$$

MTE: De ning MTE

- Now let's de ne ATE and MTE in this model
- Let Y_1 Y_0
- Conditional ATE is de ned as usual: ^{ATE} x E $\P X$ x
- MTE is de ned as the mean e ect of treatment on those for whom X x and U_D u_D V v

De nition of the MTE

The Marginal Treatment E ect is de ned as:

$$^{MTE} \quad x; u_D \quad E \quad \P X \quad x; U_D \quad u_D$$

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 V
- 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 propensity score of 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

- MTE is a mean treatment e ect for a very speci c group of people
- People with observed characteristics X and unobserved taste on treatment M
- People with observed characteristics X who would be indi erent 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 propensity score of u_D
- Di erent from LATE, it is not de ned 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 V
- People with observed characteristics X who would be indifferent between treatment or not if they were randomly assigned a value of Z, z such that $P(z) = u_D$
- That is why it is called "marginal"
  Marginal people who have just the threshold propensity score of 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 $X = x$ and unobserved taste on treatment V
- People with observed characteristics $X = x$ who would be indifferent between treatment or not if they were randomly assigned a value of $Z = z$ such that $P(z) = u_D$
- That is why it is called "marginal"
  Marginal people who have just the threshold propensity score of 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 V
- 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 $P\ z\quad u_D$
- That is why it is called "marginal"
  Marginal people who have just the threshold propensity score of 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

- MTE is a mean treatment e ect for a very speci c group of people
- People with observed characteristics X and unobserved taste on treatment M
- People with observed characteristics X who would be indi erent between treatment or not if they were randomly assigned a value Z of z such that $P \; z \quad u_D$
- That is why it is called "marginal"
  Marginal people who have just the threshold propensity score u_D of
- Di erent from LATE, it is not de ned 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 V
- 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 propensity score of 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

- MTE is a mean treatment e ect for a very speci c group of people
- People with observed characteristics X and unobserved taste on treatment V
- People with observed characteristics X who would be indi erent 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 propensity score u_D of
- Di erent from LATE, it is not de ned 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: 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:

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

- 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 z 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:

$$
\begin{aligned}
\text{LATE} \quad & E \; Y_1 \quad Y_0 \P X \quad x; D \; z \quad 1; D \; z^{\neg} \quad 0 \\
& E \; Y_1 \quad Y_0 \P X \quad x; u_D^{\neg} \; \$ \; U_D \; \& \; u_D \\
& E \int_{u_D^{\neg}}^{u_D} \text{MTE} \quad x; u \; du
\end{aligned}
$$

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

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:

$$
\begin{aligned}
LATE &= E[Y_1 - Y_0 \mid X = x, D(z) = 1, D(z') = 0] \\
&= E[Y_1 - Y_0 \mid X = x, u_D' < U_D \leq u_D] \\
&= E\left[\int_{u_D'}^{u_D} \Delta^{MTE}(x, u)\, du\right]
\end{aligned}
$$

- 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 z 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:

$$
\begin{aligned}
LATE &= E\left[Y_1 - Y_0 \mid X = x; D(z) = 1; D(z') = 0\right] \\
&= E\left[Y_1 - Y_0 \mid X = x; u_D' < U_D \le u_D\right] \\
&= E\left[\int_{u_D'}^{u_D} MTE(x; u)\, du\right]
\end{aligned}
$$

- 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 z 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:

$$
\begin{aligned}
LATE &= E\left[Y_1 - Y_0 \mid X = x; D(z) = 1; D(z') = 0\right] \\
&= E\left[Y_1 - Y_0 \mid X = x; u'_D \le U_D < u_D\right] \\
&= E\left[\int_{u'_D}^{u_D} \Delta^{MTE}(x; u)\, du\right]
\end{aligned}
$$

- 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 z but above z'

MTE: MTE as a Framework

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

$$TE_j = \int_0^1 \Delta^{MTE}(x, u_D)\,\omega_j(x, u_D)\,du_D$$

- ω_j is the weight for j

MTE: MTE as a Framework

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

$$TE_j \quad E_0^1 \quad {}^{MTE} \; x; u_D \;\; \omega_j \;\; x; u_D \;\; du_D$$

