Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India

Michael Gechter

Pennsylvania State University

June, 2019

Introduction

- What do causal effects from one place tell us about causal effects in an another place?

Introduction

- What do causal effects from one place tell us about causal effects in an another place?
- Debate: should we implement a successful policy from one context in another?
  1. Scale up what works: **extrapolate widely** (Banerjee and He, 2008; Duflo, 2004)
  2. What works in one place may not in another: **don't extrapolate** (Pritchett and Sandefur, 2013)

Introduction

- What do causal effects from one place tell us about causal effects in an another place?
- Debate: should we implement a successful policy from one context in another?
    1. Scale up what works: **extrapolate widely** (Banerjee and He, 2008; Duflo, 2004)
    2. What works in one place may not in another: **don't extrapolate** (Pritchett and Sandefur, 2013)
- "school uniforms will surely not have the same impact in Norway (where every child who needs one, no doubt, has one) that it has in Kenya. The question is where to draw the line: Is Mexico more like Norway or more like Kenya?" Banerjee and Duflo (2009)

Introduction

- What do causal effects from one place tell us about causal effects in an another place?
- Debate: should we implement a successful policy from one context in another?
  1. Scale up what works: **extrapolate widely** (Banerjee and He, 2008; Duflo, 2004)
  2. What works in one place may not in another: **don't extrapolate** (Pritchett and Sandefur, 2013)
- "school uniforms will surely not have the same impact in Norway (where every child who needs one, no doubt, has one) that it has in Kenya. The question is where to draw the line: Is Mexico more like Norway or more like Kenya?" Banerjee and Duflo (2009)
- I derive methods to bound the average effect of a treatment based on experimental findings from another context

Example

Transfers to Mexican microenterprises

- McKenzie and Woodruff (2008) gave transfers to male microentrepreneurs in Leon in 2006
- Sustained increase in monthly profits of roughly 40% of the value of the transfer
- What do the results from McKenzie and Woodruff (2008) and data from a national microenterprise survey say about the returns to transfers other cities?

Setup

- Experiment
  - Untreated outcomes and covariates
  - Treated outcomes and covariates
- Alternative context
  - Untreated outcomes and covariates (for example)

Setup

- Experiment
  - Untreated outcomes and covariates
  - Treated outcomes and covariates
- Alternative context
  - Untreated outcomes and covariates (for example)
- If average causal effects are the same for individuals with the same characteristics, the problem is easy
- However, average effects for individuals with the same characteristics typically differ across contexts
  (Allcott, 2015; Attanasio et al., 2003)

Bounds

For individuals with the same characteristics

	Untreated	Treated
Experiment	Y_0	Y_1
Alternative context	Y_0	?

1. Differences in the distributions of untreated outcomes tell us about differences in unobservables
2. Assume that $Y_1 | Y_0 = y_0$ in the alternative context is **consistent with the experimental results**
3. Based on differences in the distributions of Y_0, I derive bounds on the average causal effect in the alternative context
    - Greater differences yield wider bounds

Tighter bounds

1. Assumptions on dependence between treated and untreated outcomes tighten the bounds
2. Tighter bounds yield the assumptions needed to draw specific conclusions about the average effect in the alternative context
   - Example: the policy has a positive average effect in the new context

Preview of empirical results

Transfers to Mexican microenterprises: extrapolating from a small experiment

- The McKenzie and Woodruff (2008) experiment was small and the authors are cautious in interpreting their results
- The benchmark method makes us overconfident in extrapolating from the experiment to all of urban Mexico

Preview of empirical results

Remedial education in India: evaluating predictions

- Banerjee, Cole, Duflo, and Linden (2007) conducted evaluations of the same remedial education program in two Indian cities
- I use the results from one city to predict the average causal effect in the other
- The benchmark method leads us to conclude that one experiment tells us nothing about the causal effect in the other
- The bounds are informative and contain the observed causal effects

Outline

Contributions relative to the benchmark method

Hotz, Imbens, and Mortimer (2005)

- Two step procedure
  1. Test for external validity: is the distribution of untreated outcomes the same for individuals with the same characteristics?
  2. Based on the test results...
     - **Fail to reject:** conditional on characteristics, expected treated outcomes from the experiment generalize to the alternative context
     - **Reject:** abandon extrapolation

Contributions relative to the benchmark method

Hotz et al. (2005)

- Two step procedure
  1. Test for external validity: is the distribution of untreated outcomes the same for individuals with the same characteristics?
  2. Based on the test results...
     - **Fail to reject:** conditional on characteristics, expected treated outcomes from the experiment generalize to the alternative context
     - **Reject:** abandon extrapolation
- Pitfalls
  1. The test is often underpowered
  2. Can we really learn nothing if we reject?

Contributions relative to the benchmark method

Hotz et al. (2005)

- Two step procedure
  1. Test for external validity: is the distribution of untreated outcomes the same for individuals with the same characteristics?
  2. Based on the test results...
     - **Fail to reject:** conditional on characteristics, expected treated outcomes from the experiment generalize to the alternative context
     - **Reject:** abandon extrapolation
- Pitfalls
  1. The test is often underpowered
  2. Can we really learn nothing if we reject?
- I quantify uncertainty about generalization from two sources:
  1. Sample size
  2. Differences in outcome distributions

Literature

1. Existing methods based on invariance assumptions for individuals with the same covariates
   (Hotz et al., 2005; Attanasio et al., 2003; Angrist and Fernández-Val, 2013; Angrist and Rokkanen, 2015; Cole and Stuart, 2010; Stuart, Cole, Bradshaw, and Leaf, 2011; Pearl and Bareinboim, 2014; Dehejia, Pop-Eleches, and Samii, 2017; Bisbee, Dehejia, Pop-Eleches, and Samii, 2015; Vivalt, 2017)
2. Differences in causal effects across contexts for individuals with the same observed characteristics
   (Allcott, 2015; Attanasio et al., 2003)
3. Methods relating outcome distributions and unobservables
   (Matzkin (2013) review, Athey and Imbens (2006) estimator coincides with mine under maximum dependence)
4. Sensitivity analyses based on bounds
   (Altonji, Elder, and Taber, 2005; Altonji, Conley, Elder, and Taber, 2013; Kline and Santos, 2013)
5. Other approaches
   Kowalski (2016); Brinch, Mogstad, and Wiswall (2015); Mogstad, Santos, and Torgovitsky (2016); Chassang, Padro i Miquel, and Snowberg (2012); Banerjee, Chassang, and Snowberg (2016)

