Bounds

Advanced Econometrics 3

Oliver Snellman

07.05.2020

Contents

Introduction

In some cases the correct aim of analysis is partial identification, where an interval of possibilities is identified instead of a point estimate. This excercise deals with that kind of questions, where additionally there is a sample selection problem.

We wish to find sensible and informative upper and lower bounds for average treatment effect. These bounds should be based on the data. As we will see, without postulating additional structure on the problem, the bounds are generally not going to be too useful. I go through several assumptions, which can grant us more interesting bounds.

Treatment analysis

Treatment variable Treatment analysis studies the average causal effect of a treatment variables $t \in T$ (realized treatment $z \in T$) on an outcome variable y in the treatment group. There is also a control group, which doesn't receive the treatment t. The outcome y could be for example earnings and the treatment t could be the education level or inclusion in a government subsidy program.

Covariates Covariates $x \in X$, including other factors influencing y, can be used to divide the data into comparable subgroups. These variables can be for example IQ or parents' income. Then the causal impact of treatment t on the outcome y(t) can be teased out by comparing it's effect on the treatment and control groups (separately in each subpopulation x). The covariates can be divided further into $x = (w, v)$ s.t. $X = W \times V$, V standing for instruments and w for control variables. This notation is a bit confusing, as the treatment variable (education) in this exercise is going to be x, and there are no covariates involved.

The problem is: Given that we observe a random sample of the population and find the probability distributino for $P(x, z, y)$, how do we find the distribution for $P(y(z))$?

Sample selection Additionally, it might be the case that y is observable alongside with the covariates x only when some criterion is fulfilled, D=1, whereas only the covariates are observed otherwise, when D=0. If y is only partially observed, we are dealing with missing data, which can raise problems related to sample selection.

Data Let's check out the data for this exercise.

Table 1: Aghion, Akcigit, Hyytinen and Toivanen (2017)

Own education	Pr(invent)	% of obs	N
base	0.002	0.223	78755
2dary	0.003	0.444	156504
college	0.014	0.218	76882
MSc	0.065	0.106	37483
PhD	0.175	0.009	3044
Total	0.013	1.000	352668

The variable $y(t) := Pr(invent|t = x) \in [0, 1]$ denotes the conditional probability of a person in education group $x \in X = \{\text{base, secondary, college, MSc, PhD}\}$ becoming an inventor. The treatment t will be considered as one of the education levels in X. We do not have covariates in this exercise.

We are interested in the average treatment effect $\mathbb{E}[y_j(t)|x_j]$, where j is an index for an individual or a group, and $y(t)$:= Pr(invent | t). This treatment effect means the change in probability of becoming an inventor, for a person with education level x, if he had obtained a different education level t. In other words, what would be the probability of becoming an inventor, if everyone had educated themselves to the level t? More specifically, we are interested in finding bounds for this average treatment effect in line with AAHT (2017) paper.

The missing information problem arises, because we don't have access to the counterfactuals - how would the Pr(invent | x) have changed, if the members in group x had pursued a different level of education x'? Or in other words, what is the impact of the actual education level, apart from other influencing factors (covariates), such as innate creativity or social environment? For what we know, these covariates could be the main influencers behind both education decision and the probability of inventing something. If we could observe the probabilities from alternate realities, it would be possible to find a point estimate for the causal influence of an educational level on the probability, as we could potentially control for everything else. But alas, we must settle down with bounds as the alternate realities are still out of bounds.

The poor man's bound -solution is to use the lowest and highest possible values, inf and sup of y. This is not informative with regards to binary dependant variables, and rarely otherwise either. Next we define concepts, which add structure onto the problem and allow for construction of more informative and narrow bands.

Adding structure on the problem

We need to pose some assumptions regarding the relationship between X, T and y, in order to identify some reasonable bands.

The two most important assumptions are MTS and MTR. In essence, MTS relates to selection effect and states, that some innate ability drives people to acquire both higher outcome y and higher value on variable related to treatments (x in our case). MTR links the treatment itself to the outcome regardless of other factors.

