Bounding Average Treatment Effects using Linear Programming

Lukáš Lafférs

Department of Economics, NHH Norwegian School of Economics

March 13, 2013

This paper is about Identification.

What can we learn from **data** and **assumptions**?

What drives our results?

Which assumptions are important?

Which assumptions are not important?

Here I present a general framework for studying bounds on Average Treatment Effects.

I show how to relax various assumptions to see their identification strength.

Demonstrate it on an application of an effect of mother's on child's schooling.

Setup and Notation

Individual i: $y_i(.) : T \to Y$ (individuals do not interact)

(Potential) Treatment: $t \in T$ (mutually exclusive and exhaustive)

(Potential) Outcome: $y_i(t) \in Y$

Observed Treatment: $z_i \in T$

Observed Outcome: $y_i \equiv y_i(z_i) \in Y$

Monotone instrument: $v_i \in V$

Fundamental problem: $y_i(t)$ is not observed for $t \neq z_i$

Distribution of (y_i, z_i, v_i) is observed.

Our Goal

Learn about the probability distribution of counter-factual outcomes
$P(y(t_1), y(t_2), \ldots, y(t_m))$

What are we interested in?

- $E[y(t)]$ - average treatment response
- $E[y(t)] - E[y(s)]$ - average treatment effect

Examples

- Effect of parental schooling on child's schooling
- Effectiveness of a labor participation program
- Effect of a medical intervention

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

These assumptions 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.

Examples (Manski)

Say we are interested in $E[y(t)]$

$$E[y(t)] = E[y|z = t].P(z = t) + E[y(t)|z \neq t].P(z \neq t)$$

Example

Say we are interested in $E[y(t)]$

$$E[y(t)] = E[y|z = t].P(z = t) + E[y(t)|z \neq t].P(z \neq t)$$

Observed quantities

Unobserved quantities

Example (exogenous selection)

If we assume that $E[y(t)|z = t] = E[y(t)|z \neq t]$

$$\begin{aligned} E[y(t)] &= E[y|z = t].P(z = t) + E[y(t)|z \neq t].P(z \neq t) \\ &= E[y|z = t].P(z = t) + E[y|z = t].P(z \neq t) \\ &= E[y|z = t] \end{aligned}$$

Under this assumption, $E[y(t)]$ is **point** identified.

Example (bounded support)

Suppose that $y_{min} \leq y_i(t) \leq y_{max}$

$$
\begin{array}{rcl}
LB_{E[y(t)]} & = & E[y|z=t].P(z=t) + y_{min}.P(z \neq t) \\
& \leq & \\
E[y(t)] & = & E[y|z=t].P(z=t) + E[y(t)|z \neq t].P(z \neq t) \\
& \leq & \\
UB_{E[y(t)]} & = & E[y|z=t].P(z=t) + y_{max}.P(z \neq t)
\end{array}
$$

Under this assumption, $E[y(t)]$ is **partially** identified and the interval $(LB_{E[y(t)]}, UB_{E[y(t)]})$ is called an **identified set**.

Schooling example

Suppose we are interested in an effect of mother's education of child's education (de Haan 2012).

Outcome - College degree of child i : $y_i(.) : \{0,1\} \to \{0,1\}$

(Potential) Treatment - Mother's college: $t \in \{0,1\}$

(Potential) Outcome - Child's college: $y_i(t) \in \{0,1\}$

Observed Treatment: Observed mother's college $z_i \in \{0,1\}$

Observed Outcome: Observed child's college $y_i \equiv y_i(z_i) \in \{0,1\}$

Monotone instrument - Father's schooling level: $v_i \in \{1,2,3,4\}$

Data: Wisconsin Longitudinal Study

Different assumptions

- Monotone Treatment Response (MTR) assumption $\forall i, t_2 \geq t_1 : y_i(t_2) \geq y_i(t_1)$
- Monotone Treatment Selection (MTS) assumption $\forall t, z_2 \geq z_1 : E[y(t)|z = z_2] \geq E[y(t)|z = z_1]$
- Conditional Monotone Treatment Selection (cMTS) assumption $\forall t, z_2 \geq z_1, \forall m : E[y(t)|z = z_2, v = m] \geq E[y(t)|z = z_1, v = m]$
- Monotone Instrumental Variable (MIV) assumption $\forall t, v_2 \geq v_1 : E[y(t)|v = v_2] \geq E[y(t)|v = v_1]$

Analytical bounds on $E[y(t)]$ under MTR, MTS, MIV, MTR+MTS and MTR+cMTS+MIV are available. These then translate to bounds on $E[y(t)] - E[y(s)]$.