Outline

Stylized cash transfers to Mexican microentrepreneurs

- Suppose outcomes are binary (I will generalize later)
  - low profit or high profit
- Experimental results from city e:

	Low profit	High profit
No transfers	$\frac{2}{3}$	$\frac{1}{3}$
Transfers	$\frac{1}{3}$	$\frac{2}{3}$

- Average effect of the transfers in e: the share of high-profit entrepreneurs increases by $\frac{1}{3}$

Generalizing the experimental results

- Alternative city *a*

	Low profit	High profit
No transfers	$\frac{1}{2}$	$\frac{1}{2}$

Generalizing the experimental results

- Alternative city *a*

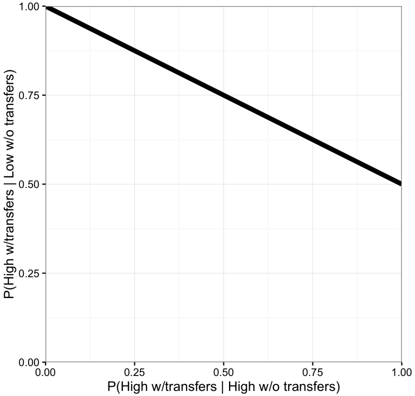
	Low profit	High profit
No transfers	$\frac{1}{2}$	$\frac{1}{2}$

- What would the average effect of transfers be in *a*?

$$\text{P(high profit with transfers} \mid \text{high profit without transfers)} \times \frac{1}{2}$$
$$+ \text{P(high profit with transfers} \mid \text{low profit without transfers)} \times \frac{1}{2}$$
$$- \frac{1}{2}$$

Assumption

- P(high profit with transfers | profits without transfers) in *a* is consistent with the experimental marginal distributions
- Each point on the line represents a possible conditional distribution



Bounds

- Assuming P(high profit with transfers | high profit without transfers) $= 1$ yields the maximum average effect: $\frac{1}{4}$

Bounds

- Assuming P(high profit with transfers $|$ high profit without transfers) $= 1$ yields the maximum average effect: $\frac{1}{4}$
- Assuming P(high profit with transfers $|$ high profit without transfers) $= 0$ yields the minimum average effect: 0

Bounds

- Assuming P(high profit with transfers | high profit without transfers) $= 1$ yields the maximum average effect: $\frac{1}{4}$
- Assuming P(high profit with transfers | high profit without transfers) $= 0$ yields the minimum average effect: 0
- Intermediate assumptions yield intermediate average effects
- The bounds are based on the Frechet (1951)-Hoeffding (1940) inequalities for a joint distribution with known marginals

Bounds

- Assuming P(high profit with transfers | high profit without transfers) $= 1$ yields the maximum average effect: $\frac{1}{4}$
- Assuming P(high profit with transfers | high profit without transfers) $= 0$ yields the minimum average effect: 0
- Intermediate assumptions yield intermediate average effects
- The bounds are based on the Frechet (1951)-Hoeffding (1940) inequalities for a joint distribution with known marginals
- **Takeaway 1:** Greater differences in untreated profits generate wider bounds
    - Example: with $\frac{2}{3}$ high-profit entrepreneurs without transfers, the bounds are -$\frac{1}{3}$ and $\frac{1}{6}$

Dependence between individuals' profits with and without transfers

- Small transfers are unlikely to cause all high-profit entrepreneurs to realize low profits
  - P(high profit with transfers | high profit without transfers) $= 0$
- It is natural to believe that entrepreneurs realizing high profits *without transfers* would also realize high profits *with transfers*

Dependence between individuals' profits with and without transfers

- Small transfers are unlikely to cause all high-profit entrepreneurs to realize low profits
  - P(high profit with transfers | high profit without transfers) $= 0$
- It is natural to believe that entrepreneurs realizing high profits *without transfers* would also realize high profits *with transfers*
- Assumptions on dependence between profits with and without transfers generate tighter bounds
  - With perfect dependence, the experiment tells me
    P(high profit with transfers | high profit without transfers) $= 1$,
    P(high profit with transfers | low profit without transfers) $= \frac{1}{2}$

Dependence between individuals' profits with and without transfers

- Small transfers are unlikely to cause all high-profit entrepreneurs to realize low profits
  - P(high profit with transfers | high profit without transfers) $= 0$
- It is natural to believe that entrepreneurs realizing high profits *without transfers* would also realize high profits *with transfers*
- Assumptions on dependence between profits with and without transfers generate tighter bounds
  - With perfect dependence, the experiment tells me P(high profit with transfers | high profit without transfers) $= 1$, P(high profit with transfers | low profit without transfers) $= \frac{1}{2}$
- **Takeaway 2:** The method tells you how much dependence you need to draw specific conclusions about the average effect of the transfers in city a
  - A zero average effect is only possible with minimum dependence: P(high profit with transfers | high profit without transfers) $= 0$
  - I.e., it has a breakdown point structure

Generalizing the example

- Dealing with covariates
  - Run the previous analysis conditioning on covariate values

Generalizing the example

- Dealing with covariates
  - Run the previous analysis conditioning on covariate values
- Non-binary outcomes
  - This adds another dimension for imperfect correlation between control and treated outcomes: leapfrogging
  - And makes computing the bounds complicated and high-dimensional
  - With discrete or discretized outcomes, the bounds can be computed as solutions to a linear programming problem

Generalizing the example

- Dealing with covariates
  - Run the previous analysis conditioning on covariate values
- Non-binary outcomes
  - This adds another dimension for imperfect correlation between control and treated outcomes: leapfrogging
  - And makes computing the bounds complicated and high-dimensional
  - With discrete or discretized outcomes, the bounds can be computed as solutions to a linear programming problem
- Sampling
  - Confidence intervals computed using the bootstrap and results from Stoye (2009)

