Bounds on Net and Mechanism Average Treatment Effects / Bounds on LATE with an Invalid IV

Carlos A. Flores

California Polytechnic State University–San Luis Obispo

Short Course in Treatment Effect Bounds
Inter-American Development Bank

Bounds on (a) Mechanism Effects; and (b) with Invalid IVs

Material in this part is based on the papers Flores and Flores-Lagunes (2010, 2013)

Motivation

- The main focus when estimating causal effects is typically on the (total) average effect of a treatment on an outcome.
- Other effects of interest, which we analyze here, are the estimation of
  - A **causal mechanism (or indirect) effect** through which the treatment works.
  - A **causal net (or direct) effect**: the effect of the treatment net of that mechanism.
- These effects allow a better understanding of the treatment, and thus are important for policy.

Motivation: Example/Empirical Application

- We consider the average effect of a training program (Job Corps, JC) on earnings and employment
  - JC is the largest and most comprehensive job training program for economically disadvantaged youth (ages 16-24) in the US
  - It offers many services: academic and vocational training, health services, counseling, job search assistance, social skill training, etc.
  - We use data from a randomized evaluation of JC (in which positive effects of JC on earnings and employment were found)

- How does JC affect the outcomes? A possible mechanism for the effect: attainment of a high school, GED or vocational degree.
  - Ask: *What part of the effect of JC on employment and earnings is due to the attainment of such a degree (relative to other components)?*
    - Treatment: random assignment to JC
    - Mechanism: attainment of a HS,GED or vocational degree
    - Outcomes: employment and weekly earnings 12 quarters after RA

Motivation

- Questions of interest:

$$\begin{tikzcd} \textit{Treatment}\ (T) \arrow[r] \arrow[rr, bend left] & \textit{Mechanism}\ (S) \arrow[r] & \textit{Outcome}\ (Y) \end{tikzcd}$$

  - What part of the effect of T on Y works through S? (mechanism or indirect effect)
  - What part of the effect of T on Y is net of S? (net of direct effect)

- It is not easy to estimate these two effects
  - Need stronger assumptions than those needed for estimation of the effect of T on Y (Robins and Greenland, 1992; Rubin, 2004).

Motivation: Sketch of Problem

- Usual Case:

$$\begin{array}{ccc} \text{Treatment } (T) & \longrightarrow & \text{Outcome } (Y) \\ \nwarrow & \text{Confounders } (U) & \nearrow \end{array}$$

  - Need to address confounders U: OK if T is Randomly Assigned (RA)

- Current Case:

$$\begin{array}{ccccc} & \swarrow & \text{Confounders } U_1 & \searrow & \\ \textit{Treatment } (T) & \longrightarrow & \textit{Mechanism } (S) & \longrightarrow & \textit{Outcome } (Y) \\ & \nwarrow \ \text{Conf. } U_2 \ \nearrow & & \nwarrow \ \text{Conf. } U_3 \ \nearrow & \end{array}$$

  - Need to address confounders U_1: OK if T is RA
  - Need to address confounders U_2: OK if T is RA
  - Need to address confounders U_3: S is endogenous even if T is RA!
    - Thus, need further (and stronger) assumptions

Outline for the Rest of the Presentation

- Definition of Parameters: *NATE* and *MATE*
- Partial Identification (Bounds)
  - Preliminary Notes
  - Assumptions
- Extensions:
  - Bounds on the Effect of the Mechanism (S) on the outcome (Y) using the Treatment (T) as an Invalid Instrument (F and F-L, 2013)
- Empirical Application
- Stata Programs and Replication Results

Definition of Parameters

- Potential Outcomes Framework (Neyman, 1923; Rubin, 1974)
  - Treatment: $T_i \in \{0, 1\}$
  - Potential outcomes: $Y_i(0)$, $Y_i(1)$
  - Mechanism variable S with potential values: $S_i(0)$, $S_i(1)$. Need not be binary
  - Observe: (Y_i, T_i, S_i), where

$$\begin{aligned} Y_i &= T_i Y_i(1) + (1 - T_i) Y_i(0) \\ S_i &= T_i S_i(1) + (1 - T_i) S_i(0) \end{aligned}$$

  - The average treatment effect (ATE) is: $E[Y(1) - Y(0)]$.