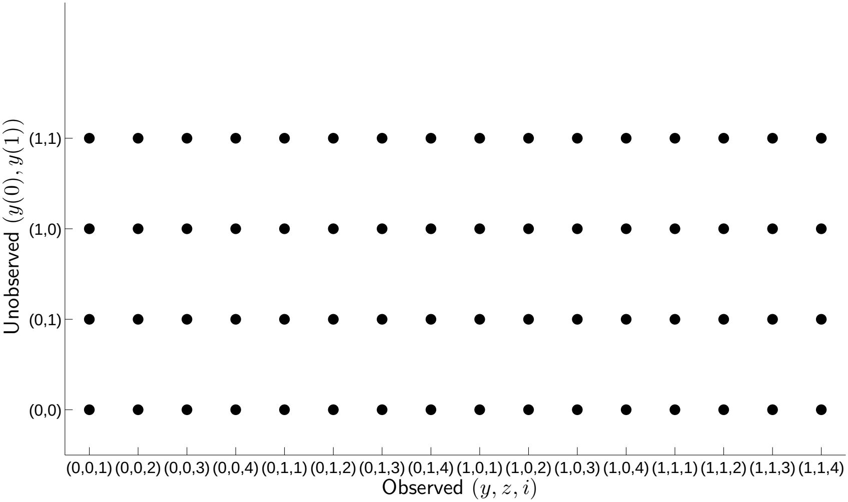
Results

Bounds on Effect of Mother's College Increase on the Probability of Child Has College Degree	
Assumptions	[Lower Bound, Upper Bound]
No Assumptions	[-0.358, 0.641]
MTS	[-0.358, 0.365]
cMTS	[-0.358, 0.214]
MTR	[0, 0.641]
MTR + MTS	[0, 0.365]
MTR + cMTS	[0, 0.214]
MTR + MTS + MIV	[0, 0.365]
MTR + cMTS + MIV	[0, 0.214]

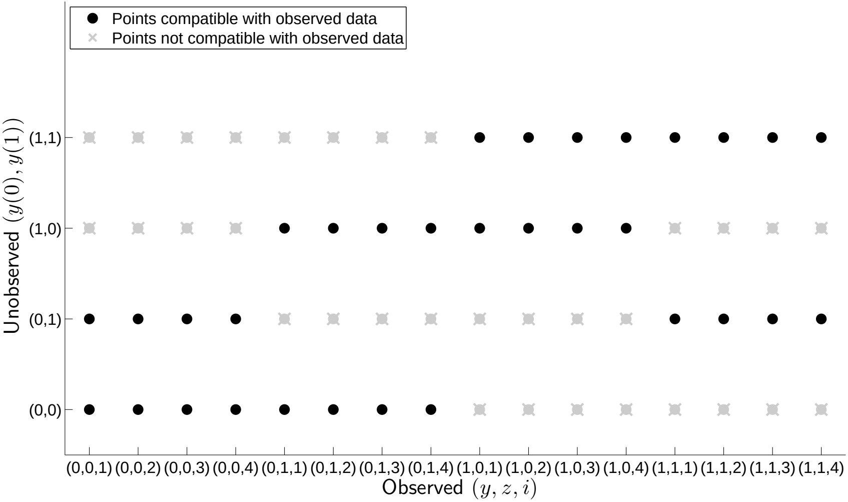
Note: Estimates are not bias corrected, $n = 16912$

How did I calculate these numbers?