Generalizing the example

- Dealing with covariates
  - Run the previous analysis conditioning on covariate values
- Non-binary outcomes
  - This adds another dimension for imperfect correlation between control and treated outcomes: leapfrogging
  - And makes computing the bounds complicated and high-dimensional
  - With discrete or discretized outcomes, the bounds can be computed as solutions to a linear programming problem
- Sampling
  - Confidence intervals computed using the bootstrap and results from Stoye (2009)
- The set of $F_{Y_1|Y_0}(y_1|Y_0 = y_0)$ distributions consistent with experimental results may not be large enough to include the conditional distribution in a
  - I.e., the distribution $Y_1|Y_0, a$ may be selected on Y_1 relative to the distributions consistent with experimental results
  - Bound the maximal selection allowed for any y_0
  - Yields a 2-dimensional sensitivity analysis structure

Outline

Formal setup

- Binary treatment: $T \in \{0, 1\}$
- Outcome $Y \in \mathcal{Y} \subseteq \mathbb{R}$
- $Y = TY_1 + (1 - T)Y_0$
  - Treated outcome: $Y_1 \in \mathcal{Y}_1 \subseteq \mathcal{Y}$
  - Untreated outcome: $Y_0 \in \mathcal{Y}_0 \subseteq \mathcal{Y}$
- Observed covariates: $X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$
- Two contexts: $D \in \{e, a\}$

Treatment assignment

Assumption.

Random assignment in population e**.** *$0 < P^e(T = 1) < 1$. $T|D = e$ is independent of all other random variables and $P^e(T = 1)$ is known.*

Treatment assignment

Assumption.

Random assignment in population e**.** $0 < P^e(T = 1) < 1$*.* $T|D = e$ *is independent of all other random variables and* $P^e(T = 1)$ *is known.*

Assumption.

Everyone in a **is untreated***.* $T = 0|D = a$*.*

- Again, for concreteness
- We can alternatively assume $T = 1|D = a$ or that individuals in a choose T

The average treatment effect in a

$$ATE^a = E^a[Y_1 - Y_0] = \underbrace{E^a[Y_1]}_{unknown} - E^a[Y]$$

The average treatment effect in a

$$ATE^a = E^a[Y_1 - Y_0] = \underbrace{E^a[Y_1]}_{unknown} - E^a[Y]$$

- If I knew $E^a[Y_1|Y_0 = y_0, X = x]$, I could identify $E^a[Y_1]$

The average treatment effect in a

$$ATE^a = E^a[Y_1 - Y_0] = \underbrace{E^a[Y_1]}_{unknown} - E^a[Y]$$

- If I knew $E^a[Y_1|Y_0 = y_0, X = x]$, I could identify $E^a[Y_1]$

$$E^a[Y_1] = \int_{\mathcal{X}} \left(\int_{\mathbb{R}} \underbrace{E^a[Y_1|Y_0 = y_0, X = x]}_{unknown} \underbrace{dF^a_{Y_0|X}(y_0|x)}_{identified} \right) \underbrace{dF^a_X(x)}_{identified}$$

The average treatment effect in a

$$ATE^a = E^a[Y_1 - Y_0] = \underbrace{E^a[Y_1]}_{unknown} - E^a[Y]$$

- If I knew $E^a[Y_1|Y_0 = y_0, X = x]$, I could identify $E^a[Y_1]$

$$E^a[Y_1] = \int_{\mathcal{X}} \left(\int_{\mathbb{R}} \underbrace{E^a[Y_1|Y_0 = y_0, X = x]}_{unknown} \underbrace{dF^a_{Y_0|X}(y_0|x)}_{identified} \right) \underbrace{dF^a_X(x)}_{identified}$$

- I will use $F^e_{Y_0|X}(y_0|x)$, $F^e_{Y_1|X}(y_1|x)$ and an assumption to bound $E^a[Y_1|Y_0 = y_0, X = x]$

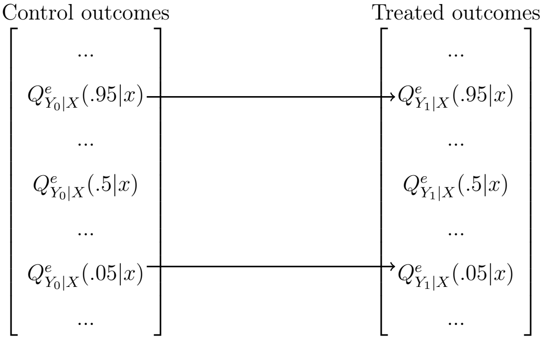
Key assumption

Consistency with experimental results

- Many joint distributions $F_{Y_0,Y_1|X}(y_0, y_1|x)$ are consistent with the identified distributions in the experiment $F^e_{Y_0|X}(y_0|x)$ and $F^e_{Y_1|X}(y_1|x)$
- The joint distributions define many conditional distributions consistent with the experimental results
- I assume that the conditional distribution $F^a_{Y_1|Y_0,X}(y_1|y_0, x)$ can be generated as one of the distributions consistent with the experimental results

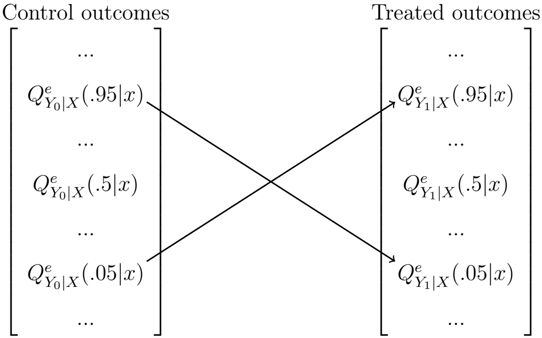
Perfect positive dependence

- $Q^e_{Y_t|X}(\alpha|x)$: the α quantile of Y_t conditional on x in e



The bounds also allow perfect negative dependence

- $Q^e_{Y_t|X}(\alpha|x)$: the α quantile of Y_t conditional on x in e



Key assumption

Consistency with experimental results: formally

Assumption.

