Lecture 23:
Generalization and External Validity I

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

May 4, 2022

Introduction

We have been studying ways to generate "specific causal facts."

Samii (2016, *JOP*) explains why such specificity is unavoidable in empirical research.

But, the scientific endeavor asks that we generalize and interpret.

Two ways to think about generalization and interpretation:

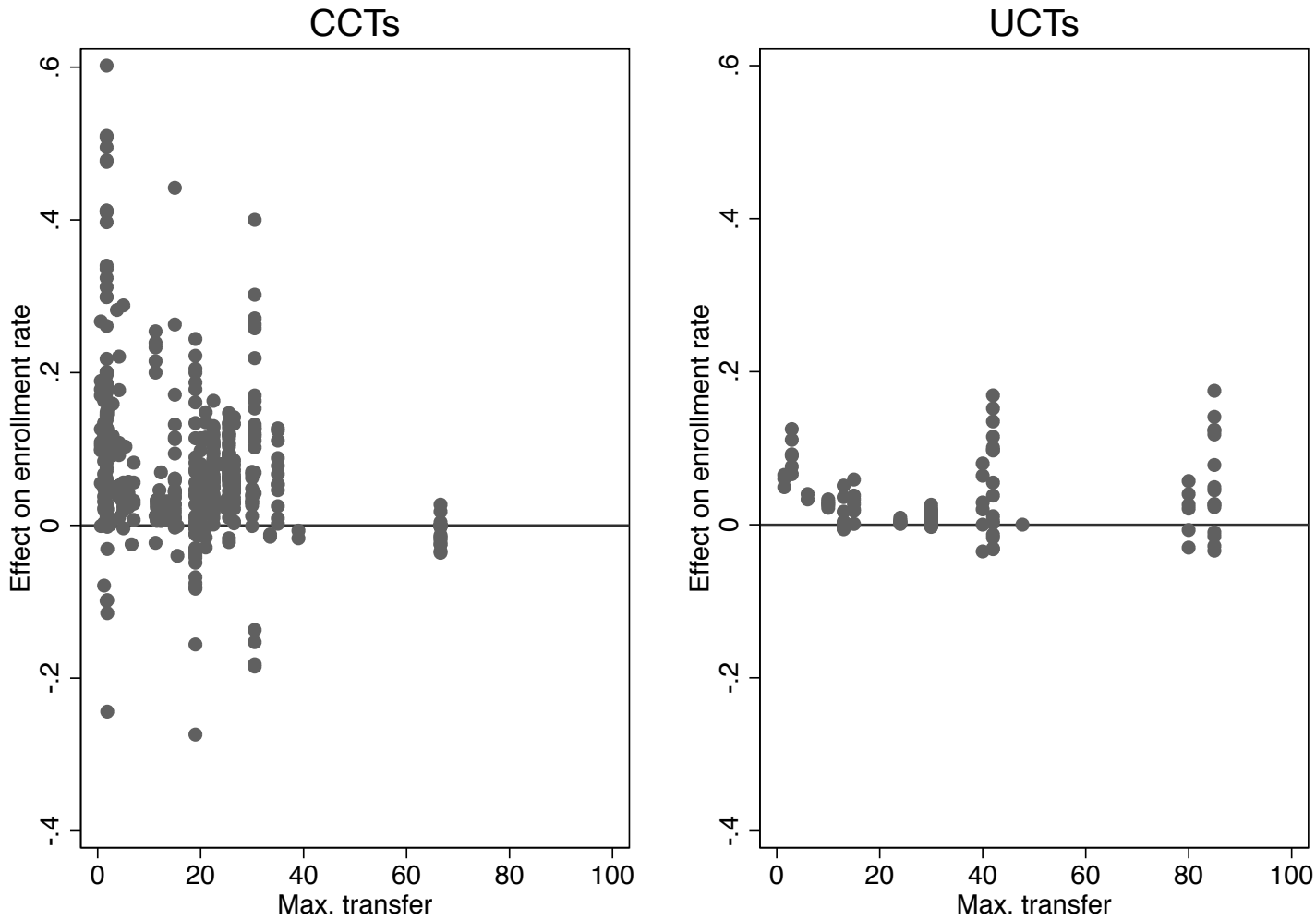
- Bare-foot empirical approach.
- Through the lens of theory.

