Comment on "Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps" by **Xuan Chen** and **Carlos A. Flores**

Eduardo Fajnzylber

Sixth Annual Meeting of the
IMPACT EVALUATION NETWORK (IEN)
Latin American and the Caribbean Economic Association (LACEA)
September 19th and 20th, 2012
Graduate School of Education, Harvard University

Summary

- The authors derive **nonparametric bounds** for ATE in the presence of sample selection and Noncompliance
  - On top of AIR (1996) assumptions, they introduce
    - Monotonicity of Selection in Treatment status (S in D)
    - Mean dominance
- They apply these bounds to evaluate the wage effects of the **Job Corps** (JC) program
- This is an **important** question, as wages (not only employment) are an important outcome in most programs

Job Corps Example

- Sample of 9090 individuals
- Randomization: Z
  - Treatment group (Z=1): 5491
    - 73.8% enrolled in JC (D=1) by week 208
    - 60.7% employed (S=1) at week 208
  - Control group (Z=0): 3599
    - 4.4% enrolled in JC (D=1) by week 208 (**Non Compliance**)
    - 56.6% employed (S=1) at week 208
- AIR(1996): LATE of D on S <u>identified</u> for compliers
- **Sample Selection**: Wage (Y) only observed if employed (S=1)

Object of interest

- Average Treatment Effect on wages for the **compliers** who would be employed regardless of treatment assignment ("**Always employed**"):

$$\Delta = E[Y^*(1) - Y^*(1)|cEE]$$

Assumptions

- AIR (1996)
  - A1: Random assignment

    (Y*,S, D) independent of Z
  - A2: Exclusion restriction

    Z does not affect directly S or D
  - A3: Non-zero effect of Z on D
  - A4: Monotonicity of D(Z): No defiers

    D(Z=1)≥D(Z=0) for every individual
- New assumptions:
  - A5: Monotonicity of S(D)

    S(D=1) ≥S(D=0) for every individual
  - A6: Mean dominance

$$E[Y(1)|cEE] \geq E[Y(1)|cNE]$$

Results

Proposition 1 *If Assumptions 1 to 5 hold, then $L_{cEE} \leq \Delta \leq U_{cEE}$. L_{cEE} and U_{cEE} are lower and upper bounds for Δ given by:*

$$L_{cEE} = LY_{1,cEE} - \overline{Y}^{001} \frac{p_{01|0}}{p_{01|0} - p_{01|1}} + \overline{Y}^{101} \frac{p_{01|1}}{p_{01|0} - p_{01|1}}$$

$$U_{cEE} = UY_{1,cEE} - \overline{Y}^{001} \frac{p_{01|0}}{p_{01|0} - p_{01|1}} + \overline{Y}^{101} \frac{p_{01|1}}{p_{01|0} - p_{01|1}},$$

Proposition 2 *If Assumptions 1 to 6 hold, then $L_{cEE} \leq \Delta \leq U_{cEE}$. L_{cEE} and U_{cEE} are lower and upper bounds for Δ, where U_{cEE} is equal to the upper bound for Δ given in Proposition 1 and L_{cEE} equals:*

$$L_{cEE} = LY_{1,cEE} - \overline{Y}^{001} \frac{p_{01|0}}{p_{01|0} - p_{01|1}} + \overline{Y}^{101} \frac{p_{01|1}}{p_{01|0} - p_{01|1}},$$

with

$$LY_{1,cEE} = \frac{p_{11|1}\overline{Y}^{111} - p_{11|0}\overline{Y}^{011}}{p_{11|1} - p_{11|0}}.$$

Comments on A5

- A5: Monotonicity of S(D)

  S(D=1) ≥S(D=0) for every individual
  - "It cannot be the case that is less employed as a result of joining JC"
  - "No one can be negatively affected (in terms of employment)"

- Differences with A4 (D(Z=1)≥D(Z=0))
  - Z is not a choice of the individual, D is a choice, an outcome
  - S could be affected by factors not controlled by the individual (like labor market)
    - It could happen that D(Z=1)<D(Z=0) for some individuals

Comments on A5

- Authors acknowledge individuals could be less employed **as a result** of enrollment:
  - "Lock-in" effect (unemployed while being trained)
  - Higher reservation wage
  - ➔These should be short term effects
- Other possibilities (more long term):
  - Training could increase skills in nonemployed sectors: Voluntary work, **marriage**