Consistency with experimental results: *Consistency of the conditional distribution of treated outcomes in the population of interest with the experimental results:*

$$F^a_{Y_1|Y_0,X}(y_1|y_0,x) = C_1(F^e_{Y_0|X}(y_0|x), F^e_{Y_1|X}(y_1|x)|x)$$

for some copula function $C \in \mathcal{C}$ *where* $C_1(v,w|x) = \frac{\partial C(v,w|x)}{\partial v}$.

- A copula is a distribution function over the unit interval, $\mathcal{C}$ denotes the set of all valid copulas
- Each copula $C \in \mathcal{C}$ defines a joint distribution consistent with the experimental results Formal definition

Consistency with experimental results: discussion

- **Necessary condition:** the distribution of untreated outcomes fully captures all relevant differences in unobserved heterogeneity across contexts
  - $F^a_{Y_0|X}(y_0|x) = F^e_{Y_0}(y_0|x) \implies E^a[Y_1|X=x] = E^e[Y_1|X=x]$
- **Sufficient condition:** the distribution of treated outcomes for individuals with $X = x$ and $Y_0 = y_0$ is the same across contexts
  - $D \perp\!\!\!\perp Y_1 | Y_0, X$
  - However, specific deviations from this condition are allowed
  - And the condition as a whole can be relaxed

Support assumptions

Unbounded Y

Assumption.

Support of $\mathbf{Y_0}|\mathbf{X} = \mathbf{x}$: *The support of $Y_0|X = x$ in a subset of the support in e for all $X \in \mathcal{X}^a$: $Supp^a(Y_0|X = x) \subseteq Supp^e(Y_0|X = x)\ \forall x \in \mathcal{X}^a$.*

- Consistency with the experimental results only gives me $E^a[Y_1|Y_0 = y_0, X = x]$ for y_0 on the support of $Y_0|X = x$ in population e

Assumption.

Support of X: *The support of X in a subset of the support in e: $\mathcal{X}^a \subseteq \mathcal{X}^e$.*

- The previous assumptions only give me $E^a[Y_1|X = x]$ for x on the support of X in population e

Bounds

Assumption.

Expectation of $\mathbf{Y_0}$: *Y_0 has finite expectation in a: $E^a\left[|Y_0|\right] < \infty$.*

Proposition 1.

Under the previous assumptions sharp bounds on ATE^a are given by:

$$E^a[Y_1|x] \in$$

$$\Bigg[\min_{C\in\mathcal{C}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y_1 dC_1(F^e_{Y|T,X}(y|T=0,x), F^e_{Y|T,X}(y|T=1,x)|x) \right) dF^a_{Y|X}(y|x),$$

$$\max_{C\in\mathcal{C}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y_1 dC_1(F^e_{Y|T,X}(y|T=0,x), F^e_{Y|T,X}(y|T=1,x)|x) \right) dF^a_{Y|X}(y|x)\Bigg],$$

$$ATE^a \in \Bigg[\int_{\mathcal{X}} (\min \; E^a[Y_1|x])\, dF^a_X(x) - E^a[Y],$$

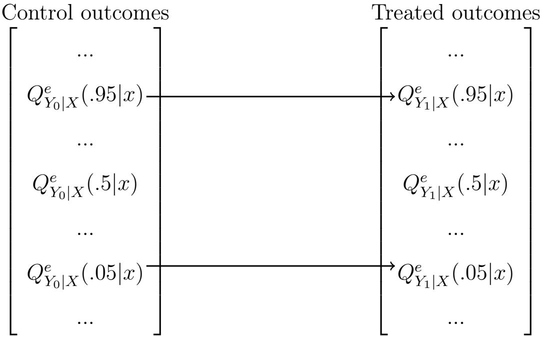
$$\int_{\mathcal{X}} (\max \; E^a[Y_1|x])\, dF^a_X(x) - E^a[Y]\Bigg].$$

Bounds with constrained dependence

- Recall from the example that negative dependence is typically unrealistic
  - Transfers will not cause all high-profit entrepreneurs without transfers to realize low profits
- We expect positive dependence between outcomes with and without the treatment

Perfect positive dependence

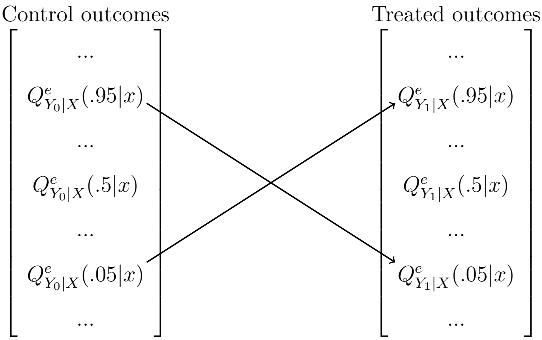
- $Q^e_{Y_t|X}(\alpha|x)$: the α quantile of Y_t conditional on x in e



- Assuming perfect positive dependence point-identifies ATE^a but is often unrealistic

The bounds also allow perfect negative dependence

- $Q^e_{Y_t|X}(\alpha|x)$: the α quantile of Y_t conditional on x in e



- We often want to exclude this case and others like it

Bounds with constrained dependence

Assumption.

Minimum dependence: *C is an element of $\mathcal{C}(\rho^L)$, the set of copula functions such that $\rho(Y_0, Y_1|X = x) \geq \rho^L$ where $\rho^L \in [0, 1]$.*

- *$\rho(Y_0, Y_1|X = x)$ is a normalized version of Spearman's non-parametric measure of dependence* ▸ Definition

Bounds with constrained dependence

Proposition 2.