Today we will focus on the first, next time on the second.

Introduction

- Causal evidence is accumulating. Where is it taking us?
- Can we achieve "external validity" goals?
- By what standard do we judge a *research program* (rather than single study) successful?

Effect heterogeneity



(Data from Vivalt, 2019)

Defining external validity goals

- In evaluating external validity for a causal effect, one asks, *to what population, settings, and variables can this effect be generalized?*

  (Campbell, 1957)
- "Do these results from India tell me anything about Nigeria?"
- Success of research program $\propto$ diversity of settings in which one can reliably predict effects.
- Successful research programs obviate need for further experimentation with a treatment (at least temporarily).

Approaches

- Meta-analysis and synthesis.
- Extrapolation to new contexts.

Meta-analysis and synthesis

- Sometimes a collection of studies can speak more clearly than any one about *a given* effect: *effect synthesis*.
- Sometimes effect heterogeneity is of interest: meta-regression, decomposition, etc. (cf. Angrist et al. paper from a few weeks ago).

Effect Synthesis

- Classical set-up:
- Set of study populations $s = 1, \ldots, S$ with estimates effects,

$$\hat{\rho}_s = \rho_s + \varepsilon_s.$$

- Classically, assume $\varepsilon_s \sim \mathrm{N}(0, v_s^2)$.

Effect Synthesis

- Classical set-up:
- Set of study populations $s = 1, \ldots, S$ with estimates effects,

$$\hat{\rho}_s = \rho_s + \varepsilon_s.$$

- Classically, assume $\varepsilon_s \sim \mathrm{N}(0, v_s^2)$.
- "Random effects" model further assumes:

$$\rho_s = \mu + u_s.$$

with $u_s \sim \mathrm{N}(0, \tau^2)$.

Effect Synthesis

- Classical set-up:
- Set of study populations $s = 1, \ldots, S$ with estimates effects,

$$\hat{\rho}_s = \rho_s + \varepsilon_s.$$

- Classically, assume $\varepsilon_s \sim \mathrm{N}(0, v_s^2)$.
- "Random effects" model further assumes:

$$\rho_s = \mu + u_s.$$

  with $u_s \sim \mathrm{N}(0, \tau^2)$.
- Could extend to a mixed effects model that brings in covariates,

$$\rho_s = x_s' \beta + u_s.$$

Effect Synthesis

- Classical set-up:
- Set of study populations $s = 1, \ldots, S$ with estimates effects,

$$\hat{\rho}_s = \rho_s + \varepsilon_s.$$

- Classically, assume $\varepsilon_s \sim \mathrm{N}(0, v_s^2)$.
- "Random effects" model further assumes:

$$\rho_s = \mu + u_s.$$

  with $u_s \sim \mathrm{N}(0, \tau^2)$.
- Could extend to a mixed effects model that brings in covariates,

$$\rho_s = x_s'\beta + u_s.$$

- ε_s is thus the within study variance, and so v_s is the standard error, and then u_s is between-population heterogeneity, with τ the standard deviation of the population effect distribution.

Synthesized Effect Estimators

- "Fixed effects" analysis characterizes $\bar{\rho}$ based on the $\hat{\rho}_s$ estimates.
- MMSE estimator is,

$$\hat{\bar{\rho}} = \frac{\sum_s \frac{\hat{\rho}_s}{\hat{v}_s^2}}{\sum_s \frac{1}{\hat{v}_s^2}}$$

Synthesized Effect Estimators

- "Fixed effects" analysis characterizes $\bar{\rho}$ based on the $\hat{\rho}_s$ estimates.
- MMSE estimator is,

