Sharp Bounds on Average Treatment Effects in the Presence of Sample Selection Bias and Survey Non-response

Roman Nedela (joint work with Lukáš Lafférs)

Matej Bel University - Dept of Mathematics

15 September 2016

What is partial identification

This presentation is about identification of parameter of interest (average treatment effect) under credible assumptions.

Assumptions + Data $\rightarrow$ Results

Charles Manski: *The Law of Decreasing Credibility:* The credibility of inference decreases with the strength of the assumptions maintained.

Assumptions have to be made in order to learn something about properties of an unobserved counter-factual distribution.

These assumption may or may not be strong enough be **point** identify the quantity of interest.

If only weak assumptions are made, the quantity of interest may be **partially** identified.

This Paper

Replicates and extends the results in

Problem1 Lee, David S. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." The Review of Economic Studies 76.3 (2009): 1071-1102.

Problem2 Behaghel, Luc, et al. "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models." Review of Economics and Statistics 97.5 (2015): 1070-1080.

Lee (2009) and Behaghel et al. (2015) derived analytic formulas for sharp bounds on ATE.

We write a computer program that

- can replicate their results (so complicated proofs are not necessary)
- allows us to study sensitivity analysis of identifying assumptions

Why is it (potentially) interesting

We want to know which assumptions are important and how important they are.

No existing analytical results are available.

If you change an assumption, you don't have to redo the proof, instead, change a few lines of code.

Problem 1: Bounds on Average Treatment Effect under sample selection bias - Lee (2009):

Job training programme in the U.S.

What is the effect of the job programme on earnings?

Two effects:

- program increase probability of employment (quantity effect)
- program increase wages (quality effect) $\leftarrow$

It is difficult to separate them.

Problem 1: Bounds on Average Treatment Effect under sample selection bias - Lee (2009):

Two main identifying assumptions are made:

1. Treatment (training program) is independent of potential selection (employment) and potential outcome (wage).
2. Treatment has nonnegative impact on selection (employment).

What we do

Formulate the problem of finding lowest-highest possible ATE as a constrained optimization.

We will conduct a search in the space of all probability distributions that are

1. consistent with the model assumptions
2. compatible with the observed data

This search reduces to a linear program $\implies$ easy to solve.

What we do

We have to carefully translate all the model's identifying assumptions into restrictions on the joint probability distribution of all the variables in the model.

$$(Y(1), Y(0), S(1), S(0), D)$$

Problem 1: Mathematical formulation

Random vector $U = (Y(1), Y(0), S(1), S(0), D) \in \mathcal{U}$. Joint probability distribution $\pi : \mathcal{U} \mapsto \langle 0, 1 \rangle$.

- Treatment (assignment to job training) $D \in \mathcal{D}$.
- Potential outcomes (discretized wage 1 year after the programme) $Y(d) \in \mathcal{Y}, \quad \forall d \in \mathcal{D}$.
- Potential selection (indicator variable of employment status) $S(d) \in \mathcal{S}, \quad \forall d \in \mathcal{D}$.

Random vector $O = (Y, R, D) \in \mathcal{O}$ (observed), with probability distribution $\pi_O : \mathcal{O} \mapsto \langle 0, 1 \rangle$.

- Observed outcome $Y \in \mathcal{Y}$.
- Observed selection $S \in \mathcal{S}$.
- Treatment $D \in \mathcal{D}$.

Problem 1: Set of model restrictions

(c1) Compatibility with observed data:
$Y = S \cdot \big(Y(1)D + Y(0)(1-D)\big)$ and
$S = S(1)D + S(0)(1-D)$.

(c2) Independence of treatment: $D \perp (Y(1), Y(0), S(1), S(0))$.

(c3) Monotonicity in D: $\pi\big(S(1) \geq S(0)\big) = 1$.

Problem 1: Goal

$$\Phi = \{\pi \in \Pi(\mathcal{U}) : \pi \text{ satisfies (c1)-(c3)}\}$$

Our goal is to find the sharp bounds on average treatment effect

$$ATE_c \equiv \mathbb{E}_\pi\big(Y(1) - Y(0) | S(1) = 1, S(0) = 1\big),$$

i.e. average treatment effect of those who would be employed in both states as treated and also as non-treated.

Problem 1: Linear program formulation

Now bounds on conditional ATE can be written as $\min_{\pi}\big(\sum_{i=1}^{n} c_i \cdot \pi_i\big)$ and $\max_{\pi}\big(\sum_{i=1}^{n} c_i \cdot \pi_i\big)$, $\forall \pi \in \Phi$, with constraints:

- Observed distribution

$$\sum_{i=1}^{n} \pi_i \cdot \mathbb{1}\big(s_i \cdot (y_i^1 d_i + y_i^0 (1-d_i)) = y', s_i^1 d_i + s_i^0 (1-d_i) = s', d_i = d'\big) =$$

$$= \pi_O(\{O = o'\}),$$

$$\forall o' = (y', s', d') \in \mathcal{O}, \text{ with } s_i = s_i^1 d_i + s_i^0 (1 - d_i).$$