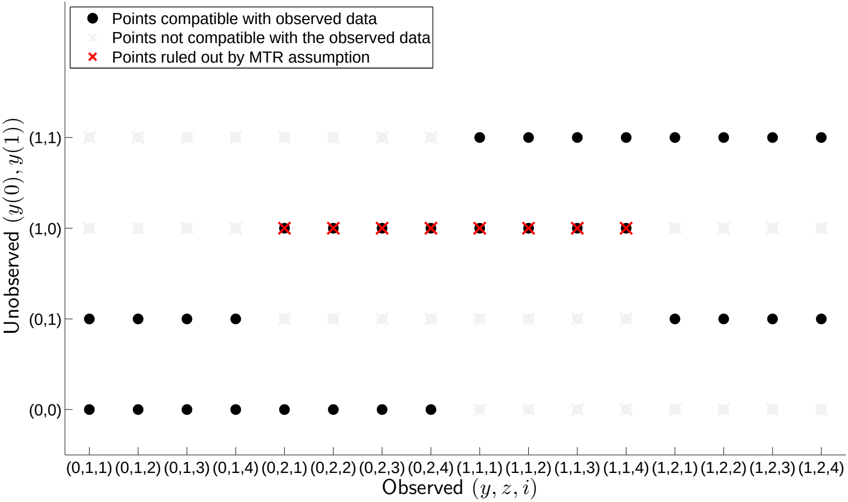
The Joint Support of $(y(0), y(1), y, z, v)$



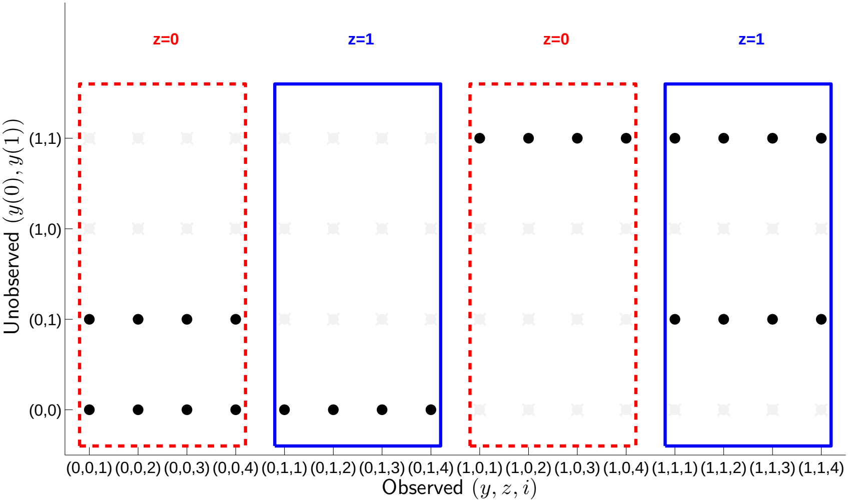
Compatibility with Observed Data: $\forall i, t : z_i = t \Rightarrow y_i = y_i(t)$



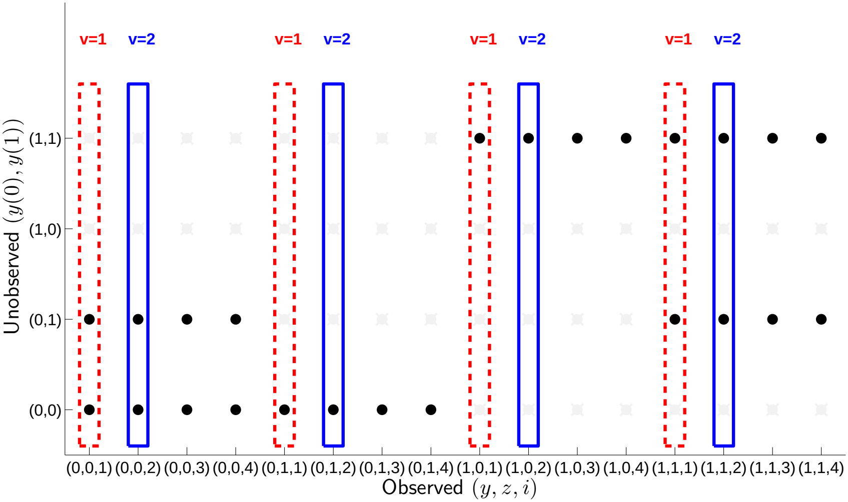
Monotone Treatment Response: $\forall i : y_i(0) \leq y_i(1)$