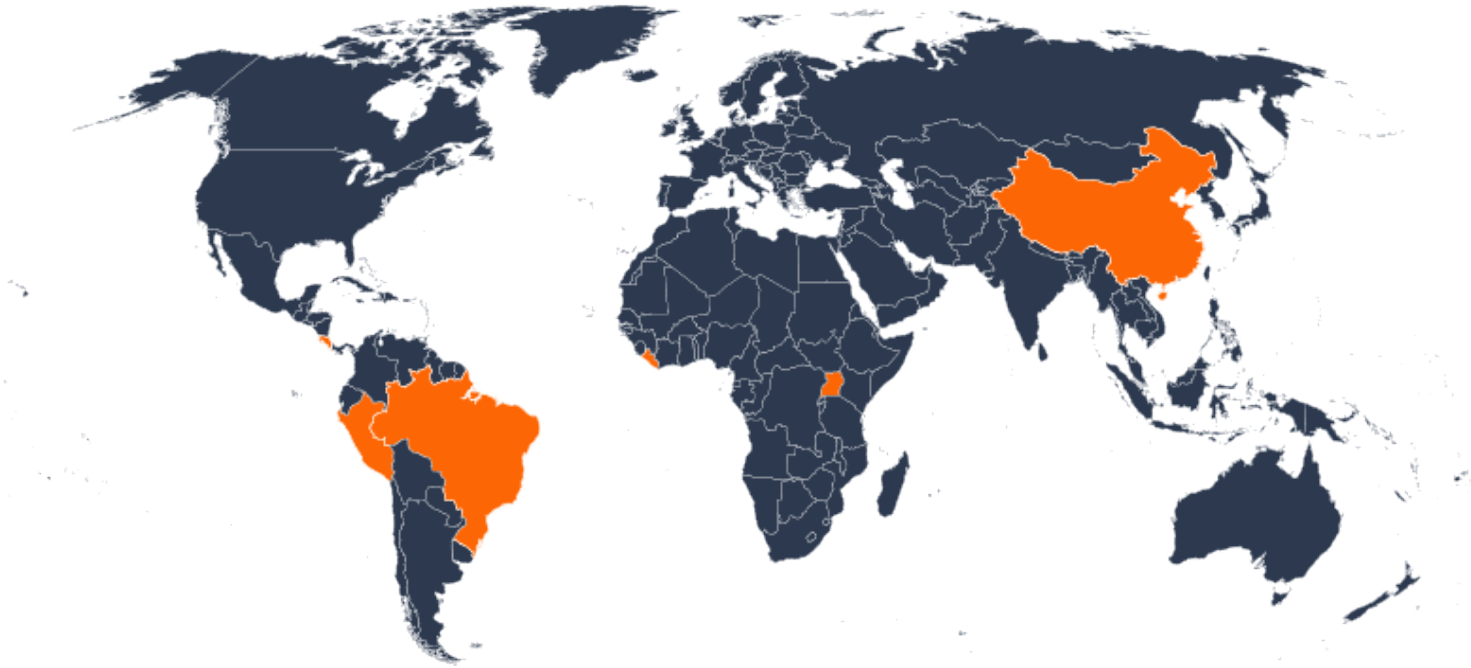
$$\hat{\bar{\rho}} = \frac{\sum_s \frac{\hat{\rho}_s}{\hat{v}_s^2}}{\sum_s \frac{1}{\hat{v}_s^2}}$$

- "Random effects" analysis characterizes the effect distribution parameterized by μ and τ.
- MMSE estimator is,

$$\hat{\mu} = \frac{\sum_s \frac{\hat{\rho}_s}{\hat{v}_s^2 + \hat{\tau}^2}}{\sum_s \frac{1}{\hat{v}_s^2 + \hat{\tau}^2}}$$

- Estimation requires iterative fitting to get $\hat{\tau}^2$ and then $\hat{\mu}$.
- Meta regressions follow similar weighting schemes for fixed-effects and mixed-(random-)effects

Illustration



- EGAP “Metaketa III” is a coordinated set of 6 RCTs on natural resource governance.
- Intervention protocols and outcome measurement coordinated in registered pre-analysis plans.
- Sought to address: can improving community monitoring capacities allow communities to more sustainability manage common pool resources?