Under the maintained assumptions

$$E^a[Y_1|x] \in$$

$$\left[\min_{C\in\mathcal{C}(\rho^L)} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y_1 dC_1(F^e_{Y|T,X}(y|T=0,x), F^e_{Y|T,X}(y|T=1,x)|x)\right) dF^a_{Y|X}(y|x),\right.$$

$$\left.\max_{C\in\mathcal{C}(\rho^L)} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y_1 dC_1(F^e_{Y|T,X}(y|T=0,x), F^e_{Y|T,X}(y|T=1,x)|x)\right) dF^a_{Y|X}(y|x)\right]$$

and

$$ATE^a \in$$

$$\left[\int_{\mathcal{X}} (\min\ E^a[Y_1|x])\, dF^a_X(x) - E^a[Y], \int_{\mathcal{X}} (\max\ E^a[Y_1|x])\, dF^a_X(x) - E^a[Y]\right]$$

Estimation

- $\mathcal{C}(\rho^L)$ is an infinite-dimensional space and, unlike in the example, the maximum and minimum causal effects do not occur at the boundaries in general
- Computing the bounds is challenging, even at the population level

Estimation

- $\mathcal{C}(\rho^L)$ is an infinite-dimensional space and, unlike in the example, the maximum and minimum causal effects do not occur at the boundaries in general
- Computing the bounds is challenging, even at the population level
- **Solution:** when outcomes and covariates are discrete, I show that optimization over $\mathcal{C}(\rho^L)$ can be represented as a variant of an optimal transportation problem (see Villani (2009))
- Discrete optimal transportation problems can be solved by linear programming (Boyd and Vandenberghe, 2004) Details
- I then proceed with estimation using sample analogs

Estimation

- $\mathcal{C}(\rho^L)$ is an infinite-dimensional space and, unlike in the example, the maximum and minimum causal effects do not occur at the boundaries in general
- Computing the bounds is challenging, even at the population level
- **Solution:** when outcomes and covariates are discrete, I show that optimization over $\mathcal{C}(\rho^L)$ can be represented as a variant of an optimal transportation problem (see Villani (2009))
- Discrete optimal transportation problems can be solved by linear programming (Boyd and Vandenberghe, 2004) Details
- I then proceed with estimation using sample analogs

Assumption.

Finite support of the potential outcomes and covariates: (i) Finite support of Y: Y takes finite values in $\mathcal{Y} = \{y_1, \ldots, y_j, \ldots, y_J\}$. (ii) Finite support of X: X takes finite values in a finite set $\mathcal{X} = \{x_1, \ldots, x_l, \ldots, x_L\}$.

Inference

- I require an additional assumption for consistent estimation, inference

Assumption.

Estimation and inference. *(i) Sampling.* $Z_i = (Y_{0i}, Y_{1i}, X_i, D_i)$ *for* $i = 1, \ldots, N$ *are i.i.d. (ii) Differentiability of the linear programming representations.* $\phi(\cdot; \rho^L)$ *is differentiable at p. (iii) Finite bounds.* $\tau^U(\rho^L) - \tau^L(\rho^L) < \infty$. *(iv) Positive, finite asymptotic variance of* $\sqrt{N}(\phi(\hat{p}; \rho^L) - \phi(p; \rho^L))$.
$[\underline{\sigma}^2, \underline{\sigma}^2]' \leq \nabla\phi(p; \rho^L)\Sigma\nabla\phi(p; \rho^L)' \leq [\bar{\sigma}^2, \bar{\sigma}^2]$ *where* $\underline{\sigma}^2$ *and* $\bar{\sigma}^2$ *are positive and finite and*

$$\Sigma = \begin{bmatrix} p_{110a}(1 - p_{110a}) & -p_{110a}p_{210a} & \cdots \\ -p_{110a}p_{210a} & p_{210a}(1 - p_{210a}) & \\ \vdots & & \ddots \end{bmatrix}.$$

- $p_{jltd} = P(Y = y_j, X = x_l, T = t, D = d)$
- $\hat{p}_{jltd} = \frac{1}{N}\sum_{i=1}^{N} 1\{Y_i = y_j, X_i = x_l, T_i = t, D_i = d\}$

Inference

- Consistency follows immediately from the consistency of p_{jltd} for $\hat{p}_{jltd}$ and the continuous mapping theorem

Proposition 3.

Let $\mathcal{P}$ be the set of distributions for which **Assumption: Estimation and Inference** *holds. Then,* $\lim_{N\to\infty} \inf_{P\in\mathcal{P},\tau(\rho^L)\in[\phi^L(\rho^L),\phi^U(\rho^L)]} P(\tau(\rho^L) \in CI_\alpha(\rho^L)) = 1-\alpha.$

$CI_\alpha(\rho^L)$

1. Generate $\{Z_1^*, \ldots, Z_N^*\}$ from $\hat{p}$.
2. Compute $\hat{p}^*$ by applying the frequency estimator to $\{Z_1^*, \ldots, Z_N^*\}$
3. Compute $[\tau^L(\rho^L)^*, \tau^U(\rho^L)^*]' = \phi(\hat{p}^*; \rho^L)$.
4. Repeat steps 1-3 B times. Compute $\hat{\sigma}^L = \sqrt{N} \times SD(\tau^L(\rho^L)^*)$, $\hat{\sigma}^U = \sqrt{N} \times SD(\tau^U(\rho^L)^*)$ and $\hat{\varrho} = Cor(\tau^L(\rho^L)^*, \tau^U(\rho^L)^*)$.
5. Form the $(1-\alpha)$-percent confidence interval for $\tau(\rho^L)$ as

$$CI_\alpha(\rho^L) = \left[\phi^L(\hat{p}; \rho^L) - \frac{\hat{\sigma}^L c^L}{\sqrt{N}}, \phi^U(\hat{p}; \rho^L) - \frac{\hat{\sigma}^U c^U}{\sqrt{N}}\right]$$

where $[c^L, c^U]$ are as described in Stoye (2009)

Outline

Empirical results

- I contrast the results using the bounds with Hotz et al. (2005)'s two step procedure
- Two empirical settings:
  1. Transfers to Mexican microenterprises: Hotz et al. (2005) is overconfident
  2. Remedial education in India: Hotz et al. (2005) is overly conservative

