The effect of parental education on children's scholastic achievement: a non-parametric bounds analysis

by

Dimitris Christelis — CSEF, CFS, and CEPAR

Julian Messina — World Bank

Maria Ana Lugo — World Bank

Aim

- Estimate the effect of parents' education on student school performance and attendance

- Use observational micro survey data

- Address endogeneity of parents' education using non-parametric bounds techniques (Manski 1989, 1997; Manski and Pepper, 2000; de Haan, 2011)

Observable and unobservable magnitudes

- Outcome Y
- Treatment D, generic value d
- Average outcome for treatment t
- $E(Y(t)) = E(Y(t)|d = t)P(d = t) + E(Y(t)|d \neq t)P(d \neq t)$
- We do not observe $E(Y(t)|d \neq t)$, i.e. the counterfactual
- We observe everything else

Observable and unobservable magnitudes (cont.)

- If $E(Y(t)|d = t)$= $E(Y(t)|d \neq t)$ then treatment is exogenous, and $E(Y(t)) = E(Y(t)|d = t)$, which is observed in the data

- In most cases, however, $E(Y(t)|d = t) \neq E(Y(t)|d \neq t)$

- Need to replace $E(Y(t)|d \neq t)$ with something observable

Observable and unobservable magnitudes (cont.)

- We thus replace $E(Y(t)|d \neq t)$ with observable magnitudes M_L and M_U such that

$$M_L \leq E(Y(t)|d \neq t) \leq M_U$$

- Then we calculate lower and upper bounds $LB_E(t)$ and $UB_E(t)$ for $E(Y(t))$

- We have

$$LB_E(t) = E(Y(t)|d = t)P(d = t) + M_L P(d \neq t)$$
$$\leq E(Y(t)) \leq$$
$$E(Y(t)|d = t)P(d = t) + M_U P(d \neq t) = UB_E(t)$$

Observable and unobservable magnitudes (cont.)

- Suppose we have 2 treatments u and t with $u > t$. We then have

$$LB_E(t) \leq E(Y(t)) \leq UB_E(t)$$
$$LB_E(u) \leq E(Y(u)) \leq UB_E(u)$$

- The average treatment effect $\mathrm{ATE}(\mathrm{u}, t) = E\big(Y(u)\big) - E\big(Y(t)\big)$ can then be bounded.

Observable and unobservable magnitudes (cont.)

- Suppose we have 2 treatments u and t with $u > t$. We then have

$$UB_E(t) - LB_E(u)$$
$$\leq$$
$$ATE(u,t) = E\big(Y(u)\big) - E\big(Y(t)\big)$$
$$\leq$$
$$UB_E(u) - LB_E(t)$$

No assumptions bounds

- In $M_L \leq E(Y(t)|d \neq t) \leq M_U$ we put $M_L = Y_{min}$ and $M_U = Y_{max}$ Then we have

$$E(Y(t)|d = t)P(d = t) + Y_{min}P(d \neq t) \\ \leq E(Y(t)) \leq \\ E(Y(t)|d = t)P(d = t) + Y_{max}P(d \neq t)$$

- These bounds tend to be very wide, which is the price for making no assumptions. Narrower bounds require additional assumptions.

Mean Monotone Treatment Response (MMTR)

- If treatment $u > t \rightarrow E\big(y(u)\big) \geq E\big(y(t)\big)$
- In our context, parents' education weakly increases test scores
- Reasons (de Haan 2011): higher income, better help with homework, better role models

Implications of MMTR

- Lower bound of $E(Y(t))$
  $$E(Y(t)|d=t)P(d=t)+Y_{min}P(d\neq t)=$$
  $$Y_{min}P(d<t)+E(Y(t)|d=t)P(d=t)$$
  $$+Y_{min}P(d>t)$$

- MMTR implies, however, that for $d<t\rightarrow$ $E(Y(d)|d<t)\leq E(Y(t)|d<t)$

- Therefore, we can substitute in the lower bound the observable $E(Y(d)|d<t)$ for the unobservable counterfactual $E(Y(t)|d<t)$

Implications of MMTR (cont.)