Definition of Parameters

- Goal: Decompose *ATE* into a net and mechanism effect.

- Let $Y_i(t,s)$ be the "composite" potential outcome where t refers to the treatment, and s is one of the potential values of S: $s \in \{S(0), S(1)\}$.

  - By definition $Y_i(1) = Y_i(1, S_i(1))$ and $Y_i(0) = Y_i(0, S_i(0))$.

  - $Y_i(1, S_i(0))$: potential outcome when treatment is received but unit receives a value of the mechanism variable of $S_i(0)$.

  - Intuition: think of $Y_i(1, S_i(0))$ as the potential outcome of a *counterfactual* experiment in which the treatment is as the original one but we block the effect of T on S by holding S fixed at $S_i(0)$.

Definition of Parameters

- The net average treatment effect ($NATE$) is defined as

$$NATE = E[Y(1, S(0)) - Y(0, S(0))]$$

and the mechanism average treatment effect ($MATE$) is defined as

$$MATE = E[Y(1, S(1)) - Y(1, S(0))]$$

  - By construction, $ATE = MATE + NATE$.
  - $NATE$ gives the part of the effect of T on Y that is *not* due to a change in S caused by T (S is fixed at $S(0)$).
  - $MATE$ gives the part of the effect of T on Y that is due to a change in S caused by T (S changes from $S(0)$ to $S(1)$).
  - If *all* the effect of T on Y works through S then $NATE = 0$ and $MATE = ATE$. If *none* of the effect works through S (either because T does not affect S or S does not affect Y) then $NATE = ATE$ and $MATE = 0$

Bounds: Preliminary Notes

- Goal: Derive bounds on $NATE$ and $MATE$ under relatively weak assumptions
- We focus our discussion on $NATE$, since $MATE = ATE - NATE$
- We let S be binary, so $S_i\left(0\right) \in \{0,1\}$ and $S_i\left(1\right) \in \{0,1\}$
- **Key issue**: S is endogenous, and it is affected by treatment assignment (T), i.e., S is a "post-treatment variable"
  - As before, we use a Principal Stratification Approach to adjust for S

Partial Identification: Preliminary Notes

- In our setting with binary T and S, we have four principal strata (which are analogous to those in Angrist, Imbens, and Rubin, 1996):
  - $\{S_i(0), S_i(1)\} = \{0, 0\}$: the not-affected at 0 ($n0$)
  - $\{S_i(0), S_i(1)\} = \{1, 1\}$: the not-affected at 1 ($n1$)
  - $\{S_i(0), S_i(1)\} = \{0, 1\}$: the affected positively (ap)
  - $\{S_i(0), S_i(1)\} = \{1, 0\}$: the affected negatively (an)
- In our empirical application:
  - $n1$ ($n0$) strata: those who always (never) attain a degree whether they are assigned to JC or not
  - ap (na) strata: those who attain a degree only if assigned (not assigned) to JC

Partial Identification: Preliminary Notes

- Define the following "local" net and mechanism effects:

$$LNATE_k = E[Y(1, S(0))|k] - E[Y(0)|k], \text{ for } k = n0, n1, ap, an$$

and

$$LMATE_k = E[Y(1)|k] - E[Y(1, S(0))|k], \text{ for } k = n0, n1, ap, an$$

- Note:
  - Since $S_i(1) = S_i(0)$ for $n0$ and $n1$, $LMATE_{n0} = LMATE_{n1} = 0$.
  - We have information on $Y(1, S(0))$ for the $n0$ and $n1$ strata. For them, if $T_i = 1$ we have $Y_i = Y(1) = Y(1, S(1)) = Y(1, S(0))$