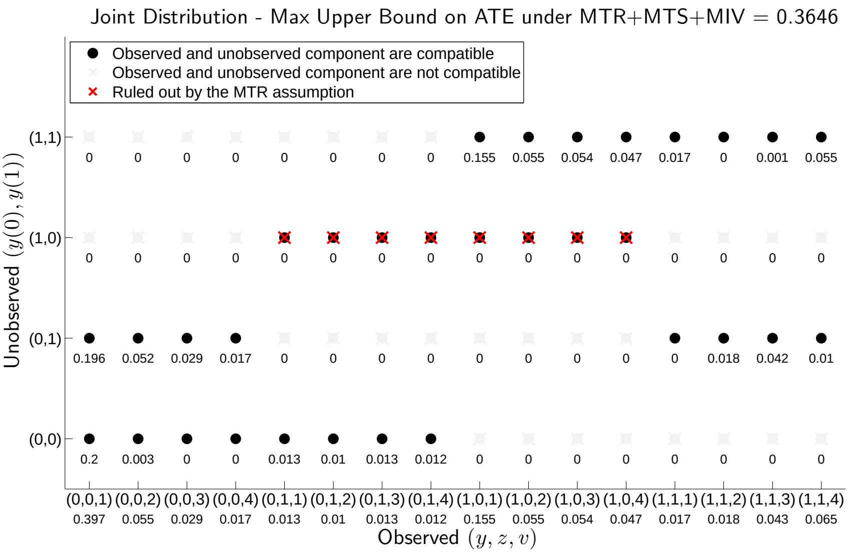
Monotone Treatment Selection: $E[y(t)|z=1] \geq E[y(t)|z=0]$



Monotone Instrumental Variable: $E[y(t)|v=2] \geq E[y(t)|v=1]$



In the empirical application



Corresponding Linear Program

$$\max_\pi \overbrace{\begin{bmatrix} 0&1&0&1&0&1&0&1&0&0&0&0&0&0&0&0&1&0&1&0&1&0&1&0 \end{bmatrix}}^{\text{Average Treatment Effect}} \times \pi$$

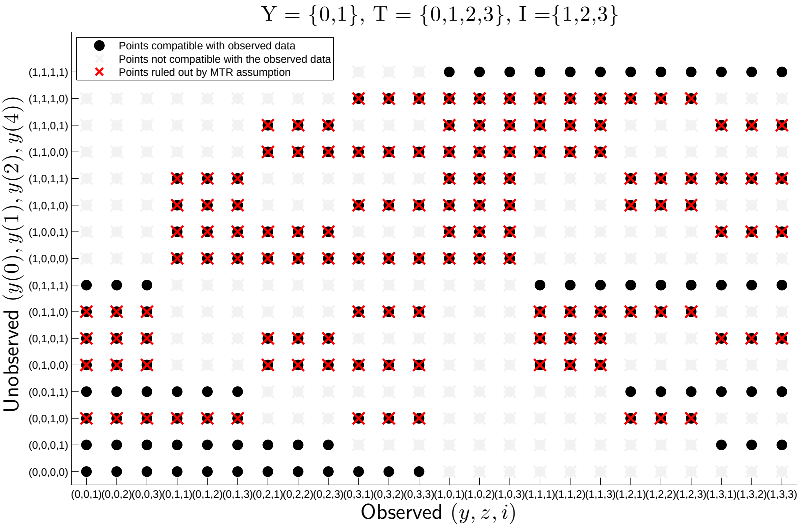
subject to

$$\text{DATA} \left\{ \begin{bmatrix}
1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1
\end{bmatrix} \times \pi = \begin{bmatrix} 0.397\\0.055\\0.029\\0.017\\0.013\\0.01\\0.013\\0.012\\0.155\\0.055\\0.054\\0.047\\0.017\\0.018\\0.043\\0.065 \end{bmatrix} \right\} \text{Observed probabilities}$$