Let's start off by stating the **IV assumption** of $v \in V \in X$: v is an IV if it has mean independence,

$$\mathbb{E}[y(t)|w, v = u] = \mathbb{E}[y(t)|w, v = u']$$

which means, that v doesn't influence the expected value of y at all.

A stronger assumption is that of MIV.

Monotone Instrumental Variable (MIV)

Let V of instruments be an ordered set such that $\forall\ u_i \in V$

$$[u_{i+1} \geq u_i] \ \rightarrow \ \Big[\mathbb{E}[y(t) \mid u_{i+1}] \geq \mathbb{E}[y(t) \mid u_i]\Big]$$

where $u_j \in V$ and the index i belongs to a suitable range with regards to the support of V.

MIV means, that an increase in variable u increases the outcome variable within the same treatment t. It is likely, that being high on openness to experience (Big 5 personality aspect) would increase the chance of becoming an inventor, when comparing to someone else who has the same (any) degree, but is lower on openness.

As we don't have covariates, MIV doesn't have an interpretation in this exercise.

Next up in MTS we switch focus from the covariates to assumption regarding the impact of variable related to the treatment.

Monotone Treatment Selection (MTS)

Under MTS the variable which determines the treatment is itself a MIV, in our example the education level x. This means, that individuals with higher schooling have weakly higher mean in their probability of becoming an inventor.

Let X of instruments be an ordered set such that $\forall\ x_i \in X$

$$[x_{i+1} \geq x_i] \ \rightarrow \ \Big[\mathbb{E}\big[\text{Pr(invent} \mid \text{t)} \mid x_{i+1}\big] \geq \mathbb{E}\big[\text{Pr(invent} \mid \text{t)} \mid x_i\big]\Big]$$

where $t \in X$. In words, person with higher education level x_{i+1} would have a higher probability of inventing in any counterfactual scenario, where both aqcuired the same aducation level t.

In our case this would mean, that people with higher innate probability of becoming an inventor also tend to pursue a higher education. The interpretation is, that a person A with a Master's degree would be an inventor more likely than a person B with a Bachelor's degree, if A would have finished studying at Bachelor's level as well.

A peculiar feature of MTS is, that it allows for education to have a negative impact on the inventor-probability. It might be, that the person A from the above example with Master's is not as likely to become an inventor with her current degree as person B is with Bachelors. This has a significant impact on the bounds, calculated under MTS, as we shall see.

In order to construct the MTS bounds, we need to assume that the outcome variable has known smallest K_0 and largest K_1 possible values. In our case this is easy, as the probability distribution is bounded between zero and one by definition.

$$\begin{aligned}
&K_0 * Pr(x < t) + \mathbb{E}[y(t)|x = t] * Pr(x \geq t) && |\ \text{Lower bound} \\
&\leq \mathbb{E}[y(t)] && |\ \text{Expected effect of treatment} \\
&\leq K_1 * Pr(x > t) + \mathbb{E}[y(t)|x = t] * Pr(x \leq t) && |\ \text{Upper bound}
\end{aligned}$$

where $K_0 = \mathbb{E}[y(x)|x]$ and $K_1 = \mathbb{E}[y(x)|x]$.

To get strict bounds, which any counterfactual observation can't surpass, we postulate that in lower bounds, all education levels lower than the treatment under consideration, x<t, have the worst possible probability, $K_0 = 0$, and observations with equal or higher education have the lowest possible probability $\mathbb{E}[y(t)|x = t]$. Likewise, on the upper bound all x>t receive the highest value, $K_1 = 1$, and $x \leq t$ reveice the highest possible probability, that is $\mathbb{E}[y(t)|x = t]$.

Here we also can see how MTS allows for negative impact of education on P(invent). Take the observations with higher education levels $x*$ than the treatment under consideration. We know the current probabilities associated with their actual education levels $K* \in [K_0, K_1]$, and now we are considering what would happen if their education level was lowered to t. In the upper bound we assign all $x > t$ the highest possible probability,

$K_1 * Pr(x > t)$, which is higher than any of their current probabilities. This means, that it is considered possible under MTS, that lowering education from x* to t could increase the outcome from K* to K_1 for some observations. This is only possible with MTS, but not with MTR.