Illustration

Table 1. Features of the research contexts and experimental designs

	Brazil	China	Costa Rica	Liberia	Peru	Uganda
Contextual features of CPR						
Resource	Groundwater	Surface water	Groundwater	Forest	Forest	Forest
Community	Rural villages	Urban microneighborhoods	Rural villages	Villages	Indigenous communities	Villages
Primary threat to resource	Drought, overuse	Individual, industrial pollution	Drought, overuse	Overcutting by residents	Extraction by outsiders	Overcutting by residents
Components of harmonized interventions						
Community workshops	✓	—	✓	✓	✓	✓
Monitor selection, training, incentives	✓	✓	✓	✓	✓	✓
Monitoring of the resource	✓	✓	✓	✓	✓	✓
Dissemination to citizens	✓	✓	✓	✓	✓	✓
Dissemination to management bodies	—	(Alternate arm)	✓	✓*	✓*	✓*
Experimental design						
Alternate treatment arm	Conservation plan making	Dissemination to government	—	Negotiation training	—	SMS reminders
Experimental design	Three-arm†	2×2 factorial	Two-arm	2×2 factorial	Two-arm	Three-arm†
No. of monitoring communities (N_M)	80	80	81	60	39	60
No. of nonmonitoring communities ($N_{\neg M}$)	40	80	80	60	37	50
Common outcome measurement						
Duration of implementation, mo	12	15	12	12	13	12
Primary compliance measure	SMS reports received	Dissemination posters	Reports submitted	Monitoring walks completed	Reports submitted	Reports submitted
Primary resource outcome	Well electricity usage	Pollutant concentration in water	Well electricity usage, water quality	Deforestation	Deforestation	Deforestation, forest quality
Endline citizen survey	✓	✓	✓	✓	✓	✓

N_M denotes the number of communities assigned to any treatment condition with community monitoring, and $N_{\neg M}$ denotes the number assigned to any treatment condition without community monitoring.

*In the forest studies, the community constitutes at least one of the possibly overlapping management bodies.

†In both three-arm designs, communities assigned to the alternative treatment arm received both monitoring and the alternative treatment.