- Since $E(Y(d)|d < t) \geq Y_{min}$ the lower bound becomes larger and the width of the range of the bounds decreases

- The lower bound becomes

$$E(Y(d)|d < t)P(d < t) + E(Y(t)|d = t)P(d = t) + Y_{min}P(d > t)$$

- The upper bound is affected in a similar way

Monotone Treatment Selection (MTS)

- If treatment $u > t \rightarrow E(y(w)|d = u) \geq E(y(d)|d = t)\ \forall w$

- Parents with higher education create an environment in which children will do better in any circumstances, due e.g. to genetics

Implications of MMTR + MTS

- Lower bound of $E(Y(t))$

$$E(Y(d)|d < t)P(d < t) + E(Y(t)|d = t)P(d = t) + Y_{min}P(d > t)$$

- MTS implies, however, that for $d > t \rightarrow E(Y(t)|d > t) \geq E(Y(t)|d = t)$

- Therefore, we can substitute in the lower bound the observable $E(Y(t)|d = t)$ for the unobservable counterfactual $E(Y(t)|d > t)$

Implications of MMTR + MTS (cont.)

- Since $E(Y(t)|d = t) \geq Y_{min}$ the lower bound becomes larger and the width of the range of the bound decreases further

- The lower bound thus becomes equal to

$$E(Y(d)|d < t)P(d < t) + E(Y(t)|d = t)(P(d = t) + P(d > t))$$

- Analogously, the upper bound becomes equal to $E(Y(t)|d = t)\big(P(d = t) + P(d < t)\big) + E(Y(d)|d > t)P(d > t)$

MMTR + MTS are testable

- If treatment $u > t \rightarrow$

$$E(y(u)|d = u) \geq$$
$$E(y(u)|d = t) \geq$$
$$E(y(t)|d = t)$$

- The first inequality is due to MTS, the second to MMTR
- Hence $E(y(u)|d = u) \geq E(y(t)|d = t)$ and both these quantities are observable. Hence we can test the joint MMTR + MTS assumptions, and in our data they are not refuted

Instruments

- One can use instruments to further narrow the bounds

- Instruments are useful if one wants the lower bound of the ATE to become larger than zero. With only MMTR+MTS this is not possible (Manski, 2000).

Exogenous instruments

- An instrument z is exogenous if
$$E(y(d)|z = m) = E\big(y(d)\big), \forall m$$

- Hence one can increase the lower bound by maximizing over all values of the instrument, and decrease the upper bound by minimizing over all values of the instrument, i.e.
$$\max_m LB_E(d|z = m) \leq E\big(y(d)\big) \leq \min_m UB_E(d|z = m)$$

Monotone instruments

- Exogenous instrument are hard to find; we will not use any

- A weaker instrument is a monotone one. An instrument z is monotone if
$$m \leq l \rightarrow E(y(d)|z = m) \leq E(y(d)|z = l)$$

- Hence, for any value m of the instrument one can increase the lower bound by maximizing over all values of the instrument smaller than m, and decrease the upper bound by minimizing over all values smaller than m, and then take the weighted average of these bounds

Monotone instruments (cont.)

- One can use another variable as a second monotone instrument.

- Maximization or minimization is done over combinations of instrument values, and only if one combination weakly dominates another

Data

- PISA survey 2009
- Age of students: 15 - 16
- Countries: Argentina, Brazil, Chile, Colombia Mexico, Panama, Peru, Uruguay
- Scores in math and reading tests
- Education of parents: primary or less, secondary, post-secondary
- Use number of household physical assets/appliances (car, computer, internet connection, cable TV) as instruments
- Use the other parent’s education as a second instrument

Results on test scores: Chile
Descriptive Statistics

Score	Mean	Median	Std. Dev	Min	Max	No. of obs
Math	420.7	417.6	80.1	194.7	782.1	5,669
Reading	449.6	451.4	82.8	145.8	711.0	5,669

Results on test scores: Chile
ATE (Post-Secondary, Primary)