$$\begin{matrix}\text{MTS}\\ \\ \text{MIV}\end{matrix}\;\begin{bmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&.19&.19&.19&.19&0&-.80&0&-.80&0&-.80&0&-.80\\
0&.19&0&.19&0&.19&0&.19&0&0&0&0&.19&.19&.19&.19&-.80&-.80&-.80&-.80&-.80&-.80&-.80&-.80\\
0&0&0&0&0&0&0&0&0&0&0&0&.13&-.58&0&0&0&.13&0&-.58&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&.13&-.13&0&0&0&0&.13&0&-.13&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&.14&-.13&0&0&0&0&0&.14&0&-.13\\
0&.13&0&-.58&0&0&0&0&0&0&0&0&.13&-.58&0&0&.13&.13&-.58&-.58&0&0&0&0\\
0&0&0&.13&0&-.13&0&0&0&0&0&0&0&.13&-.13&0&0&0&.13&.13&-.13&-.13&0&0\\
0&0&0&0&0&.14&0&-.13&0&0&0&0&0&0&.14&-.13&0&0&0&0&.14&.14&-.13&-.13
\end{bmatrix} \times \pi \le \begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\0\end{bmatrix}$$

$$\pi \ge \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix},$$

Pencil and Paper vs Computer?



How Robust are these results?

Using weaker assumptions our goal is to get more robust results. But are they?

Relax different assumptions and see their identification strength.

- Mis-measurement of Outcomes or Treatments (MOT)
- Relaxed Monotone Treatment Response (rMTR)
- Relaxed Monotone Treatment Selection (rMTS)
- Relaxed Monotone Instrumental Variable (rMIV)
- Missing Data

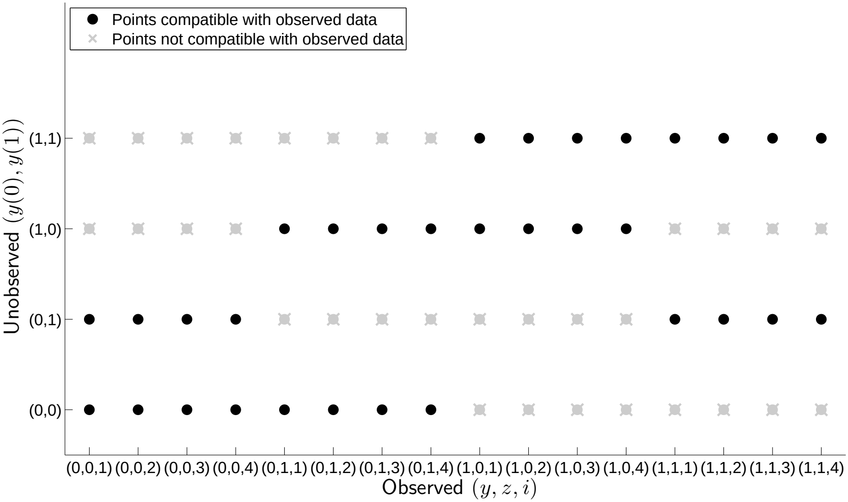
Mis-measurement of Outcomes or Treatments (MOT)

$$P[z_i = t \Rightarrow y_i = y_i(t)] \geq 1 - \alpha_{MOT}$$

Data collected mostly using phone interviews.

Instead of assuming that every individual's outcome or treatment was recorded correctly, we assume that not more than $100\alpha_{MOT}\%$ could have been recorded incorrectly.