Transfers to Mexican microenterprises

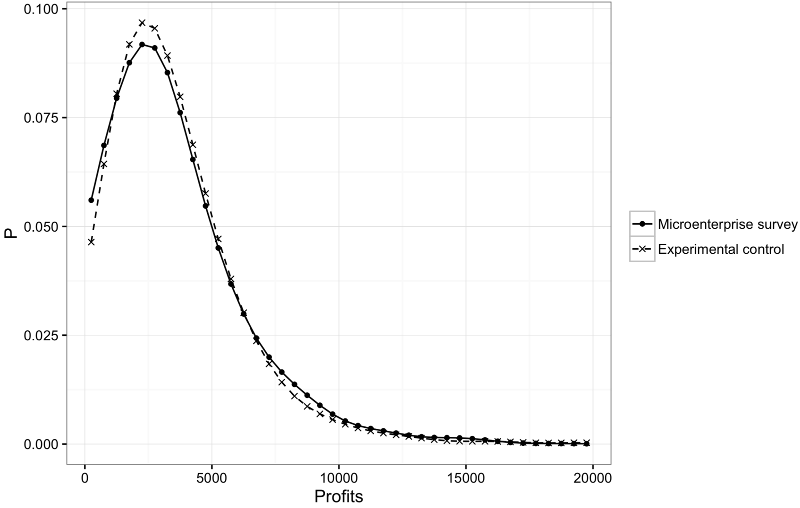
(McKenzie and Woodruff, 2008)

- Experiment carried out in 2006 (baseline Oct. 2005) in Leon
- The outcome of interest is monthly profits
- Treatment was a 1,500 ($\approx$ 140 USD) peso transfer (50% "in-kind")
- The average effect in Leon $\approx$ 600 pesos
- Sample:
  - 22-55 year old male entrepreneurs
  - Working in retail
  - Capital stock $\leq$ 10,000 pesos
  - No paid employees
  - Working 35+ hours per week in microenterprise
  - Urban
- 207 entrepreneurs - very small

The alternative population of interest

- The experimental questionnaire was based on the national microenterprise survey: Encuesta Nacional de Micronegocios (ENAMIN)
- ENAMIN is also available in 2012
- What does the experiment tell us about returns to cash transfers for microentrepreneurs in urban Mexico in 2012?
- I select a sample using the same criteria from the 2012 ENAMIN
    - 907 entrepreneurs

Untreated profits



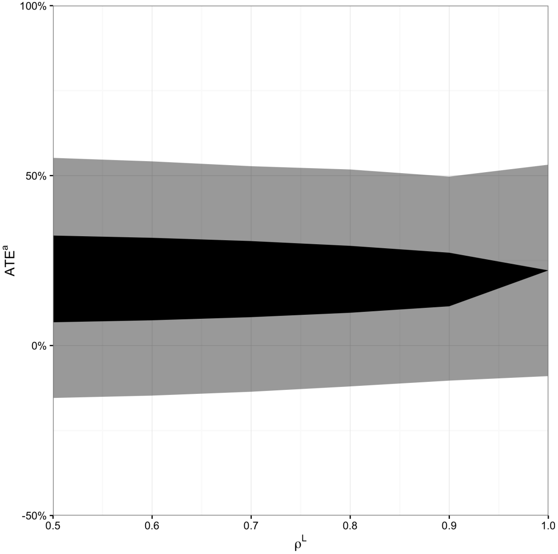
- Heaping is an issue in reported profits, smoothed out using a kernel density estimator before discretizing in 500 peso bins

Implications

Hotz et al. (2005) method

- Some differences in the distributions of untreated profits, particularly at the low end
- The small experiment means we cannot reject equal profit distributions (p-value: 0.92)
- We then apply the mean treated outcome and confidence interval to all urban Mexico
- Counterintuitively, we are **more** confident in our prediction for all Mexico than our conclusion from the experiment

Bounds on the return to transfers



- block bootstrap at the firm level

Transfers to Mexican microenterprises

Summing up

- The representativeness of the experimental site is promising, but we should be cautious in extrapolating due to the sample size

Remedial education in India

(Banerjee et al., 2007)

- Treatment is assignment of a remedial education teacher to a student's class
  - The remedial education teacher works with 15-20 students identified as falling behind for about half the school day
- Two experiments conducted over three years (2001-2003)
  - Mumbai
  - Vadodara
- I can use the results from one city to try predicting the average effect in the other

Implementation

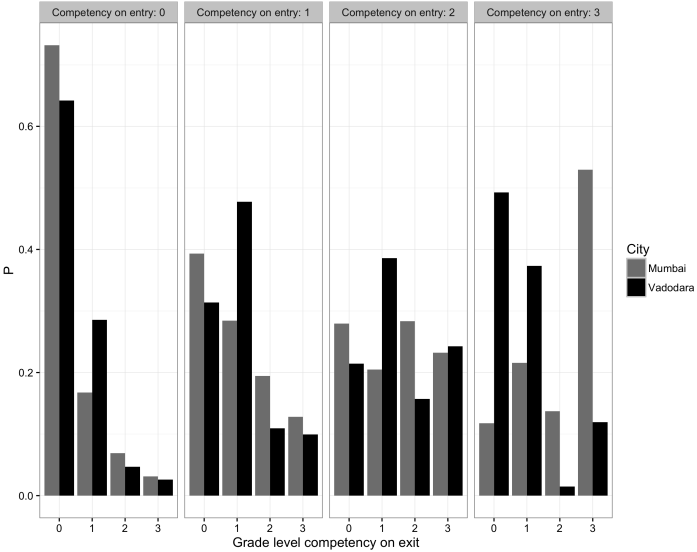
- Outcome of interest: math proficiency
- Researchers administered different math tests in the two cities
- I therefore measure the outcome of interest with the student's maximum grade-level competency in math
- Year 2 in Mumbai featured compliance problems: 1/3 of schools refused the remedial education teachers
- I want to be plausibly looking at the effect of the same policy, so
    - Location *e*: Vadodara (both years, 3rd graders)
    - Location *a*: Mumbai (year 1)

Summary statistics

	(1)	(2)
	Vadodara	Mumbai
Pre-test: maximum competency	0.29	0.54
	(0.57)	(0.79)
Male	0.29	0.47
	(0.45)	(0.50)
Number of students in class	63.94	89.51
	(27.81)	(40.23)
Observations	10049	4429

