Frontier Topics in Empirical Economics: Week 6
IV beyond LATE

Zibin Huang [1]

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

October 26, 2025

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 for everyone
- If the agent chooses something when the price is higher ($D(z=0)=1$)
- Then he/she will definitely choose it when the price is lower ($D(z=1)=1$)

IV beyond LATE: Choice Model and IV

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

IV beyond LATE: Choice Model and IV

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

IV beyond LATE: Choice Model and IV

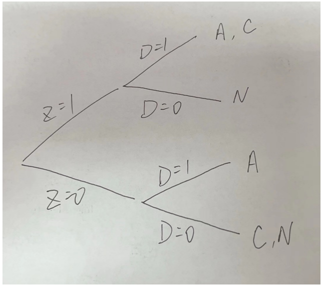
- You have already used it in the LATE Theorem: Monotonicity
- The idea of monotonicity comes from assuming treatment is a normal good for everyone
- If the agent chooses something when the price is higher ($D(z = 0) = 1$)
- Then he/she will definitely choose it when the price is lower ($D(z = 1) = 1$)

IV beyond LATE: Choice Model and IV

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

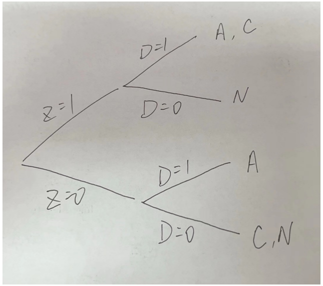
Generalization of LATE: Multiple IV

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



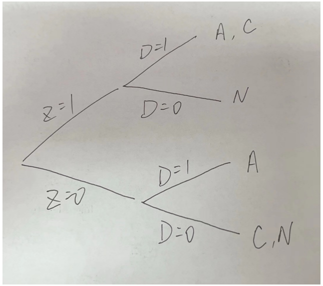
Generalization of LATE: Multiple IV

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



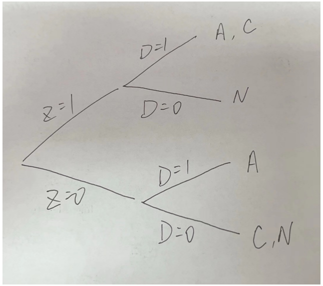
Generalization of LATE: Multiple IV

- In LATE theorem, we assume that both IV and treatment are single and binary
- Then it gives you $2 \times 2 = 4$ 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 $2 \times 2 = 4$ 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)

$$
\begin{aligned}
&E(Y|Z=1,D=1) = P(A|Z=1,D=1)E(Y_1|Z=1,D=1,A) + P(C|Z=1,D=1)E(Y_1|Z=1,D=1,C) \\
&E(Y|Z=1,D=0) = E(Y_0|Z=1,D=0,N) \\
&E(Y|Z=0,D=1) = E(Y_1|Z=0,D=1,A) \\
&E(Y|Z=0,D=0) = P(N|Z=0,D=0)E(Y_0|Z=0,D=0,N) + P(C|Z=0,D=0)E(Y_0|Z=0,D=0,C)
\end{aligned}
$$

- 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 (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 (final nodes, three types)

$$
\begin{aligned}
E(Y|Z=1,D=1) &= P(A|Z=1,D=1)E(Y_1|Z=1,D=1,A) + P(C|Z=1,D=1)E(Y_1|Z=1,D=1,C)\\
E(Y|Z=1,D=0) &= E(Y_0|Z=1,D=0,N)\\
E(Y|Z=0,D=1) &= E(Y_1|Z=0,D=1,A)\\
E(Y|Z=0,D=0) &= P(N|Z=0,D=0)E(Y_0|Z=0,D=0,N) + P(C|Z=0,D=0)E(Y_0|Z=0,D=0,C)
\end{aligned}
$$

- 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 (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 (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 (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 (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 (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

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

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

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

Generalization of LATE: Multiple IV

- First, consider we have multiple binary IV and binary treatment
- 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

- $LATE_1, LATE_2$ are LATEs for instrument z_1 and z_2
- $\psi \in (0, 1)$, determined by first stage relation
- Larger weight is given to IV with larger first 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:

$$\rho_{2SLS} = \psi LATE_1 + (1 - \psi) LATE_2$$

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

Generalization of LATE: Multivalued Treatment

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- Second, 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

- 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:

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

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

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:

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

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

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 $s = 2, 3, \ldots, \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, \ldots, \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
    - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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 $s = 2, 3, ..., \bar{s}$ when $z = 1$
  - who choose $s = 1$ when $z = 0$, but choose $s = 2, 3, ..., \bar{s}$ 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...
- Attention, here we still do not consider multivalued IV

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...
- Attention, here we still do not consider multivalued IV

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...
- Attention, here we still do not consider multivalued IV

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...
- Attention, here we still do not consider multivalued IV

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...
- Attention, here we still do not consider multivalued IV

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...
- Attention, here we still do not consider multivalued IV