Illustration

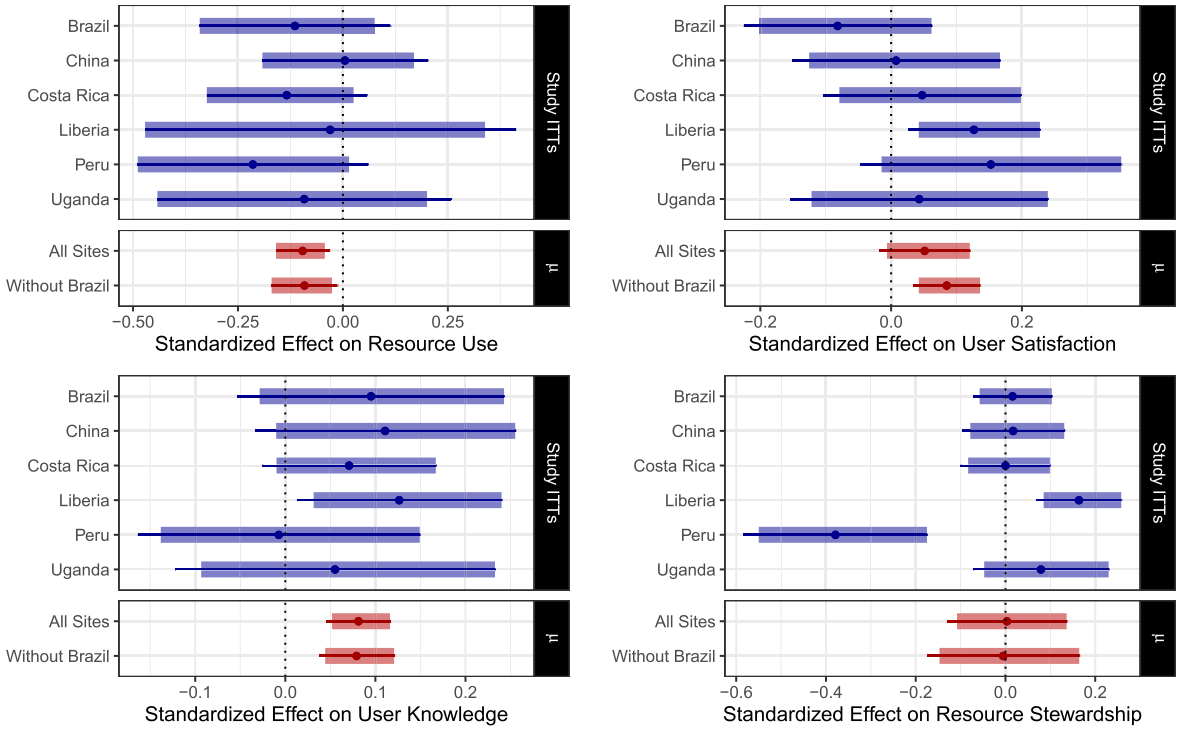


Fig. 2. Estimated site-level ITT effects (top six estimates) and mean ITT effects (bottom two estimates) (μ) across sites for each of our main hypotheses. The thin segments represent 95% CIs. The thick segments indicate the direction of the prespecified one-tailed hypotheses; where these segments do not bound zero, we reject null hypotheses at the $\alpha = 0.05$ level.

Extrapolation

Identification

- Conditions for valid extrapolation to new populations analogous to conditional independence assumptions for identifying causal effects.
- Contexts, $c = 1, ..., C+1$, with $c = 1$ "target."
- Treatment: $T = 0, 1$.
- Potential outcomes: $(Y(1), Y(0))$.
- Observed outcomes: $Y = TY(1) + (1-T)Y(0)$.
- Indicator $D = 0, 1$ for selection of target context, $D = 1$, in which case for reference contexts, $D = 0$.
- Macro (V) and micro (W) covariates.

Identification

Hotz, Imbens, and Mortimer (2005) define identifying conditions:

- C0. Effects are identified in the reference sample:

$$T \perp\!\!\!\perp (Y(1), Y(0)) | (V, W), D = 0.$$

- C1. "Unconfounded location":

$$D \perp\!\!\!\perp (Y(1), Y(0)) | (V, W).$$

- C2. Common support across locations:

$$\delta < \Pr[D = 0 | V = v, W = w] < 1 - \delta$$

for some $\delta > 0$ and for all v, w in the support of V, W.

Identification

Given C0-C2, our target quantity is

$$\begin{aligned}\tau_1 &\equiv \mathrm{E}\left[Y(1)-Y(0)|D=1\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[Y(1)-Y(0)|V,W,D=1\right]|D=1\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[Y(1)-Y(0)|V,W,D=0\right]|D=1\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[Y|T=1,V,W,D=0\right]|D=1\right] \\ &\qquad -\mathrm{E}\left[\mathrm{E}\left[Y|T=0,V,W,D=0\right]|D=1\right],\end{aligned}$$

where the last quantity is identified.

Identification

If unconfounded location holds, it motivates the following:

- Differences in ATEs across locations or populations are due to differences in covariates.
- To evaluate generalizability to such locations/populations, use the following steps:
  - Figure out which covariates moderate your treatment effects.
  - Look at how the distribution of these covariates varies across sites.
- See work by Egami and Hartman (2020) for more.

Partial identification (bounds)

Gechter (2016) bounds:

- Like CIA, we may worry about unconfounded location.
- Unconfounded location implies that conditional $Y(0)$ distributions match in both reference and target locations:

$$D \perp\!\!\!\perp Y(0)|X.$$

- Often you have the $Y(0)$ distribution in the target setting, in which case this is testable.