Conditional distributions of grade level competency

Control groups



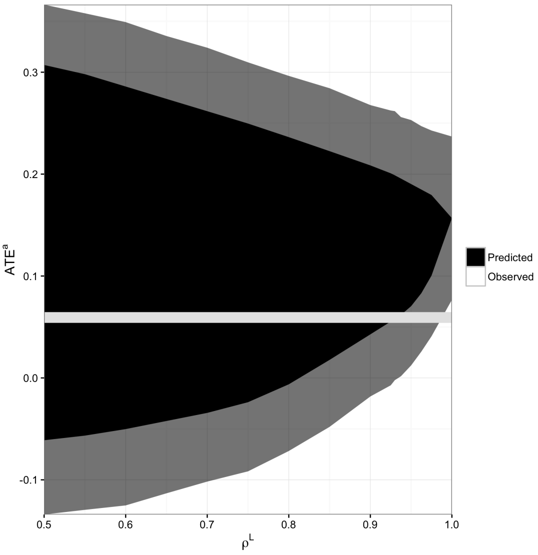
Conditional distributions of grade level competency

Control groups

- We reject equality of the conditional distributions of grade level competency for the control groups
- In the Hotz et al. (2005) framework, we conclude that we cannot extrapolate

Bounds on the increase in average grade level competency

Using Vadodara to predict Mumbai (block bootstrap at the classroom-year level)



Summing up

Summing up

- I can reject a zero ATE for Mumbai for ρ^L exceeding .937 of the maximum
- This is enough for the bounds to include the point estimate of the ATE in Mumbai

Conclusions

- This paper develops a framework for quantifying the uncertainty associated with generalizing the results from a randomized experiment to a specific target population
  - The strength of assumptions required to draw specific conclusions about the average effect in the target population
- Conceptually, Imbens and Wooldridge (2009) and others claim that failure to point-identify the distribution of treatment effects is not policy-relevant
- I argue that transporting reduced form treatment effects across contexts makes this first-order

Thank you!

Definition of copula

A copula function $C : [0,1]^2 \to [0,1]$ satisfies:

1. Boundary conditions:
   1. $C(0,v) = C(u,0) = 0 \; \forall \; u,v \in [0,1]$
   2. $C(u,1) = u$ and $C(1,v) = v \; \forall \; u,v \in [0,1]$
2. Monotonicity condition:
   1. $C(u,v) + C(u',v') - C(u,v') - C(u',v) \geq 0 \; \forall \; u,v,u',v'$ s.t. $u \leq u', v \leq v'$

Back

For any two random variables V and W:

$$\rho(V, W) = \frac{Cov_C(R(V), R(W))}{Cov_M(R(V), R(W))}$$

where $R(V) = F_V(v)$ when V is continuously distributed and $R(V) = \frac{F_V(v)+F_V(v-)}{2}$ when V takes a finite number of values. $F_V(v-)$ denotes $P(V < v)$. $Cov_C(R(V), R(W))$ denotes the covariance between $R(V)$ and $R(W)$ under copula C and $Cov_M(R(V), R(W))$ the maximum covariance between $R(V)$ and $R(W)$.

Back

Linear programming representation

Objective function (conditioning on $X = x$ implicit)

$$\tau^U(\rho^L) = \max_{\mathcal{C}(\rho^L)} \; E^a[Y_1 - Y_0]$$

$$\phi^U\left(\left\{P^e(y_{0j})\right\}_{j=1,\dots,J}, \left\{P^e(y_{1k})\right\}_{k=1,\dots,K}, \left\{P^a(y_{0j})\right\}_{j=1,\dots,J}; \rho^L\right)$$

$$= \max_{\{P^e(y_{0j},y_{1k})\}_{j=1,\dots,J}^{k=1,\dots,K}} \sum_{j=1}^{J}\sum_{k=1}^{K} y_{1k}\frac{P^a(y_{0j})}{P^e(y_{0j})} \times P^e(y_{0j}, y_{1k}) - \underbrace{\sum_{j=1}^{J} y_{0j}P^a(y_{0j})}_{normalization}$$

- Maximization is over the elements of the joint distribution of the potential outcomes

Linear programming representation

Constraints: consistency with experimental results

$$\sum_{k=1}^{K} P^e(y_{0j}, y_{1k}) = P^e(y_{0j}) \ \forall j \in \{1, ..., J\}$$

$$\sum_{j=1}^{J} P^e(y_{0j}, y_{1k}) = P^e(y_{1k}) \ \forall k \in \{1, ..., K\}$$

Linear programming representation

Constraint: dependence

$$
\begin{aligned}
&\rho^L + 4\sum_{j=1}^{J}\sum_{k=1}^{J}\pi_{jk}\left(\frac{P^e(Y \leq y_j|T=0) + P^e(Y \leq y_{j-1}|T=0) - 1}{2}\right) \\
&\qquad \times \left(\frac{P^e(Y \leq y_j|T=1) + P^e(Y \leq y_{j-1}|T=1) - 1}{2}\right) \\
&\geq \rho^L \sum_{j=1}^{J}\sum_{k=1}^{J}\Bigg[P^e(y_j|T=0)P^e(y_k|T=1) \\
&\qquad \times \Big(\min\left\{P^e(Y \leq y_j|T=0), P^e(Y \leq y_j|T=1)\right\} \\
&\qquad + \min\left\{P^e(Y < y_j|T=0), P^e(Y < y_j|T=1)\right\} \\
&\qquad + \min\left\{P^e(Y < y_j|T=0), P^e(Y \leq y_j|T=1)\right\} \\
&\qquad + \min\left\{P^e(Y \leq y_j|T=0), P^e(Y < y_j|T=1)\right\}\Big)\Bigg]
\end{aligned}
$$