Method	Math Score				Reading Score			
	Point Estimate		Low 95% CI	High 95% CI	Point Estimate		Low 95% CI	High 95% CI
Exogenous Treatment Selection	70.7		64.8	76.5	71.8		65.8	77.8
Methods using bounds	Lower Bound	Upper Bound	Low 95% CI	High 95% CI	Lower Bound	Upper Bound	Low 95% CI	High 95% CI
No assumptions	-537.6	570.7	-542.3	574.9	-494.8	512.7	-499.7	517.3
MMTR	0.0	570.7	0.0	574.9	0.0	512.7	0.0	517.3
MMTR + MTS	0.0	70.9	0.0	76.6	0.0	72.0	0.0	77.4
MMTR + MTS + 1 instr.	0.0	55.9	0.0	62.2	0.0	55.7	0.0	62.7
MMTR + MTS + 2 instr.	0.3	36.1	0.0	47.7	0.7	38.2	0.0	47.8

Probability of Attendance (selection)

- In some countries, a considerable share of students leaves school before age 15-16
- PISA does not provide information on this
- Use data from SERCE 2009
- Match variables (parents' education, instruments) with PISA
- In Mexico: 65% attendance when the mother has primary education, 94% when secondary, 99% when post-secondary

Results on probability of school attendance: Mexico

Method	Probability of Selection			
	Point Estimate		Low 95% CI	High 95% CI
Exogenous Treatment Selection	0.332		0.310	0.354
Methods using bounds	Lower Bound	Upper Bound	Low 95% CI	High 95% CI
No assumptions	-0.529	0.596	-0.541	0.608
MMTR	0.000	0.596	0.000	0.608
MMTR + MTS	0.000	0.335	0.000	0.357
MMTR + MTS + 1 instr.	0.004	0.205	0.001	0.233
MMTR + MTS + 2 instr.	0.006	0.170	0.000	0.224

Discussion

- Used non-parametric methods based on bounds in order to assess the importance of parents' education on children test scores and school attendance

- Cautionary tale: ATEs can be much smaller (but still relevant) and more uncertain than the estimates assuming exogeneity suggest

Discussion (cont.)

- Advantages of the method
  - Non-parametric: no assumptions about functional form
  - Deals with all forms of endogeneity of the treatment
  - Does not involve any other variables
  - Estimates the average treatment effect (and not the local average treatment effect as in most IV estimation contexts)
  - Allows for complete heterogeneity of the treatment effect
  - Uses relatively mild assumptions, some of which are testable (MMTR + MTS)

Discussion (cont.)

- Advantages of the method
  - Is completely transparent about how each assumption affects the outcome
  - The ATE can be computed for quantiles as well
  - Makes no assumption about behavior or expectations
  - Allows the use of instruments if available, and they need not be exogenous
  - Analysis can be done within subsamples of interest
  - Relatively easy to compute – means or kernel densities used. No need to maximize likelihoods
  - Cross-sectional data not a problem

Discussion (cont.)

- Disadvantages of the method
  - Not easy to bound the ATE away from zero – importance depends on the context
  - Instruments are needed to move the lower bound of the ATE away from zero
  - Sometimes the width of the range defined by the bounds can be large – but one should be wary of unwarranted certainty
  - Estimation is done in subsamples defined by the treatment/instrument combination – large samples might be needed
  - Continuous instruments need to be discretized

Discussion (cont.)

- Treatment of selection
  - In order to estimate the unconditional ATE taking into account selection, we need to make selection a function of the treatment
  - Involves calculating probabilities of treatment conditional on selection
  - Starting point: Blundell et. al (2007)

Discussion (cont.)

- $E(Y|d = t) = E(Y|d = t, S = 1)P(S = 1) + E(Y|d = t, S = 0)\big(1 - P(S = 1)\big)$

- Need to estimate (with no help from randomization) the unconditional expectation $E\big(Y(t)\big) = E(Y(t)|S(t) = 1)P(S(t) = 1) + E((Y(t)|S(t) = 0)(1 - P(S(t) = 1))$

- None of the above are fully observable – need to take into account counterfactual parts