Generalization of LATE: Multivalued IV

- We have already checked multiple IV and multivalued treatment cases
- We now consider multivalued IV
- This can be a little bit tricky
- There is a simple first thought of this issue: transit multivalued IV to multiple 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 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 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 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 $z = 0, 1, 2$, 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 $z = 0, 1, 2$, 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 $z = 0, 1, 2$, 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 $z = 0, 1, 2$, 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 $z = 0, 1, 2$, 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 $z_1 = 0$
- 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 $z_1 = 0$
- 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 $z_1 = 0$
- 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 $z_1 = 0$
- 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 $z_1 = 0$
- 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

Homework:

- Actually, in three-value IV, we can decompose it to dummies in another way to avoid the issues above
- Think about this, how to design this dummy decomposition?
- Can this be extrapolated to high levels? (number of values more than three)

Generalization of LATE: Multivalued IV

Homework:

- Actually, in three-value IV, we can decompose it to dummies in another way to avoid the issues above
- Think about this, how to design this dummy decomposition?
- Can this be extrapolated to high levels? (number of values more than three)

Generalization of LATE: Multivalued IV

Homework:

- Actually, in three-value IV, we can decompose it to dummies in another way to avoid the issues above
- Think about this, how to design this dummy decomposition?
- Can this be extrapolated to high levels? (number of values more than three)

Generalization of LATE: Multivalued IV

Homework:

- Actually, in three-value IV, we can decompose it to dummies in another way to avoid the issues above
- Think about this, how to design this dummy decomposition?
- Can this be extrapolated to high levels? (number of values more than three)

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

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 ($x_A \succsim_R x_B$)
- 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 ($x_A \succsim_R x_B$)
- 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 ($x_A \succsim_R x_B$)
- 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 ($x_A \succsim_R x_B$)
- 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 $(p^1, w^1), ..., (p^N, w^N)$ 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 $(p^1, w^1), ...(p^N, w^N)$ 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 $(p^1, w^1), ...(p^N, w^N)$ 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 $(p^1, w^1), ...(p^N, w^N)$ 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 $(p^1, w^1), \ldots (p^N, w^N)$ 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 $(p^1, w^1), ...(p^N, w^N)$ 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

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

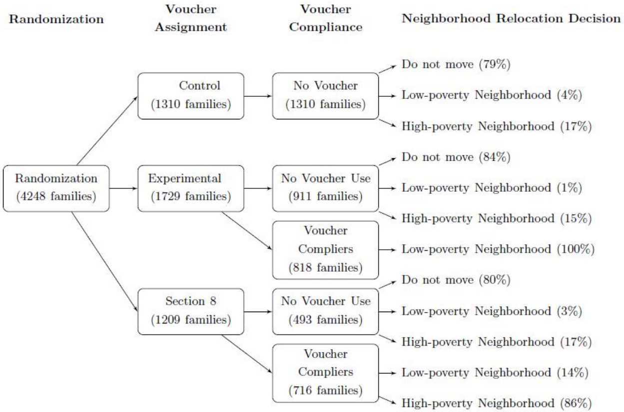
- Moving to Opportunity (MTO) is a housing experiment to encourage low-income families to move to neighborhood with low poverty rate
- There are three policy groups (three values of IV)
  - Control group: No vouchers (z_1)
  - Experimental group: Vouchers, available only for housing lease in low poverty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

Generalization of LATE: Multivalued IV

- 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

Figure 1: **Neighborhood Relocation by Voucher Assignment and Compliance**



Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- How to eliminate types as we do in monotonicity? ARP

Generalization of LATE: Multivalued IV

- 3 moving decisions under 3 possible vouchers
- Thus, we have 27 types of agents in total: $3^3 = 27$
- Only 9 available equations for observed expectations: $3 \times 3 = 9$
- It is impossible to invert a linear system of 9 equations to identify any causal effect with 27 behavior types
- Generally, if you have n decisions with m IV values:
  Number of agent types is n^m, number of equations is $n \times m$
- Behavior types of agent grow exponentially when you have more IV/decision values
- 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 $C(z)$ 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 $C(z)$ 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 $C(z)$ 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 $C(z)$ given different vouchers

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

$$W_\omega(z_1, 2) \subsetneq W_\omega(z_2, 2) = W_\omega(z_3, 2) \quad (1)$$

$$W_\omega(z_1, 3) = W_\omega(z_2, 3) \subsetneq W_\omega(z_3, 3) \quad (2)$$

$$W_\omega(z_1, 1) = W_\omega(z_1, 2) = W_\omega(z_1, 3) = W_\omega(z_2, 1) = W_\omega(z_2, 3) = W_\omega(z_3, 1) \quad (3)$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group ($z = z_1$) 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:

$$W_\omega(z_1, 2) \subsetneq W_\omega(z_2, 2) = W_\omega(z_3, 2) \tag{1}$$

$$W_\omega(z_1, 3) = W_\omega(z_2, 3) \subsetneq W_\omega(z_3, 3) \tag{2}$$

$$W_\omega(z_1, 1) = W_\omega(z_1, 2) = W_\omega(z_1, 3) = W_\omega(z_2, 1) = W_\omega(z_2, 3) = W_\omega(z_3, 1) \tag{3}$$