Monotone Treatment Response (MTR)

MTR relates the effect straigth to the treatment itself, education level. Under MTR, each person's probability function is weakly increasing in conjectured education level. Let X of instruments be an ordered set such that $\forall\ t_i \in X$

$$[t_{i+1} \geq t_i] \ \rightarrow\ \Big[\Pr(\text{invent}|t_{i+1}) \geq \Pr(\text{invent}|t_i)\Big]$$

Now higher education level increases the probability of being an inventor in itself, regardless of the selection effects. In effect, MTR denies the possibility of education having a negative effect on the inventor-probability.

MTR would mean, that education x itself has a causal effect on P(Innovate). Our data could conform to MTR, because the probability of becoming an inventor does increase monotonically by education level. However, we only have aggregated data, and can't tell if this holds for each individual, as is required.

Combining MTS and MRS together yields the following bounds

$$\begin{aligned}
&\sum_{x<t} \Big(\mathbb{E}[y(x)|x] * Pr(x < t)\Big) + \mathbb{E}[y(t) \mid x] * Pr(x \geq t) && \mid \text{Lower bound} \\
&\leq \mathbb{E}[y(t)] && \mid \text{Expected effect of treatment} \\
&\leq \sum_{x>t} \Big(\mathbb{E}[y(x)|x] * Pr(x > t)\Big) + \mathbb{E}[y(t) \mid x] * Pr(x \leq t) && \mid \text{Upper bound}
\end{aligned}$$

Here the order of influence is preserved in the sense, that lowering one's education in counterfactual from level x to x', with $x > x\prime$, will lower the output at most to y(x') but can't increase the potential output beyond y(x). There can be a large difference in bounds under MTS and MTR, if all realized outcomes y() are small relative to the highest possible value of y, as we'll see shortly.

Finding the actual bounds

Bounds with MTS

a) Secondary

In the first scenario we ask, what would be (the bounds for) the probability of becoming an inventor, if everyone had finished their education after the secondary degree?

The MTS bounds are obtained by

$$
\begin{aligned}
&0 * Pr(\text{invent} \mid \text{base}) + \mathbb{E}[\Pr(\text{invent} \mid \text{Secondary})] \ * Pr(x \in X_{/base}) && \mid \text{Lower bound} \\
&\leq \mathbb{E}[\Pr(\text{invent} \mid \text{secondary})] && \mid \text{Everyone with the treatment} \\
&\leq 1 * Pr(x \in X_{/\{\text{base, secondary}\}}) + \mathbb{E}[\Pr(\text{invent} \mid \text{secondary})] * Pr(x \in \{\text{Base, Secondary}\}) && \mid \text{Upper bound}
\end{aligned}
$$

and with numbers (see the actual values gathered in table below)

$$
\begin{aligned}
&0.003 * (1 - 0.223) && \mid \text{Lower bound} \\
&\leq 0.003 \\
&\leq (0.218 + 0.106 + 0.009) + 0.003 * (0.223 + 0.444) && \mid \text{Upper bound}
\end{aligned}
$$

b) college

$$
\begin{aligned}
&\mathbb{E}[\Pr(\text{invent} \mid \text{College})] \ * Pr(x \in \{\text{College, MSc, PhD}\}) && \mid \text{Lower bound} \\
&\leq \mathbb{E}[\Pr(\text{invent} \mid \text{College})] && \mid \text{Expected treatment effect} \\
&\leq 1 * Pr(x \in \{\text{MSc, PhD}\}) + \mathbb{E}[\Pr(\text{invent} \mid \text{college})] * Pr(x \in \{\text{Base, Secondary, College}\}) && \mid \text{Upper bound}
\end{aligned}
$$