-
  - We never observe $Y(1, S(0))$ for compliers

Partial Identification: Preliminary Notes

- Our approach to bound $NATE$ and $MATE$:
  1. Get bounds for the expectations in the local effects
  2. Use those bounds to construct bounds on $NATE$ and $MATE$. For example:

$$NATE = \pi_{n1} LNATE_{n1} + \pi_{n0} LNATE_{n0} + \pi_{ap} LNATE_{ap}$$

Partial Identification: Basic Assumptions

Assumption A1. (*Randomly Assigned Treatment*).

$$\{Y(1), Y(0), Y(1, S(0)), S(1), S(0)\} \text{ is independent of } T$$

- It holds by design in our application
- It allows point identification of $E[Y(1)]$, $E[Y(0)]$, $E[S(1)]$ and $E[S(0)]$ (e.g., $E[Y(1)] = E[Y|T = 1]$)

Partial Identification: Basic Assumptions

Assumption A2. (*Individual-Level Monotonicity of* T on S*).*
$S_i(1) \geq S_i(0)$ *for all* i.

- Effect of T on S is non-negative for all units. It rules out the *an* stratum.
- In our application: No one would get a degree if not assigned to JC and would not get it if assigned to JC
- Similar assumptions are common in the literature (Imbens and Angrist, 1994; Angrist et al., 1996; Manski and Pepper; 2000; Shaikh and Vytlacil, 2005; Zhang et al., 2008; Lee, 2009; Sjölander, 2009)

Partial Identification: Basic Assumptions

- Under Assumption A2:

		T_i = 0	1
S_i	0	$n0, ap$	$n0$
	1	$n1$	$n1, ap$

- **Under Assump. A1 and A2** we can point identify several quantities:
  - $E[Y(0)|n1] = E[Y|T=0, S=1]$
    $E[Y(1)|n0] = E[Y|T=1, S=0]$
  - The proportion of each strata in the population, call them π_{n0}, π_{n1}, π_{ap}, and π_{an}. For example, $\pi_{n0} = \Pr(S=0|T=1)$.
- And, we can bound $E[Y(0)|n0]$, $E[Y(1)|n1]$, $E[Y(0)|ap]$ and $E[Y(1)|ap]$ using the trimming procedure in, e.g., Lee (2009).
- Hence, we can also bound $LNATE_{n0}$ and $LNATE_{n1}$.
- However, cannot bound $NATE$ and $MATE$ without more assumptions ($Y(1, S(0))$ never observed for compliers).

Bounds on NATE and MATE: Additional Assumptions

Assumption B. (*Weak Monotonicity of Mean Potential Outcomes Within Strata).*

B1. $E[Y(1)|ap] \geq E[Y(1, S(0))|ap]$

B2. $E[Y(1, S(0))|k] \geq E[Y(0)|k]$, for $k = n0, n1, ap$

- B1 assumes $LMATE_{ap} \geq 0$. On average, the mechanism effect is non-negative (Recall: $MATE = \pi_{ap} LMATE_{ap}$)
- B2 assumes $LNATE \geq 0$ for all strata: On average, the net effect is non-negative for all strata

Bounds on NATE and MATE: Additional Assumptions

In the context of our application:

- Assump. B1: attainment of degree has a non-negative **average** effect on employment and earnings for *ap* stratum (consistent with economic theory)

- Assump. B2: *LNATE* for all strata are non-negative$\longrightarrow$ the combination of the rest of the channels through which assignment to JC affects outcomes (e.g., health services, counseling, job search assistance) has a non-negative effect on **average** outcomes
  - May get evidence about its plausibility by looking at the bounds on $LNATE_{n1}$, $LNATE_{n0}$ and $LNATE_{ap}$ (without using Assump. B)