- 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:

$$W_\omega(z_1, 2) \subsetneq W_\omega(z_2, 2) = W_\omega(z_3, 2) \quad (1)$$

$$W_\omega(z_1, 3) = W_\omega(z_2, 3) \subsetneq W_\omega(z_3, 3) \quad (2)$$

$$W_\omega(z_1, 1) = W_\omega(z_1, 2) = W_\omega(z_1, 3) = W_\omega(z_2, 1) = W_\omega(z_2, 3) = W_\omega(z_3, 1) \quad (3)$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group ($z = z_1$) 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:

$$W_\omega(z_1, 2) \subsetneq W_\omega(z_2, 2) = W_\omega(z_3, 2) \quad (1)$$

$$W_\omega(z_1, 3) = W_\omega(z_2, 3) \subsetneq W_\omega(z_3, 3) \quad (2)$$

$$W_\omega(z_1, 1) = W_\omega(z_1, 2) = W_\omega(z_1, 3) = W_\omega(z_2, 1) = W_\omega(z_2, 3) = W_\omega(z_3, 1) \quad (3)$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group ($z = z_1$) 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:

$$W_\omega(z_1, 2) \subsetneq W_\omega(z_2, 2) = W_\omega(z_3, 2) \quad (1)$$

$$W_\omega(z_1, 3) = W_\omega(z_2, 3) \subsetneq W_\omega(z_3, 3) \quad (2)$$

$$W_\omega(z_1, 1) = W_\omega(z_1, 2) = W_\omega(z_1, 3) = W_\omega(z_2, 1) = W_\omega(z_2, 3) = W_\omega(z_3, 1) \quad (3)$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

- Now we translate three subsidizing rules to budget set:
  - Control group ($z = z_1$) 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:

$$W_\omega(z_1,2) \subsetneq W_\omega(z_2,2) = W_\omega(z_3,2) \tag{1}$$

$$W_\omega(z_1,3) = W_\omega(z_2,3) \subsetneq W_\omega(z_3,3) \tag{2}$$

$$W_\omega(z_1,1) = W_\omega(z_1,2) = W_\omega(z_1,3) = W_\omega(z_2,1) = W_\omega(z_2,3) = W_\omega(z_3,1) \tag{3}$$

- 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:

$$W_\omega(z_1, 2) \subsetneq W_\omega(z_2, 2) = W_\omega(z_3, 2) \quad (1)$$

$$W_\omega(z_1, 3) = W_\omega(z_2, 3) \subsetneq W_\omega(z_3, 3) \quad (2)$$

$$W_\omega(z_1, 1) = W_\omega(z_1, 2) = W_\omega(z_1, 3) = W_\omega(z_2, 1) = W_\omega(z_2, 3) = W_\omega(z_3, 1) \quad (3)$$

- What are the meanings of these three relations?

Generalization of LATE: Multivalued IV

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

- 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:

$$
\begin{aligned}
&1. C_\omega(z_1) = 2 \Rightarrow C_\omega(z_2) = 2, C_\omega(z_3) \neq 1 \\
&2. C_\omega(z_1) = 3 \Rightarrow C_\omega(z_2) \neq 1, C_\omega(z_3) \neq 1 \\
&3. C_\omega(z_2) = 1 \Rightarrow C_\omega(z_1) = 1, C_\omega(z_3) \neq 2 \\
&4. C_\omega(z_2) = 3 \Rightarrow C_\omega(z_1) = 3, C_\omega(z_3) = 3 \\
&5. C_\omega(z_3) = 1 \Rightarrow C_\omega(z_1) = 1, C_\omega(z_2) = 1 \\
&6. C_\omega(z_3) = 2 \Rightarrow C_\omega(z_2) = 2
\end{aligned}
$$

- Test yourself, explain all these six inequalities

Generalization of LATE: Multivalued IV

- Then we derive the following choice rule

Lemma L-1 Pinto (2015)

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

$$
\begin{aligned}
&1. C_\omega(z_1) = 2 \Rightarrow C_\omega(z_2) = 2, C_\omega(z_3) \neq 1\\
&2. C_\omega(z_1) = 3 \Rightarrow C_\omega(z_2) \neq 1, C_\omega(z_3) \neq 1\\
&3. C_\omega(z_2) = 1 \Rightarrow C_\omega(z_1) = 1, C_\omega(z_3) \neq 2\\
&4. C_\omega(z_2) = 3 \Rightarrow C_\omega(z_1) = 3, C_\omega(z_3) = 3\\
&5. C_\omega(z_3) = 1 \Rightarrow C_\omega(z_1) = 1, C_\omega(z_2) = 1\\
&6. C_\omega(z_3) = 2 \Rightarrow C_\omega(z_2) = 2
\end{aligned}
$$

- Test yourself, explain all these six inequalities

Generalization of LATE: Multivalued IV

- Then we derive the following choice rule

Lemma L-1 Pinto (2015)