$$
\begin{aligned}
&0.014 * (0.218 + 0.106 + 0.009) && \mid \text{Lower bound} \\
&\leq 0.014 \\
&\leq (0.106 + 0.009) + 0.014 * (0.218 + 0.223 + 0.444) && \mid \text{Upper bound}
\end{aligned}
$$

c) MSc

$$
\begin{aligned}
&\mathbb{E}[\Pr(\text{invent} \mid \text{MSc})] \ * Pr(x \in \{\text{MSc, PhD}\}) && \mid \text{Lower bound} \\
&\leq \mathbb{E}[\Pr(\text{invent} \mid \text{MSc})] && \mid \text{Expected treatment effect} \\
&\leq 1 * Pr(x = \text{PhD}) + \mathbb{E}[\Pr(\text{invent} \mid \text{MSc})] * Pr(x \in \{\text{Base, Secondary, College, MSc}\}) && \mid \text{Upper bound}
\end{aligned}
$$

$$
\begin{aligned}
&0.065 * (0.106 + 0.009) && \mid \text{Lower bound} \\
&\le 0.065 \\
&\le 0.009 + 0.065 * (0.106 + 0.218 + 0.223 + 0.444) && \mid \text{Upper bound}
\end{aligned}
$$

The bounds

```
calculations <- as.matrix(data[,-1])

secondary_lower <- calculations[2,1] * (1 - calculations[1,2])
secondary_upper <- sum(calculations[3:5,2]) + calculations[2,1] * sum(calculations[1:2,2]) # ssh oauhiu

college_lower <- calculations[3,1] * sum(calculations[3:5,2])
college_upper <-  sum(calculations[4:5,2]) + calculations[3,1]*sum(calculations[1:3,2])

msc_lower <- calculations[4,1] * sum(calculations[4:5,2])
msc_upper <- calculations[5,2] + calculations[4,1] * sum(calculations[1:4,2])

bounds <- round(cbind(c(secondary_upper, secondary_lower), c(college_upper, college_lower), c(msc_upper
rownames(bounds) <- c("Upper", "Lower")
colnames(bounds) <- c("Secondary", "College", "MSc")

kable(bounds, "latex", booktabs=T, caption="MTS bounds") %>%
      kable_styling(latex_options = "hold_position")
```

Table 2: MTS bounds

	Secondary	College	MSc
Upper	0.335	0.127	0.073
Lower	0.003	0.005	0.007

Interpretation

Take b), bounds for treatment effect of college education. In counterfactual, where everyone in our sample finished off their education no sooner or later than college, the expected probability of a college educated person in that reality becoming an inventor is at minimum 0.003, and at maximum 0.335, based on our data. That is the best data driven estimate for this partial identification, under the assumption of MTS. The a) and c) have similar stories.

Looking at the MTS bounds formula more closely, we can also say something about the counterfactuals between education levels. Take c), everyone with an education level below MSc could at maximum reach a probability of P(invent | MSc) = 0.065, if they continued to studying until MSc, hence max(P(invent | lower than MSc upgrading to MSc)) = P(invent | MSc). Because we only assumed MTS and not MTR, the lowest possible probability that they could get after reaching MSc could be 0, even though it was larger before the re-education. Under MTR the lower bound would be their respective probability before the counterfactual. On the other hand, people with PhD, if downgraded, could at minimum get the same probability as current

MSc, P(invent | MSc), and at maximum the wollopping max(P(invent | PhD to MSc)) = 1. Here we see the drastic difference between MTS and MTR, of which the latter would have given an upper bound of only P(invent | PhD)=0.175. Similar reasoning applies to a) and b).

Upper bound for a treatment secondary -> PhD

Under MTS assumption, the counterfactual of someone with only secondary education, continuing until reaching PhD, would yield a lower bound of zero and upper bound of max(P(invent | secondary -> PhD)) = P(invent | PhD) = 0.175. So this extreme treatment could potentially increase the probability of this person at most by P(invent | PhD) - P(invent | secondary) = 0.172 %.

If both MTS and MTR are assumed, the upper bound remains the same, but the lower bound increases from zero to P(invent | secondary).