Comments on A5

- In empirical application, important to **understand** why Hispanics showed negative effects of JC on employment and earnings
  - Dropping Hispanics from sample might not be enough to guarantee validity of A5

Comments on A6

- A6: Mean dominance

$$E[Y(1)|cEE] \geq E[Y(1)|cNE]$$

  - "Mean $Y(1)$ of always-employed compliers is greater than or equal to that of those who would be employed only if they enrolled in JC."
  - Not clear what the **intuition** might be behind this assumption
  - Shouldn't $E[Y(1)|cNE]$ be $E[Y^*(1)|cNE]$?
  - Authors suggest evaluating baseline characteristics of these two strata, particularly baseline Y
    - Results for JC are not encouraging (though imprecisely estimated)

Appendix Table A1: Average Baseline Characteristics for the cEE and cNE Strata

	Entire Sample			Non-Hispanics		
	cEE	cNE	$cEE - cNE$	cEE	cNE	$cEE - cNE$
Female	.396	.630	-.234	.390	.544	-.154
	(.015*)	(.165*)	(.174)	(.016*)	(.174*)	(.145)
Age at Baseline	18.44	19.19	-.749	18.39	19.18	-.786
	(.056*)	(.690*)	(.736)	(.069*)	(.509*)	(.632)
White, Non-hispanic	.299	.260	.039	.369	.259	.110
	(.012*)	(.196*)	(.133)	(.015*)	(.196*)	(.135)
Black, Non-Hispanic	.445	.622	-.177	.550	.624	-.074
	(.013*)	(.158*)	(.166)	(.015*)	(.176*)	(.147)
Has Child	.161	.229	-.068	.151	.210	-.059
	(.011*)	(.112*)	(.110)	(.012*)	(.111**)	(.110)
Number of children	.215	.356	-.141	.209	.280	-.071
	(.018*)	(.187**)	(.200)	(.019*)	(.170)	(.102)
Personal Education	10.22	10.34	-.123	10.24	10.27	-.036
	(.040*)	(.504*)	(.520)	(.048*)	(.402*)	(.434)
Ever Arrested	.230	.223	.007	.228	.292	-.064
	(.012*)	(.138**)	(.136)	(.013*)	(.113*)	(.121)
At Baseline						
Have job	.241	.174	.068	.244	.159	.084
	(.011*)	(.108)	(.115)	(.012*)	(.102)	(.110)
Weekly hours worked	24.07	25.27	-1.196	24.05	25.23	-1.187
	(.583*)	(6.265*)	(6.766)	(.612*)	(5.768*)	(6.160)
Weekly earnings	113.86	120.08	-6.219	115.48	142.57	-27.09
	(2.987*)	(39.90*)	(40.90)	(3.500*)	(34.21*)	(36.51)
Had job, Prev. Yr.	.714	.585	.129	.718	.588	.130
	(.013*)	(.141*)	(.151)	(.014*)	(.136*)	(.136)
Months Employed, Prev.Yr.	4.346	3.286	1.060	4.435	2.935	1.500
	(.122*)	(1.201*)	(1.280)	(.137*)	(1.105*)	(1.201)
Earnings, Prev.Yr.	3396.2	3136.2	260.02	3377.6	2879.7	497.88
	(128.63*)	(1185.0*)	(1250.6)	(138.55*)	(1000.5*)	(1005.0)

Additional comments

- Assumption A5: Implies assuming that program has an effect on D for all individuals
  - Could be **polemic** to assume an effect for the program you are evaluating but
  - This can be corroborated by estimating LATE on different subgroups
    - Requires the ability to estimate LATE on different subgroups
  - These are still **mean comparisons**, not necessarily enough to capture monotonicity for all individuals
- Is it possible to perform **Montecarlo** studies to quantify departures from the asumptions?
- Would it be possible to develop bounds if <u>a fraction λ</u> does not comply with A5?
  - Or what value of λ would bring the lower bound to zero?

Additional comments

- What does it mean to be "always employed"? Or "Employed only if in treatment group"?
  - Is it an individual attribute?
  - Employment is not only a function of the characteristics of the individual
- Some difficulties with **notation**
  - $Y(1)$ refers to $Y(D=1)$
  - $S(1)$ refers to
    - $S(Z=1)$ in the definition of EE,NE, EN,NN
    - $S(D=1)$ in the definition of A5