Linear programming representation

Expanding the set of possible $F^a_{Y_1|Y_0}(y_1|y_0)$

$$\max_{\substack{\pi^e \in [0,1]^{J^2},\\ \pi^a \in [0,1]^{J^2}}} \sum_{j=1}^{J}\sum_{k=1}^{J} y_{1k}\pi^a_{k|j}P^a(y_j) - \sum_{j=1}^{J} y_{0j}P^a(y_j)$$

subject to

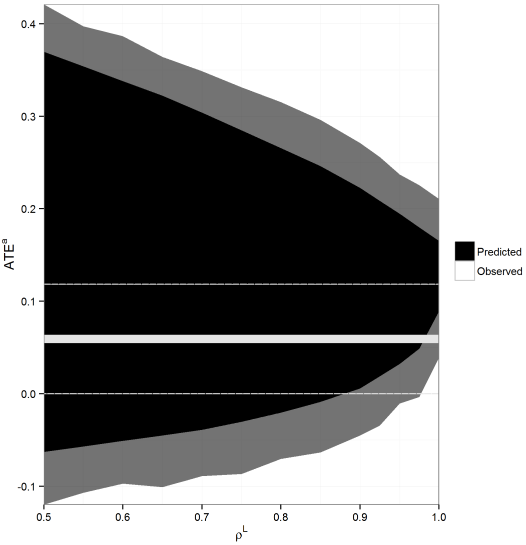
All previous constraints and...

$$\kappa \geq \sum_{k=1}^{J} \pi^a_{k|j} \log\left(\frac{\pi^a_{k|j}}{\left(\frac{\pi^e_{jk}}{P^e(y_j|T=0)}\right)}\right) \forall j \in \{1,\ldots,J\}$$

Back

Bounds on the increase in average grade level competency

Using Vadodara to predict Mumbai



Using Vadodara to predict Mumbai

Summing up

- Lack of students with grade level competency 3 on entering third grade: a violation of one common support assumption
- I make a very conservative assumption: with remedial education, these students could have gotten the worst outcome (competency 0) or the best (competency 3) Back

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *Quarterly Journal of Economics 130*(3), 1117–1165.

Altonji, J., T. Conley, T. Elder, and C. Taber (2013). Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables. *Mimeo*.

Altonji, J., T. Elder, and C. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy 113*(1), 151–184.

Angrist, J. and I. Fernández-Val (2013). ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. In *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress, Volume III: Econometrics*. Econometric Society Monographs.

Angrist, J. and M. Rokkanen (2015). Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff. *Journal of the American Statistical Association 110*(512), 1331–1344.

Athey, S. and G. Imbens (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica 74*(2), 431–497.

Attanasio, O., C. Meghir, and M. Szekely (2003). Using Randomised Experiments and Structural Models for 'Scaling Up': Evidence from the PROGRESA Evaluation. *Mimeo*.

Banerjee, A., S. Chassang, and E. Snowberg (2016). Decision Theoretic Approaches to Experiment Design and External Validity. In *Handbook of Field Experiments, forthcoming*.

Banerjee, A., S. Cole, E. Duflo, and L. Linden (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics 122*(3), 1235–1264.

Banerjee, A. and R. He (2008). Making aid work. In W. Easterly (Ed.), *Reinventing Foreign Aid*, Chapter 2.

Banerjee, A. V. and E. Duflo (2009). The experimental approach to development economics. *Annual Review of Economics* (1), 151–78.

Bisbee, J., R. Dehejia, C. Pop-Eleches, and C. Samii (2015). LOCAL INSTRUMENTS, GLOBAL EXTRAPOLATION: EXTERNAL VALIDITY OF THE LABOR SUPPLY-FERTILITY LOCAL AVERAGE TREATMENT EFFECT. *Mimeo 10012*(2014).

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.

Brinch, C., M. Mogstad, and M. Wiswall (2015). Beyond LATE with a discrete instrument. *Journal of Political Economy, forthcoming*.

Chassang, S., G. Padro i Miquel, and E. Snowberg (2012). Selective Trials : A Principal-Agent Approach to Randomized Controlled Experiments. *American Economic Review*.

Cole, S. and E. Stuart (2010). Generalizing Evidence from Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *American Journal of Epidemiology 172*(1), 107–15.

Dehejia, R., C. Pop-Eleches, and C. Samii (2017). From Local to Global: External Validity in a Fertility Natural Experiment. *NBER Working Paper 21459*.

Duflo, E. (2004). Scaling up and evaluation. In *Annual World Bank Conference on Development Economics*, pp. 341–369.

Frechet, M. (1951). Sur Les Tableux de Correlation Dont Les Marges Sont Données. *Ann. University: Lyon 3*(14), 53–77.

Hoeffding, W. (1940). Masstabinvariate Korrelationstheorie. *Schriften des Mathematischen Instituts und des Institutes für Angewandte Mathematik der Universität Berlin*, 179–251.

Hotz, V. J., G. Imbens, and J. Mortimer (2005). Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations. *Journal of Econometrics 125*, 241–270.

Imbens, G. W. G. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Kline, P. and A. Santos (2013). Sensitivity to Missing Data Assumptions: Theory and an Evaluation of the US Wage Structure. *Quantitative Economics 4*(2013), 231–267.

Kowalski, A. E. (2016). Doing More When You're Running LATE: Applying Marginal Treatment Effect Methods to Experiments. *Working Paper*.

Matzkin, R. L. (2013, jan). Nonparametric Identification in Structural Economic Models. *Annual Review of Economics 5*(1), 457–486.

McKenzie, D. and C. Woodruff (2008). Experimental Evidence on Returns to Capital and Access to Finance in Mexico. *The World Bank Economic Review 22*(3), 457–482.

Mogstad, M., A. Santos, and A. Torgovitsky (2016). Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters. *Working Paper*.

Pearl, J. and E. Bareinboim (2014). External Validity: From do-calculus to Transportability across Populations. *Statistical Science 29*(4), 579–595.

Pritchett, L. and J. Sandefur (2013). Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix. *Journal of Globalization and Development 4*(2), 161–197.

Stoye, J. (2009). More on Confidence Intervals for Partially Identified Parameters. *Econometrica 77*(4), 1299–1315.

Stuart, E., S. Cole, C. Bradshaw, and P. Leaf (2011). The Use of Propensity Scores to Assess the Generalizability of Results from Randomized Trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 174*(2), 369–386.

Villani, C. (2009). *Optimal Transport: Old and New*. Springer.

Vivalt, E. (2017). How Much Can We Generalize From Impact Evaluations? *Mimeo*.