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

$$
\begin{aligned}
&1. C_\omega(z_1) = 2 \Rightarrow C_\omega(z_2) = 2, C_\omega(z_3) \neq 1 \\
&2. C_\omega(z_1) = 3 \Rightarrow C_\omega(z_2) \neq 1, C_\omega(z_3) \neq 1 \\
&3. C_\omega(z_2) = 1 \Rightarrow C_\omega(z_1) = 1, C_\omega(z_3) \neq 2 \\
&4. C_\omega(z_2) = 3 \Rightarrow C_\omega(z_1) = 3, C_\omega(z_3) = 3 \\
&5. C_\omega(z_3) = 1 \Rightarrow C_\omega(z_1) = 1, C_\omega(z_2) = 1 \\
&6. C_\omega(z_3) = 2 \Rightarrow C_\omega(z_2) = 2
\end{aligned}
$$

- 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 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 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 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 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

- Let's take 1 as an example:
  $C_\omega(z_1) = 2$ means that under z_1, we choose $t = 2$ over $t = 1, 3$ when no action is financed.
  Then under z_2, moving to rich $t = 2$ is financed, 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 financed, extending the budget (but may be in different 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:
  $C_\omega(z_1) = 2$ means that under z_1, we choose $t = 2$ over $t = 1, 3$ when no action is financed.
  Then under z_2, moving to rich $t = 2$ is financed, 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 financed, extending the budget (but may be in different 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

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

- 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 identification
- Otherwise, you really do not know what results your IV is giving you
- Now you see the power of economic theory to guide your identification
- When statistics tools are exhausted, remember you are an economist
- Do not think first year Micro and Macro are useless!!!

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

Generalization of LATE: Multivalued IV

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

MTE: Choice Model

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

MTE: Choice Model

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

MTE: Choice Model

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

MTE: Choice Model

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

MTE: Choice Model

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

MTE: Choice Model

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

MTE: Choice Model

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

$$Y_1 = \mu_1(X, U_1) \quad (4)$$
$$Y_0 = \mu_0(X, U_0) \quad (5)$$

- 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

$$Y_1 = \mu_1(X, U_1) \tag{4}$$
$$Y_0 = \mu_0(X, U_0) \tag{5}$$

- 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

$$Y_1 = \mu_1(X, U_1) \quad (4)$$
$$Y_0 = \mu_0(X, U_0) \quad (5)$$

- 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

$$Y_1 = \mu_1(X, U_1) \tag{4}$$
$$Y_0 = \mu_0(X, U_0) \tag{5}$$

- 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

$$Y_1 = \mu_1(X, U_1) \tag{4}$$
$$Y_0 = \mu_0(X, U_0) \tag{5}$$

- 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

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

- Z is an instrument that can change individual's choices, V is an unobserved factor
- For instance, Y is wage, D is college enrollment, Z is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe (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

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

- Z is an instrument that can change individual's choices, V is an unobserved factor
- For instance, Y is wage, D is college enrollment, Z is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe (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

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

- Z is an instrument that can change individual's choices, V is an unobserved factor
- For instance, Y is wage, D is college enrollment, Z is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe (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

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

- Z is an instrument that can change individual's choices, V is an unobserved factor
- For instance, Y is wage, D is college enrollment, Z is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe (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

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

- Z is an instrument that can change individual's choices, V is an unobserved factor
- For instance, Y is wage, D is college enrollment, Z is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe (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

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

- Z is an instrument that can change individual's choices, V is an unobserved factor
- For instance, Y is wage, D is college enrollment, Z is a policy to subsidize students from poor regions
- Agents observe everything. Econometricians observe (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) $E(|Y_1|), E(|Y_0|)$ are finite
- (A-5) $0 < Pr(D = 1|X) < 1$
  Possible to have $D = 1$ or $D = 0$ at any point of X

MTE: Choice Model

- We invoke five assumptions for this model
- (A-1) (U_0, U_1, V) are independent of Z conditional on X
  Independence 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 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 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 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 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 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 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 (Example - Roy Model)

- A simple example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payoff, an observed relative working cost $C = Z_1$ in modern sector
- Agents choose a sector with higher utility (payoff abstract from cost)
- Positive sorting: People with higher return sort to modern sector (treatment)

MTE: Choice Model (Example - Roy Model)

- A simple example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payoff, an observed relative working cost $C = Z_1$ in modern sector
- Agents choose a sector with higher utility (payoff abstract from cost)
- Positive sorting: People with higher return sort to modern sector (treatment)

MTE: Choice Model (Example - Roy Model)

- A simple example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payoff, an observed relative working cost $C = Z_1$ in modern sector
- Agents choose a sector with higher utility (payoff abstract from cost)
- Positive sorting: People with higher return sort to modern sector (treatment)

MTE: Choice Model (Example - Roy Model)

- A simple example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payoff, an observed relative working cost $C = Z_1$ in modern sector
- Agents choose a sector with higher utility (payoff abstract from cost)
- Positive sorting: People with higher return sort to modern sector (treatment)

MTE: Choice Model (Example - Roy Model)