Bounds on NATE and MATE: Additional Assumptions

Assumption C (*Weak Monotonicity of Mean Pot. Out. Across Strata).*

$$C1.\ E\left[Y\left(1\right)|n1\right] \geq E\left[Y\left(1,S\left(0\right)\right)|ap\right] \geq E\left[Y\left(1\right)|n0\right]$$

$$C2.\ E\left[Y\left(0\right)|n1\right] \geq E\left[Y\left(0\right)|ap\right] \geq E\left[Y\left(0\right)|n0\right]$$

$$C3.\ E\left[Y\left(1\right)|n1\right] \geq E\left[Y\left(1\right)|ap\right] \geq E\left[Y\left(1\right)|n0\right]$$

- It formalizes the notion that some strata are likely to have more favorable characteristics and thus better potential outcomes
- It does not impose restrictions on the signs of *LNATE* and *LMATE*.
- Assump. A1, A2 and C give two *testable implications* that can be used to falsify the assumptions:
  $E[Y|T=1,S=1] \geq E[Y|T=1,S=0]$ and
  $E[Y|T=0,S=1] \geq E[Y|T=0,S=0]$
- Can get indirect evidence about its plausibility by looking at average baseline characteristics of the strata (e.g., pre-treatment outcomes)

Assumption C in our empirical application

- Assump. C: **average** potential outcomes of those who obtain a degree only if assigned to JC is no less (no greater) than the corresponding average potential outcomes of those who never (always) obtain a degree regardless of assignment to JC.
  - Likely to hold given the characteristics of the individuals expected to belong in each strata
- Average characteristics at baseline (e.g., labor market outcomes) of individuals in each stratum conform to the above assumption
  - For example, the probability of being employed in the year *prior* to randomization is $(0.15;$ s.e. $0.01)$ for $n0$ stratum, $(0.21;$ s.e. $0.01)$ for ap stratum, and $(0.22;$ s.e. $0.01)$ for $n1$ stratum
- The testable implications of the assumptions are satisfied. E.g., $E[Y|T=1,S=1]-E[Y|T=1,S=0]=0.15$ (std. er. 0.014)≥ 0

Extension: Bound effect of S on Y using T as invalid IV

- **Idea:** We can use the bounds on $MATE$ to bound an *average effect of S on Y for the ap strata* (under $T=1$)
- In a related paper (F and F-L, 2013), we show that:

$$LATE_{ap}^{SY} \equiv E\left[Y(1,1) - Y(1,0)|ap\right] = \frac{MATE}{E\left[S(1) - S(0)\right]}$$

  - Thus, can get bounds on $LATE_{ap}^{SY}$ by dividing the bounds on $MATE$ by the effect of T on S

- Relation to Instrumental Variables (IV):
  - Let T be an IV, S be a treatment, and Y be the outcome.
  - The *ap* stratum are the compliers
  - We allow the IV to have a direct effect on $Y \Rightarrow$ we do not need the exclusion restriction assumption, which assumes all the effect of the IV on the outcome works through the treatment, i.e., the net effect of the IV is zero (always hard to justify!)
  - The bounds on $LNATE_{n0}$ and $LNATE_{n1}$ can shed light on the validity of the exclusion restriction, which implies $LNATE_{n0} = LNATE_{n1} = 0$

Link to Mechanism Analysis and Invalid IVs

- Question of interest in causal mediation or mechanism analysis

$$\begin{tikzcd} \textit{Treatment } (T) \arrow[r] \arrow[rr, bend left] & \textit{Mechanism } (S) \arrow[r] & \textit{Outcome } (Y) \end{tikzcd}$$

  - What part of the effect of T on Y works through S? (*mechanism effect*)
  - What part of the effect of T on Y is net of S? (*net effect*)

- **Invalid IV link:** Disentangle the part of the effect of the IV (Z) on the outcome (Y) that works through the treatment (D) (*mechanism effect*) from the part that works through the other channels (*net effect*).

$$\begin{tikzcd} \textit{Instrument } (Z) \arrow[r] \arrow[rr, bend left] & \textit{Treatment } (D) \arrow[r] & \textit{Outcome } (Y) \end{tikzcd}$$

- In this setting, the mechanism or indirect effect plays the role of the reduced-form effect of Z on Y when the *ER* holds.