Problem 1: Linear program formulation

- Independence

$$\sum_{i=1}^{n} \pi_i \cdot \left(\frac{\mathbb{1}\left(y_i^1 = y_1', y_i^0 = y_0', s_i^1 = s_1', s_i^0 = s_0', d_i = 1\right)}{\pi_O\left(\{O \in \mathcal{O} : D = 1\}\right)} - \right.$$

$$\left. - \frac{\mathbb{1}\left(y_i^1 = y_1', y_i^0 = y_0', s_i^1 = s_1', s_i^0 = s_0', d_i = 0\right)}{\pi_O\left(\{O \in \mathcal{O} : D = 0\}\right)} \right) = 0,$$

$$\forall y_1', y_0' \in \mathcal{Y}, \forall s_1', s_0' \in \mathcal{S}.$$

- Monotonicity

$$\sum_{i=1}^{n} \pi_i \cdot \mathbb{1}\left(s_i^1 \geq s_i^0\right) = 1.$$

Problem 1: Relaxing assumptions

1. Relaxed independence $|\pi(D)\pi\big(Y(1), Y(0), S(1), S(0)\big) - \pi\big(Y(1), Y(0), S(1), S(0), D\big)| \leq \alpha_I$
2. Relaxed monotonicity: $\pi\big(S(1) \geq S(0)\big) \geq 1 - \alpha_M$

Problem 2: Bounds on Average Treatment Effect under sample nonresponse - Behaghel et al. (2015)

Job training programme in France.

Individuals were repeatedly attempted to reach before obtaining response in a phone survey (questionarie).

Treatment affected individuals survey response behaviour $\implies$ respondents sample was biased.

Question of interest: Difference of average employnment status between job training participants and non-participants.

Problem 2: Mathematical formulation

Random vector $U = (Y(1), Y(0), R(1), R(0), V, N, Z) \in \mathcal{U}$. Joint probability distribution $\pi : \mathcal{U} \mapsto \langle 0, 1 \rangle$.
The components of $\mathcal{U}$ are the following:

- Treatment (assignment to a programme) $Z \in \mathcal{Z}$.
- Potential outcomes (indicator variable of employment status) $Y(z) \in \mathcal{Y}, \quad \forall z \in \mathcal{Z}$.
- Potential responses (indicator variable of survey response) $R(z) \in \mathcal{R}, \quad \forall z \in \mathcal{Z}$.
- Reluctance to respond $V \in \mathcal{V}$.
- Number of call attempts $N \in \mathcal{N}$, with $w_\infty = w_{max} + 1$ meaning no response until w_{max}. A fixed constant w_{max} stands for the maximum number of call attempts.

Problem 2: Set of model restrictions

(C1) Compatibility with observed data: $Y = Y(1)Z + Y(0)(1 - Z)$ and $R = R(1)Z + R(0)(1 - Z)$.

(C2) Latent variable threshold-crossing response model: $\forall z \in \mathcal{Z} : R(z) = \mathbb{1}\big(V < p(w_{max}, z)\big)$.

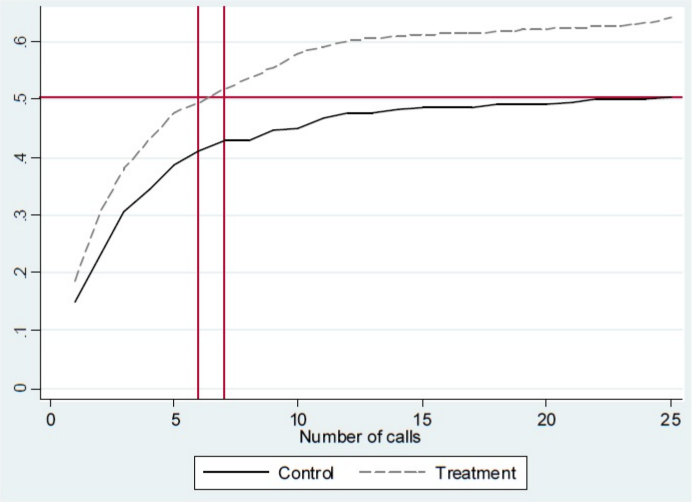
(C3) Independence of treatment: $Z \perp (Y(0), Y(1), V)$.

(C4) Monotonicity in w: $\forall z \in \mathcal{Z}, \forall w_1 \geq w_2 \in \mathcal{N} : p(w_1, z) \geq p(w_2, z)$.

(C5) Uniformity of unobserved reluctance: $V \sim \text{Unif}(0, 1)$.

(C6) Definition of N: $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N} : \big(V < p(w, z), Z = z \iff N \leq w, Z = z\big)$.

p(w,z) function



Problem 2: A complication

Variable V is continuous. So finite-dimensional formulation is not possible.

We discretize V to V^* and provide a proof that this discretization does not affect bounds on ATE.

Problem 2: Discretized model restrictions

Restrictions that define the model with discrete V^*:

(D1) Compatibility with observed data: $Y = Y(1)Z + Y(0)(1 - Z)$ and $R = R(1)Z + R(0)(1 - Z)$.

(D2) Latent variable threshold-crossing response model: $\forall z \in \mathcal{Z}, R(z) = \mathbb{1}\big(V^* \leq p(w_{max}, z)\big)$.

(D3) Independence of treatment: $Z \perp Y(0), Y(1), V^*$.

(D4) Monotonicity in w: $\forall z \in \mathcal{Z}, \forall w_1 \geq w_2 \in \mathcal{N} : p(w_1, z) \geq p(w_2, z)$.

(D5) Distribution of discretized reluctance: $\forall v^* \in \mathcal{V}^*, F^*_{V^*}(v^*) = v^*$.

(D6) Equivalence of observed calls: $\forall z \in \mathcal{Z}, \forall w \in \mathcal{N}, \big(V^* \leq p(w, z), Z = z \iff N \leq w, Z = z\big)$.

Similarly, we define

$$\Phi^* = \{\pi^* \in \Pi(\mathcal{U}^*) : \pi \text{ satisfies (D1)-(D6)}\}.$$

Results

- Linear program formulation of the problems
- Replication existing analytical results via linprog
- Extension to sensitivity analysis

Thank you for your attention.

roman.nedela@gmail.com