- A simple example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payoff, an observed relative working cost $C = Z_1$ in modern sector
- Agents choose a sector with higher utility (payoff abstract from cost)
- Positive sorting: People with higher return sort to modern sector (treatment)

MTE: Choice Model (Example - Roy Model)

- A simple example of this model setting is the Roy Model (sorting model)
- We have two sectors Agriculture=0 and Modern=1
- Y is working payoff, an observed relative working cost $C = Z_1$ in modern sector
- Agents choose a sector with higher utility (payoff abstract from cost)
- Positive sorting: People with higher return sort to modern sector (treatment)

MTE: Choice Model (Example - Roy Model)

- Assume μ additively separable in U

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0$$
$$D^* = \mu_1(X) + U_1 - [\mu_0(X) + U_0] - Z_1 - V_C; \quad D = 1 \text{ if } D^* \geq 0; \; D = 0 \text{ otherwise}$$

- We can transform this equation to have:

$$D^* = \underbrace{\mu_1(X) - \mu_0(X) - Z_1}_{\mu_D} + \underbrace{[U_1 - U_0 - V_C]}_{-V}$$

- In this case, we have $V = -[U_1 - U_0 - V_C]$
- Positive sorting: People with high $\mu_1 - \mu_0$ or large $U_1 - U_0 \Longrightarrow$ Go to sector 1

MTE: Choice Model (Example - Roy Model)

- Assume μ additively separable in U

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0$$
$$D^* = \mu_1(X) + U_1 - [\mu_0(X) + U_0] - Z_1 - V_C, \quad D = 1 \text{ if } D^* \geq 0;\ D = 0 \text{ otherwise}$$

- We can transform this equation to have:

$$D^* = \underbrace{\mu_1(X) - \mu_0(X) - Z_1}_{\mu_D} + \underbrace{[U_1 - U_0 - V_C]}_{-V}$$

- In this case, we have $V = -[U_1 - U_0 - V_C]$
- Positive sorting: People with high $\mu_1 - \mu_0$ or large $U_1 - U_0 \Longrightarrow$ Go to sector 1

MTE: Choice Model (Example - Roy Model)

- Assume μ additively separable in U

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0$$
$$D^* = \mu_1(X) + U_1 - [\mu_0(X) + U_0] - Z_1 - V_C, \quad D = 1 \text{ if } D^* \geq 0; \; D = 0 \text{ otherwise}$$

- We can transform this equation to have:

$$D^* = \underbrace{\mu_1(X) - \mu_0(X) - Z_1}_{\mu_D} + \underbrace{[U_1 - U_0 - V_C]}_{-V}$$

- In this case, we have $V = -[U_1 - U_0 - V_C]$
- Positive sorting: People with high $\mu_1 - \mu_0$ or large $U_1 - U_0 \Longrightarrow$ Go to sector 1

MTE: Choice Model (Example - Roy Model)

- Assume μ additively separable in U

$$Y_1 = \mu_1(X) + U_1$$
$$Y_0 = \mu_0(X) + U_0$$
$$D^* = \mu_1(X) + U_1 - [\mu_0(X) + U_0] - Z_1 - V_C, \quad D = 1 \text{ if } D^* \geq 0;\ D = 0 \text{ otherwise}$$

- We can transform this equation to have:

$$D^* = \underbrace{\mu_1(X) - \mu_0(X) - Z_1}_{\mu_D} + \underbrace{[U_1 - U_0 - V_C]}_{-V}$$

- In this case, we have $V = -[U_1 - U_0 - V_C]$
- Positive sorting: People with high $\mu_1 - \mu_0$ or large $U_1 - U_0 \Longrightarrow$ Go to sector 1

MTE: Choice Model (Example - Roy Model)

- Assume μ additively separable in U

$$
\begin{aligned}
Y_1 &= \mu_1(X) + U_1 \\
Y_0 &= \mu_0(X) + U_0 \\
D^* &= \mu_1(X) + U_1 - [\mu_0(X) + U_0] - Z_1 - V_C, \quad D = 1 \text{ if } D^* \geq 0;\ D = 0 \text{ otherwise}
\end{aligned}
$$

- We can transform this equation to have:

$$D^* = \underbrace{\mu_1(X) - \mu_0(X) - Z_1}_{\mu_D} + \underbrace{[U_1 - U_0 - V_C]}_{-V}$$

- In this case, we have $V = -[U_1 - U_0 - V_C]$
- Positive sorting: People with high $\mu_1 - \mu_0$ or large $U_1 - U_0 \Rightarrow$ Go to sector 1