Partial identification (bounds)

- Moreover, consider the following decomposition of the *treated* potential outcome means in the target location:

$$\mathrm{E}_{D=1}[Y(1)] = \int_{\mathscr{X}} \left(\int_{\mathbb{R}} \left[\int_{\mathbb{R}} y(1) \underbrace{dF_{Y(1)|Y(0),X,D=1}(y(1)|y(0),x)}_{A} \right] \underbrace{dF_{Y(0)|X,D=1}(y(0)|x)}_{B} \right) \underbrace{dF_{X|D=1}(x)}_{C}$$

- B and C are observable.
- A is not. It tells us how the $Y(1)$s vary in $Y(0)$ and X.
- What do we know about A? Generally speaking, not much.
- But suppose the following "constant copula" condition holds (Gechter, Assn. 3):

$$F_{Y(1)|Y(0),X,D=1}(y(1)|y(0),x) = C_1\left(F_{Y(0)|X,D=0}(y(0)|x), F_{Y(1)|X,D=0}(y(1)|x)\right),$$

for some copula C yielding the conditional distribution C_1.

- Go through all copulas satisfying this property. Then, implied maximum and minimum for $E_{D=1}[Y(1)]$ yield our extrapolation bounds.

Extrapolate or Experiment?

Dehejia, Pop-Eleches, and Samii (2019) decision problem:

Extrapolate or Experiment?

Dehejia, Pop-Eleches, and Samii (2019) decision problem:

- Suppose a policy maker wants to make an evidence-based decision about whether to implement a program.
- Can either use existing evidence base or run a new experiment.

Extrapolate or Experiment?

Dehejia, Pop-Eleches, and Samii (2019) decision problem:

- Suppose a policy maker wants to make an evidence-based decision about whether to implement a program.
- Can either use existing evidence base or run a new experiment.
- Suppose program is worth implementing if its effects exceed some critical threshold, c^*.
- Decision reduces to something like an hypothesis test: does evidence base provide a reliable enough estimate relative to c^*?

Extrapolate or Experiment?

- Target context, which we will call S_1, has covariate distribution, $C(S_1)$.
- Policy maker decides existing evidence is adequate if 95% prediction interval excludes c^*.
- (This is arbitrary. The "right" interval would depend on costs/benefits of program.)
- Existing evidence allows us to estimate effect in target context: $\hat{\tau}_1$ where error of prediction is ζ_1.

Extrapolate or Experiment?

Context	Covariates	τ_c	$= \hat{\tau}_c - \zeta_c$
0	$C(S_1)$	?	$= \hat{\tau}_1 - ?$
1	$C(S_2)$	τ_1	$= \hat{\tau}_2 - \zeta_2$
$\vdots$	$\vdots$	$\vdots$	
$C+1$	$C(S_{C+1})$	τ_{C+1}	$= \hat{\tau}_{C+1} - \zeta_{C+1}$

Extrapolate or Experiment?

- Prediction error:

$$\zeta_1 = \hat{\tau}_1 - \tau_1 .$$

- Distribution is governed by variances of $\hat{\tau}_1$ and τ_1.
- Variance of first is *estimation* variability.
  - Shrinks in amount of data from reference contexts.
- Variance of second is *intrinsic* variability of effects at $C(S_1)$ in the covariate distribution space.
  - Shrinks in extent to which covariates explain outcome heterogeneity.

Extrapolate or Experiment?