Compatibility with Observed Data: $\forall i, t : z_i = t \Rightarrow y_i = y_i(t)$



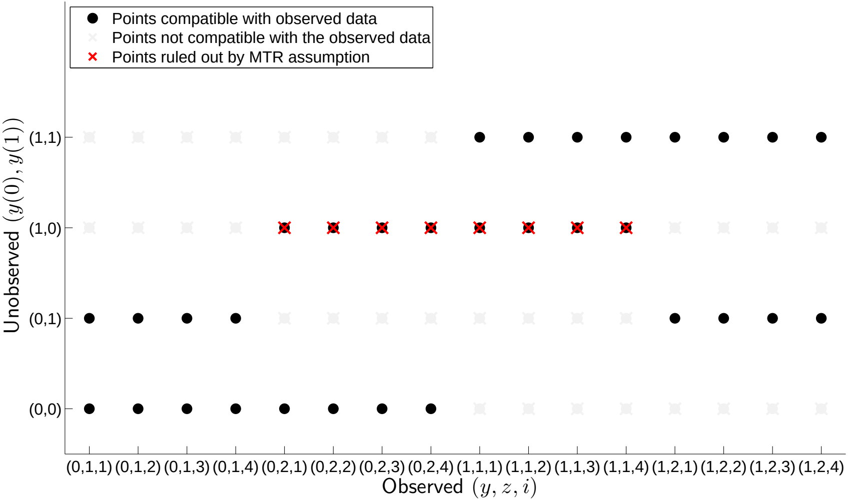
Relaxed Monotone Treatment Response (rMTR)

$$P[t_2 \geq t_1 \Rightarrow y_i(t_2) \geq y_i(t_1)] \geq 1 - \alpha_{MTR},$$

Behrman and Rosenzweig (AER, 2002) suggest that more educated woman spend less time with their children.

We allow that at most $100\alpha_{MTR}\%$ childrens' potential education is not increasing function of mother's education.

Monotone Treatment Response: $\forall i : y_i(0) \leq y_i(1)$



Relaxed Monotone Treatment Selection (rMTS)

$$\forall z_2 \geq z_1 : E[y(t)|z = z_1] - E[y(t)|z = z_2] \leq \delta_{MTS}$$

We assume that children of lower educated women may have higher potential schooling but not more than by δ_{MTS}.

Relaxed Monotone Instrumental Variable (rMIV)

$$\forall v_2 \geq v_1 : E[y(t)|v = v_1] - E[y(t)|v = v_2] \leq \delta_{MIV}$$

We assume that children of lower educated men may have higher potential schooling but not more than by δ_{MIV}.

Missing Data

Responsiveness' rates very good, around 90%.

But

- Data not missing-at-random.
- Systematic Non-responsiveness (Hauser 2005) in our dataset.

Hence cannot be ignored.

We remain agnostic about the process that drives the missingness.

Results - Robustness

Bounds on Effect of Mother's College Increase on the Probability of Child Has College Degree

MTR+cMTS+MIV

[Lower bound, Upper Bound] = **[0, 21.44%]**

	Lower Bound	Upper Bound			
	α_{MTR}	α_{MOT}	α_{MTS}	α_{MIV}	α_{MISS}
Optimistic	0.01	0.001	0.01	0.01	0.01
	-1%	23.36%	22.44%	21.44%	27.31%
Pessimistic	0.05	0.01	0.05	0.05	0.10
	-5%	35.66%	26.44%	21.44%	38.15%
Optimistic	$[-1\%, 28.62\%]$				
Pessimistic	$[-5\%, 44.10\%]$				

Note: Estimates are not bias corrected, $n = 16912$

What is the source of identification?

cMTS:
$\forall t, z_2 \geq z_1, \forall m : E[y(t)|z = z_2, v = m] \geq E[y(t)|z = z_1, v = m]$

Bounds on ATE

$MTR + cMTS$	$[0, 21.44\%]$
$MTR + cMTS + MIV$	$[0, 21.44\%]$

If cMTS holds for $v \in \{2, 3, 4\}$ only:

Bounds on ATE

$MTR + cMTS$	$[0, 46.71\%]$
$MTR + cMTS + MIV$	$[0, 27.54\%]$

What is the source of identification? (2)