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $P(Z) \equiv Pr(D = 1|Z, X)$
- This is the propensity score (PS) to get treated for agent with Z
- In this model, it also means $P(\mu_D(Z) > V|Z, X)$
- Let $F_{V|X}(\cdot)$ denote the distribution of V conditional on X
- This PS can be further expressed as: $P(Z) \equiv Pr(D = 1|Z, X) = F_{V|X}(\mu_D(Z))$
- $F_{V|X}(\mu_D(Z))$ means the probability of V (given X) that is below $\mu_D(Z)$
- This is also just the choice probability (propensity score)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \iff \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- Thus, for a choice function, an agent can be characterized by (X, V) or (X, U_D)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \quad \Longleftrightarrow \quad \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- Thus, for a choice function, an agent can be characterized by (X, V) or (X, U_D)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \iff \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- Thus, for a choice function, an agent can be characterized by (X, V) or (X, U_D)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \iff \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- Thus, for a choice function, an agent can be characterized by (X, V) or (X, U_D)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \iff \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- Thus, for a choice function, an agent can be characterized by (X, V) or (X, U_D)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \iff \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- Thus, for a choice function, an agent can be characterized by (X, V) or (X, U_D)

MTE: Choice Model

- Let $U_D = F_{V|X}(V)$, we have $U_D \sim Unif[0,1]$
- $U_D = F_{V|X}(V)$ 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 $F_{V|X}(\mu_D(Z)) > F_{V|X}(V) = U_D$ (larger than this threshold) to get treated
- We have a clear one-to-one mapping between V and U_D
- And a clear equivalence of conditions to get treated

$$\mu_D(Z) > V \iff \underbrace{F_{V|X}(\mu_D(Z))}_{\text{PS mapped from } Z} > \underbrace{F_{V|X}(V)}_{\text{PS mapped from } V} = U_D$$

- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- 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)
- Specifically, why does this model also imply monotonicity?
- The intuition is simple: V could not affect $\mu_D(Z)$
- $D^* = \mu_D(Z) - V \Rightarrow$ additively separable for Z and V
- Thus, given z and z', $\forall V \Rightarrow D^*(z) \geq D^*(z')$ or $D^*(z) \leq D^*(z')$
- This model explicitly describes the decision-making process in a structural way, which allows us to investigate more causal questions

MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let $\Delta = Y_1 - Y_0$
- Conditional ATE is defined as usual: $\Delta^{ATE}(x) \equiv E(\Delta | X = x)$
- MTE is defined as the mean effect of treatment on those for whom $X = x$ and $U_D = u_D (V = v)$

Definition of the MTE

The Marginal Treatment Effect is defined as:

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta | X = x, U_D = u_D)$$

MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let $\Delta = Y_1 - Y_0$
- Conditional ATE is defined as usual: $\Delta^{ATE}(x) \equiv E(\Delta | X = x)$
- MTE is defined as the mean effect of treatment on those for whom $X = x$ and $U_D = u_D(V = v)$

Definition of the MTE

The Marginal Treatment Effect is defined as:

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta | X = x, U_D = u_D)$$

MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let $\Delta = Y_1 - Y_0$
- Conditional ATE is defined as usual: $\Delta^{ATE}(x) \equiv E(\Delta|X = x)$
- MTE is defined as the mean effect of treatment on those for whom $X = x$ and $U_D = u_D (V = v)$

Definition of the MTE

The Marginal Treatment Effect is defined as:

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta|X = x, U_D = u_D)$$

MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let $\Delta = Y_1 - Y_0$
- Conditional ATE is defined as usual: $\Delta^{ATE}(x) \equiv E(\Delta | X = x)$
- MTE is defined as the mean effect of treatment on those for whom $X = x$ and $U_D = u_D (V = v)$

Definition of the MTE

The Marginal Treatment Effect is defined as:

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta | X = x, U_D = u_D)$$

MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let $\Delta = Y_1 - Y_0$
- Conditional ATE is defined as usual: $\Delta^{ATE}(x) \equiv E(\Delta|X = x)$
- MTE is defined as the mean effect of treatment on those for whom $X = x$ and $U_D = u_D(V = v)$

Definition of the MTE

The Marginal Treatment Effect is defined as:

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta|X = x, U_D = u_D)$$

MTE: Defining MTE

- Now let's define ATE and MTE in this model
- Let $\Delta = Y_1 - Y_0$
- Conditional ATE is defined as usual: $\Delta^{ATE}(x) \equiv E(\Delta | X = x)$
- MTE is defined as the mean effect of treatment on those for whom $X = x$ and $U_D = u_D(V = v)$

Definition of the MTE

The Marginal Treatment Effect is defined as:

$$\Delta^{MTE}(x, u_D) \equiv E(\Delta | X = x, U_D = u_D)$$

MTE: Defining MTE

- MTE is a mean treatment effect for a very specific group of people
- People with observed characteristics 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 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 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 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 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 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 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 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: 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:

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

- 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:

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

- 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:

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

- 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:

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

- 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:

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

- 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

- 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

$\text{ATE}(x) = E(Y_1 - Y_0 \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\, du_D$

$\text{TT}(x) = E(Y_1 - Y_0 \mid X = x, D = 1) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{TT}}(x, u_D)\, du_D$

$\text{TUT}(x) = E(Y_1 - Y_0 \mid X = x, D = 0) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{TUT}}(x, u_D)\, du_D$

Policy relevant treatment effect: $\text{PRTE}(x) = E(Y_{a'} \mid X = x) - E(Y_a \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{PRTE}}(x, u_D)\, du_D$ for two policies a and a' that affect the Z but not the X

$\text{IV}_J(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{IV}}^J(x, u_D)\, du_D$, given instrument J

$\text{OLS}(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{OLS}}(x, u_D)\, du_D$

Source: Heckman and Vytlacil (2005).