- ω_j is the weight for j

MTE: MTE as a Framework

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

$$TE_j = E\left[\int_0^1 \Delta^{MTE}(x;u_D)\,\omega_j(x;u_D)\,du_D\right]$$

- ω_j is the weight for j

MTE: De ning MTE

- Selection on MTE in a positive sorting Roy model
- MTE and TT/TUT weights

MTE: De ning MTE

- Selection on MTE in a positive sorting Roy model
- MTE and TT/TUT weights

MTE: Estimate MTE Using LIV

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

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

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

MTE: Estimate MTE Using LIV

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

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- A policy shock changes Z $\to$ changes propensity score $P(Z)$ $\to$ 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 x
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expection of Y w.r.t $P(Z) = p$:

$$\Delta^{LIV}(p) = \frac{\partial E(Y \mid P(Z) = p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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 x
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expection of Y w.r.t $P(Z) = p$:

$$\Delta^{LIV}(p) = \frac{\partial E(Y \mid P(Z) = p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- A policy shock changes Z $\rightarrow$ changes propensity score $P(Z)$ $\rightarrow$ 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 x
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expection of Y w.r.t $P(Z) = p$:

$$\Delta^{LIV}(p) = \frac{\partial E(Y \mid P(Z) = p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- A policy shock changes Z $\rightarrow$ changes propensity score $P(Z)$ $\rightarrow$ 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 x
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expection of Y w.r.t $P(Z) = p$:

$$\Delta^{LIV}(p) = \frac{\partial E(Y \mid P(Z) = p)}{\partial p}$$

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- A policy shock changes Z $\rightarrow$ changes propensity score $P(Z)$ $\rightarrow$ 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 x
- How to identify it? Local instrumental variable (LIV)
- LIV is the derivative of the conditional expection of Y w.r.t $P(Z) = p$:

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

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

MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

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

- 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 $P(Z) = p$ driven by an exogenous variation on instrument Z

MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\text{MTE}(p) = \text{LIV}(p) = \frac{\partial E[Y \mid P(Z) = p]}{\partial p}$$

- 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 $P(Z) = p$ 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 $P(Z) = p$ driven by an exogenous variation on instrument Z

MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\Delta^{MTE}(p) = \Delta^{LIV}(p) = \frac{\partial E[Y|P(Z) = p]}{\partial p}$$

- 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 $P(Z) = p$ driven by an exogenous variation on instrument Z

MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\Delta^{MTE}(p) = \Delta^{LIV}(p) = \frac{\partial E[Y|P(Z) = p]}{\partial p}$$

- 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 $P(Z) = p$ driven by an exogenous variation on instrument Z

MTE: Estimate MTE Using LIV

- Under A1-A5, we can show that

$$\Delta^{MTE}(p) = \Delta^{LIV}(p) = \frac{\partial E[Y|P(Z)=p]}{\partial p}$$

- 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 $P(Z) = p$ 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), nd 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 coe cient 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), nd 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 coe cient 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), find 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), nd 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 coe cient 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), find 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), nd 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 coe cient 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 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 e ect
- 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 e ect
- 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 e ect
- 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 X and some unobserved treatment taste V (or treatment threshold p)
- 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 X and some unobserved treatment taste V (or treatment threshold p)
- 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 e ect for people with speci c characteristics X and some unobserved treatment taste V (or treatment threshold p)
- It is externally valid and not IV-speci c
- 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 X and some unobserved treatment taste V (or treatment threshold p)
- 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 X and some unobserved treatment taste V (or treatment threshold p)
- 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 e ect for people with speci c characteristics X and some unobserved treatment taste V (or treatment threshold p)
- It is externally valid and not IV-speci c
- 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 X and some unobserved treatment taste V (or treatment threshold p)
- 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 X and some unobserved treatment taste V (or treatment threshold p)
- 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-bene t 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-bene t 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-bene t 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-bene t 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-bene t 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 nd large impact from observational studies, but small e ect from RCT. Does it mean that this HS is ine ective?
- 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 peopl 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 has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t analysis
- Taking into consideration the gov's monetary savings when people transfer from other programs to HS

Application: Kline and Walters (2016)

- The treatment has three values: no program, other program, HS program
- Kline and Walters (2016) rst categorize people to all behavior types and use ARP to eliminate some of them
- Then they varify various causal parameters needed for di erent 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-bene t 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 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 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

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

Conclusion

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

Conclusion

- First, we use ARP and other reasonable economic assumptions to simplify the identi cation in complicated multivalued IV cases
- Second, we introduce MTE framework to deal with external validity issues
- MTE is the treatment e ect of a small group of people with speci c value of characteristics X and treatment taste V (or treatment threshold U_D)
- It can be identi ed 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

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

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.