Binding constraints under MTR+cMTS+MIV and
Lagrange multipliers:

$$
cMTS \left\{ \begin{array}{lcll}
E[y(0)|z=1,v=1] & \geq & E[y(0)|z=0,v=1] & 0.0303 \\
E[y(1)|z=1,v=1] & \geq & E[y(1)|z=0,v=1] & 0.5505 \\
E[y(0)|z=1,v=2] & \geq & E[y(0)|z=0,v=2] & 0.0282 \\
E[y(1)|z=1,v=2] & \geq & E[y(1)|z=0,v=2] & 0.1106 \\
E[y(0)|z=1,v=3] & \geq & E[y(0)|z=0,v=3] & 0.0554 \\
E[y(1)|z=1,v=3] & \geq & E[y(1)|z=0,v=3] & 0.0823 \\
E[y(0)|z=1,v=4] & \geq & E[y(0)|z=0,v=4] & 0.0766 \\
E[y(1)|z=1,v=4] & \geq & E[y(1)|z=0,v=4] & 0.0637
\end{array} \right.
$$

Non-binding constraints:

$$
MIV \left\{ \begin{array}{lcll}
E[y(0)|v=2] & \geq & E[y(0)|v=1] & 0 \\
E[y(1)|v=2] & \geq & E[y(1)|v=1] & 0 \\
E[y(0)|v=3] & \geq & E[y(0)|v=2] & 0 \\
E[y(1)|v=3] & \geq & E[y(1)|v=2] & 0 \\
E[y(0)|v=4] & \geq & E[y(0)|v=3] & 0 \\
E[y(1)|v=4] & \geq & E[y(1)|v=3] & 0
\end{array} \right.
$$

What is the source of identification? (3)

Binding constraints under MTR+cMTS+MIV:
(cMTS for $v \in \{2, 3, 4\}$) and Lagrange multipliers

cMTS	$E[y(0)\|z = 1, v = 2]$	$\geq$	$E[y(0)\|z = 0, v = 2]$	0.0282
	$E[y(1)\|z = 1, v = 2]$	$\geq$	$E[y(1)\|z = 0, v = 2]$	0.5768
	$E[y(0)\|z = 1, v = 3]$	$\geq$	$E[y(0)\|z = 0, v = 3]$	0.0554
	$E[y(1)\|z = 1, v = 3]$	$\geq$	$E[y(1)\|z = 0, v = 3]$	0.0823
	$E[y(0)\|z = 1, v = 4]$	$\geq$	$E[y(0)\|z = 0, v = 4]$	0.0766
	$E[y(1)\|z = 1, v = 4]$	$\geq$	$E[y(1)\|z = 0, v = 4]$	0.0637
MIV	$E[y(1)\|v = 2]$	$\geq$	$E[y(1)\|v = 1]$	0.5821

Non-binding constraints:

MIV	$E[y(0)\|v = 2]$	$\geq$	$E[y(0)\|v = 1]$	0
	~~$E[y(1)\|v = 2]$~~	~~$\geq$~~	~~$E[y(1)\|v = 1]$~~	
	$E[y(0)\|v = 3]$	$\geq$	$E[y(0)\|v = 2]$	0
	$E[y(1)\|v = 3]$	$\geq$	$E[y(1)\|v = 2]$	0
	$E[y(0)\|v = 4]$	$\geq$	$E[y(0)\|v = 3]$	0
	$E[y(1)\|v = 4]$	$\geq$	$E[y(1)\|v = 3]$	0

Statistical Inference

This is work in progress.

If bounds are smooth functions of observed probabilities

- Bootstrap - Efron and Tibshirani (1993)
- Normal approximation - Imbens and Manski (2004)
- Bayes - Moon and Schorfheide (2012), Kitagawa (2012)
- ... many others

If bounds are not smooth

- Intersection Bounds - Chernozhukov, Lee and Rosen (2013)
- Subsampling - Romano and Shaikh (2012)
- ...

My attempt: Optimize across set of observed probabilities that would not have been rejected by non-parametric test of equality of distributions.

Thank you for your attention. Any comments are very welcome.

http://sites.google.com/site/lukaslaffers

lukas.laffers@nhh.no