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

$\text{ATE}(x) = E(Y_1 - Y_0 \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\, du_D$
$\text{TT}(x) = E(Y_1 - Y_0 \mid X = x, D = 1) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{TT}}(x, u_D)\, du_D$
$\text{TUT}(x) = E(Y_1 - Y_0 \mid X = x, D = 0) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{TUT}}(x, u_D)\, du_D$
Policy relevant treatment effect: $\text{PRTE}(x) = E(Y_{a'} \mid X = x) - E(Y_a \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{PRTE}}(x, u_D)\, du_D$ for two policies a and a' that affect the Z but not the X
$\text{IV}_J(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{IV}}^J(x, u_D)\, du_D$, given instrument J
$\text{OLS}(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{OLS}}(x, u_D)\, du_D$

Source: Heckman and Vytlacil (2005).

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

$\text{ATE}(x) = E(Y_1 - Y_0 \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\, du_D$

$\text{TT}(x) = E(Y_1 - Y_0 \mid X = x, D = 1) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{TT}}(x, u_D)\, du_D$

$\text{TUT}(x) = E(Y_1 - Y_0 \mid X = x, D = 0) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{TUT}}(x, u_D)\, du_D$

Policy relevant treatment effect: $\text{PRTE}(x) = E(Y_{a'} \mid X = x) - E(Y_a \mid X = x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{PRTE}}(x, u_D)\, du_D$ for two policies a and a' that affect the Z but not the X

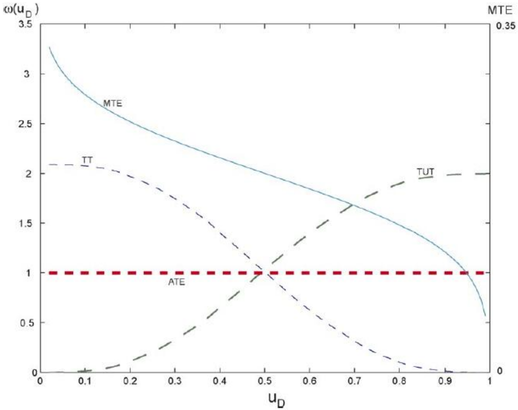
$\text{IV}_J(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{IV}}^J(x, u_D)\, du_D$, given instrument J

$\text{OLS}(x) = \int_0^1 \Delta^{\text{MTE}}(x, u_D)\omega_{\text{OLS}}(x, u_D)\, du_D$

Source: Heckman and Vytlacil (2005).

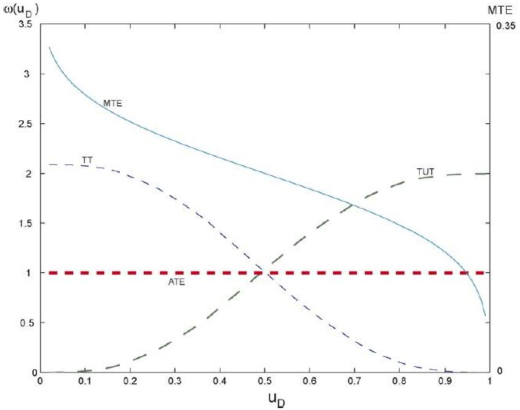
MTE: Defining MTE

- Selection on MTE in a generalized positive sorting Roy model
- Large u_D means smaller return (MTE), fewer treated, more untreated



MTE: Defining MTE

- Selection on MTE in a generalized positive sorting Roy model
- Large u_D means smaller return (MTE), fewer treated, more untreated



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 $P(Z)$
- Policy shock changes $Z \Longrightarrow$ changes propensity score $P(Z) \Longrightarrow$ 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 $P(Z) = p$:

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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 $P(Z) = p$:

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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 $P(Z) = p$:

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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 $P(Z) = p$:

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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 $P(Z) = p$:

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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 $P(Z) = p$:

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

- LIV is the mean response to a policy change embodied in changes in $P(Z)$
- 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

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

- 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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

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
- This marginal change in outcome (LIV) comes from only marginal people (MTE), others would not respond