Extension: Bound effect of S on Y using T as invalid IV

When are bounds on *LATE* using an invalid IV helpful?

- **First,** in experiments or quasi-experiments where the instrument may not satisfy the exclusion restriction
  - Consider AIR's example, where the Vietnam-era draft lottery is used as an *IV* to estimate the effect of military service on earnings.
  - Draft could have affected earnings through channels other than military service (e.g., by affecting schooling decisions to postpone conscription)
  - Our bounds can be used to bound *LATE* without imposing the exclusion restriction assumption
  - Also, our bounds can be used to shed light on the validity of the *IV*: If our bounds on $LNATE_{n0}$ and $LNATE_{n1}$ exclude zero, this implies the IV has a net effect on Y
- **Second,** to exploit existing experiments to learn about treatment effects other than those for which the experiments where originally designed for; e.g., our current empirical application

Empirical Application

- In our application:
  - T: random assignment to JC
  - S: attainment of a HS,GED or vocational degree
  - Y: we consider two outcomes, employment and weekly earnings 12 quarters after RA
- Sample size is 8,020: 2,975 control and 5,045 treated individuals.
- Note: there is no-perfect compliance in the NJCS (some people in the treatment group do not enroll in JC, while some in the control group do). Thus, ATEs below are really "intention-to-treat" (ITT) effects.

Empirical Application

- Some relevant point estimates:

Average Treatment Effects (under A1)	*Estimate*	*Std. Error*
ATE JC on employment	0.04	(0.011)
ATE JC on weekly earnings	\$18.11	(4.759)
ATE JC on attainment of degree	0.21	(0.011)
Strata Proportions (under A1 & A2)		
π_{n0}	0.34	(0.007)
π_{n1}	0.45	(0.009)
π_{ap}	0.21	(0.007)

For reference:

Probability of being employed for control group: 0.61

Average weekly earnings for control group: \$171

Probability of attaining a degree for control group: 0.45

Estimated Bounds and 95 percent C.I. for LNATE

Outcome: Employment

Assumptions:	A1, A2		A1, A2, B		A1, A2, C		A1, A2, B, C	
Parameters↓	LB	UB	LB	UB	LB	UB	LB	UB
$LNATE_{n1}$	-0.10	0.34	0	0.34	$\boxed{0.04}$	0.34	0.04	0.34
	(-0.13, 0.36)		(0, 0.36)		$\boxed{(0.02, 0.036)}$		(0.02,0.36)	
$LNATE_{n0}$	-0.37	0.24	0	0.24	-0.02	0.24	0	0.24
	(-0.42, 0.29)		(0, 0.29)		(-0.05, 0.29)		(0, 0.29)	

Outcome: Weekly earnings (in dollars)

Assumptions:	A1, A2		A1, A2, B		A1, A2, C		A1, A2, B, C	
$LNATE_{n1}$	-98.8	114.7	0	114.7	$\boxed{15.4}$	114.7	15.35	114.7
	(-112.2, 131.8)		(0, 131.8)		$\boxed{(3.49, 131.8)}$		(3.49, 131.8)	
$LNATE_{n0}$	-96.4	111.1	0	111.1	-6.3	111.1	0	111.1
	(-113.5, 122.5)		(0, 122.5)		(-41.8, 17.2)		(0, 16.7)	

Half-median unbiased estimates. CIs in parenthesis cover the parameter with 95% prob.

- Under A1, A2 and C, $LNATE_{n1}$ is statistically greater than zero $\Rightarrow$ there are other benefits of JC besides degree attainment (i.e., the other JC services also matter), at least for this stratum. It also supports Assumption B.