MTE: Estimate MTE Using LIV

- Then the question becomes how to estimate LIV?
- There is a non-parametric way:
    - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(Z)$
    - (2) Estimate Y given X and $p(Z)$ using non/semi-parametric methods
    - (3) Estimate derivatives by small perturbation
      Or it would be the regression coefficient if assume a linear model for Y
- We can also 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?
- There is a non-parametric way:
  - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(z)$
  - (2) Estimate Y given X and $p(z)$ using non/semi-parametric methods
  - (3) Estimate derivatives by small perturbation
    Or it would be the regression coefficient if assume a linear model for Y
- We can also 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?
- There is a non-parametric way:
  - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(z)$
  - (2) Estimate Y given X and $p(z)$ using non/semi-parametric methods
  - (3) Estimate derivatives by small perturbation
    Or it would be the regression coefficient if assume a linear model for Y
- We can also 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?
- There is a non-parametric way:
  - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(z)$
  - (2) Estimate Y given X and $p(z)$ using non/semi-parametric methods
  - (3) Estimate derivatives by small perturbation
    Or it would be the regression coefficient if assume a linear model for Y
- We can also 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?
- There is a non-parametric way:
  - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(z)$
  - (2) Estimate Y given X and $p(z)$ using non/semi-parametric methods
  - (3) Estimate derivatives by small perturbation
    Or it would be the regression coefficient if assume a linear model for Y
- We can also 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?
- There is a non-parametric way:
  - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(z)$
  - (2) Estimate Y given X and $p(z)$ using non/semi-parametric methods
  - (3) Estimate derivatives by small perturbation
    Or it would be the regression coefficient if assume a linear model for Y
- We can also 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?
- There is a non-parametric way:
  - (1) Estimate a treatment choice function (Probit or logit) to have PS $p(z)$
  - (2) Estimate Y given X and $p(z)$ using non/semi-parametric methods
  - (3) Estimate derivatives by small perturbation
    Or it would be the regression coefficient if assume a linear model for Y
- We can also 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 effect
- There is a slides of user guidance by Martin Andresen

MTE: Estimate MTE Using LIV

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

MTE: Estimate MTE Using LIV

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

MTE: Estimate MTE Using LIV

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

MTE: Estimate MTE Using LIV

```
. mtefe_gendata, obs(10000) districts(10)
.
. mtefe lwage exp exp2 i.district (col=distCol)
Parametric normal MTE model                              Observations : 10000
Treatment model: Probit
Estimation method: Local IV
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
beta0						
exp	.0358398	.0064408	5.56	0.000	.0232145	.0484651
exp2	-.0008453	.0002019	-4.19	0.000	-.0012411	-.0004496
district						
2	.2352456	.0680412	3.46	0.001	.1018712	.36862
3	.6294914	.0701091	8.98	0.000	.4920634	.7669194
4	.0131179	.0597721	0.22	0.826	-.1040474	.1302832
5	.0338606	.0705835	0.48	0.631	-.1044974	.1722186
6	.1699366	.0605086	2.81	0.005	.0513275	.2885458
7	-.1899241	.060115	-3.16	0.002	-.3077617	-.0720865
8	-.1842254	.0676843	-2.72	0.007	-.3169003	-.0515504
9	-.7908301	.0578436	-13.67	0.000	-.9042153	-.677445
10	-.4432749	.0597237	-7.42	0.000	-.5603455	-.3262044
_cons	3.164706	.0650331	48.66	0.000	3.037228	3.292184
beta1-beta0						

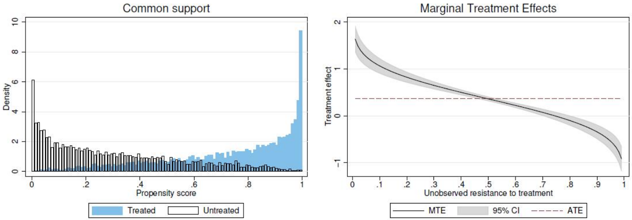
MTE: Estimate MTE Using LIV

exp	-.0386384	.010241	-3.77	0.000	-.0587128	-.018564
exp2	.0012967	.0003288	3.94	0.000	.0006523	.0019412
district						
2	.265112	.107039	2.48	0.013	.0552939	.4749301
(*output omitted*)						
10	.3143661	.1072555	2.93	0.003	.1041237	.5246085
_cons	.4255863	.0983572	4.33	0.000	.2327863	.6183863
k						
mills	-.4790282	.0611081	-7.84	0.000	-.5988124	-.359244
effects						
ate	.3283373	.0242932	13.52	0.000	.2807177	.3759568
att	.5369432	.0388809	13.81	0.000	.4607287	.6131576
atut	.1195067	.0384691	3.11	0.002	.0440995	.194914
late	.3279726	.0245142	13.38	0.000	.2799198	.3760254
mprte1	.3463148	.0256971	13.48	0.000	.2959433	.3966862
mprte2	.3309428	.024298	13.62	0.000	.2833137	.3785719
mprte3	-.016257	.0498984	-0.33	0.745	-.1140679	.0815538

Test of observable heterogeneity, p-value 0.0000
Test of essential heterogeneity, p-value 0.0000

Note: Analytical standard errors ignore the facts that the propensity score,
(*output omitted*)

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 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 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 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-benefit analysis for social programs when some close substitutes exist

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Application: Kline and Walters (2016)

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

Conclusion

- LATE is the most popular way to interpret IV estimate
- However, it has two important limitations
  - Usually not feasible when you have multivalued IV → too many types
  - Not 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 $\Longrightarrow$ 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 $\Longrightarrow$ 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 $\Longrightarrow$ 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 $\Longrightarrow$ 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 $\Longrightarrow$ 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 $\Longrightarrow$ 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

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

Conclusion

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

Conclusion

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

Conclusion

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

Conclusion

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

References

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.