Estimated Bounds and 95 % C.I. for NATE & MATE

Outcome: Employment

Assumptions→	A1, A2, B		A1, A2, C		A1, A2, B, C	
Parameters↓	LB	UB	LB	UB	LB	UB
MATE	0	0.04	-0.09	$\boxed{0.03}$	0	$\boxed{0.028}$
	(0, 0.06)		(-0.10, 0.037)		(0, 0.037)	
NATE	0	0.04	$\boxed{0.01}$	0.13	$\boxed{0.014}$	0.04
	(0, 0.062)		(-0.012, 0.157)		$\boxed{(0.003, 0.062)}$	

Outcome: Weekly earnings (in dollars)

	A1, A2, B		A1, A2, C		A1, A2, B, C	
MATE	0	19.5	-35.9	$\boxed{14.7}$	0	$\boxed{13.9}$
	(0, 28.6)		(-42.1, 17.2)		(0, 17.1)	
NATE	0	18.1	$\boxed{3.4}$	54.1	$\boxed{4.02}$	18.1
	(0, 27.1)		(−5.3, 65.9)		(0, 27.2]	

- *ATE* on: employment $(0.04$; s.e. $0.01)$; earnings $(\$18.11$; s.e. $4.75)$
- Under A1, A2, B & C, *NATE* is statistically greater than zero for employment

Empirical Application: Results

The bounds for *NATE* and *MATE* provide valuable information:

- For example, under A1, A2, B and C, the part of the *ATE* of JC on employment (earnings) that is due to the attainment of a HS, GED or vocational degree is *at most* 68% (77%)

- Hence, the other components of Job Corps besides academic and vocational training (e.g., counseling, health services, job search assistance) are beneficial to the improvement of these two labor market outcomes.
  - This is important for policy purposes

Bounds and 95 % C.I. for effect of S on Y

Outcome: Employment						
Assumptions→	A1, A2, B		A1, A2, C		A1, A2, B, C	
Parameters↓	LB	UB	LB	UB	LB	UB
MATE	0	0.04	-0.09	$\boxed{0.03}$	0	$\boxed{0.028}$
	(0, 0.06)		(-0.10, 0.037)		(0, 0.037)	
$LATE_{ap}^{SY}$	0	0.21	-0.45	$\boxed{0.15}$	0	$\boxed{0.133}$
	(0, 0.308)		(-0.47, 0.175)		(0, 0.166)	
Outcome: Weekly earnings (in dollars)						
MATE	0	19.5	-35.9	$\boxed{14.7}$	0	$\boxed{13.9}$
	(0, 28.6)		(-42.1, 17.2)		(0, 17.1)	
$LATE_{ap}^{SY}$	0	86.7	-169.9	$\boxed{70.5}$	0	$\boxed{66.2}$
	(0, 129.2)		(-185.5, 80.2)		(0, 77.5)	

- Average effect of JC on degree attainment $\big(0.21$; s.e. $0.01\big)$
- $LATE_{ap}^{SY}$: average effect for *ap* stratum of attaining a HS, GED, or vocational degree on the outcome, when assigned to participate in JC

Empirical Application: Results

Our bounds on $LATE_{ap}^{SY}$ are informative:

- For example, under Assumptions A1, A2, B and C, the *average effect of attaining a HS, GED or vocational degree on employment (earnings) for the ap stratum* is at most 0.133 ($66.2), about a 21% (38%) increase

- Here, we use assignment to JC as an instrument for attaining a degree *without* using the exclusion restriction assumption (i.e., using an *invalid* instrument)
  - In fact, $LNATE_{n1} = 0.04 > 0$ under A1, A2 and C, which implies the instrument is not valid (i.e., it has a direct or net effect on the outcome)

Stata Programs

Stata Programs to replicate results: See file "Documentation" in Dropbox.
Notes:

- As before, programs work with current direction of inequalities.
- If you need programs for other directions/cases, please e-mail me/us (cflore32@calpoly.edu).
    - For these bounds we have already derived all cases (different ranking of strata, different direction of monotonicity) and have most (all?) programs in stata in some working form (some cleaner than others).