Given $\hat{\tau}_1$, a 95% prediction interval for τ_1 is,

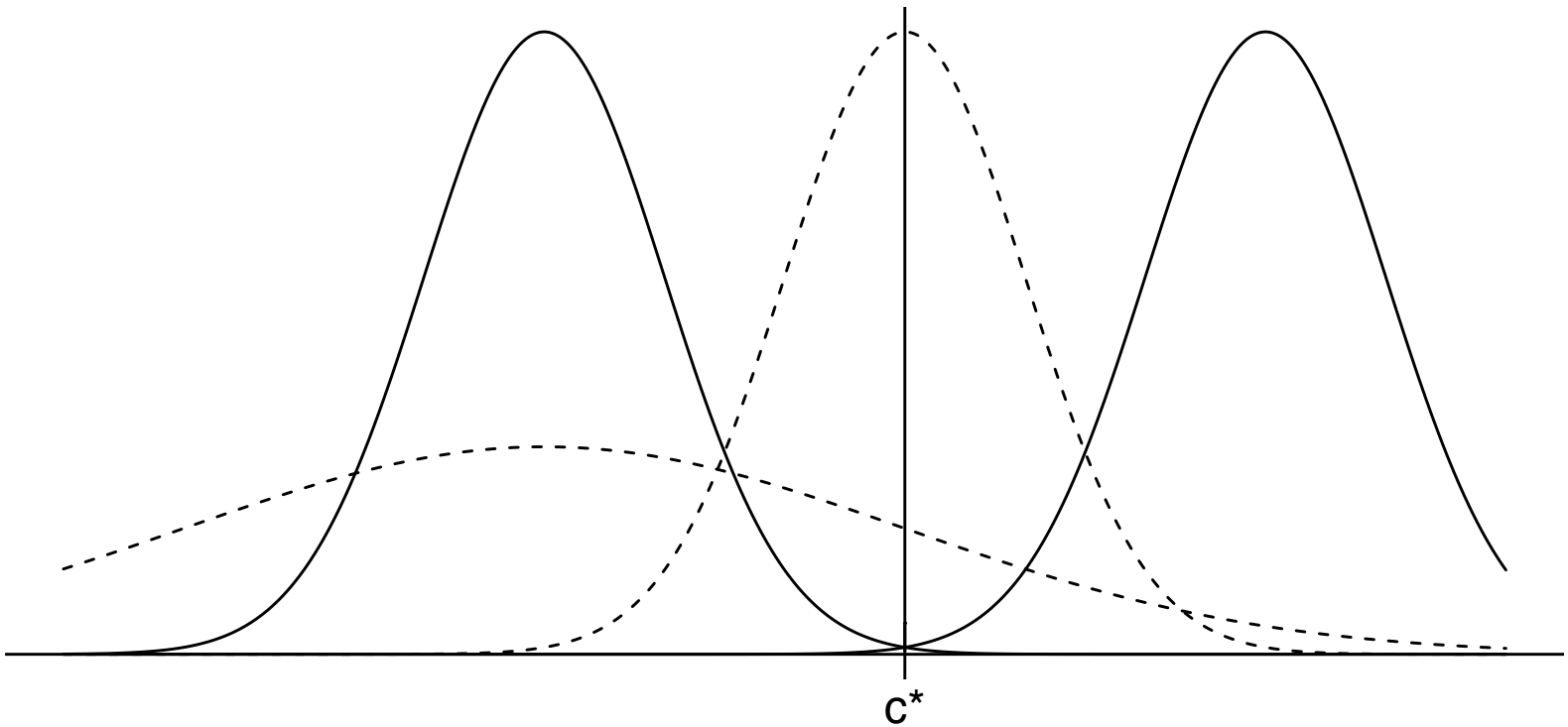
$$PI_1 = [\hat{\tau}_1 - t_{.025}\sqrt{\text{Var}\left[\zeta_c|C(S_1)\right]}, \hat{\tau}_1 + t_{.025}\sqrt{\text{Var}\left[\zeta_c|C(S_1)\right]}],$$

where we use normal quantiles for $t_{(.)}$. Thus the solution to the decision problem is:

- experiment if $c^* \in PI_1$, and
- accept verdict of existing evidence otherwise.

Extrapolate or Experiment?

Figure 23: To experiment or extrapolate? A graphical illustration of the decision problem



Notes: Solid line = experiment not warranted. Dashed line = experiment warranted.

Decision problem

Estimation:

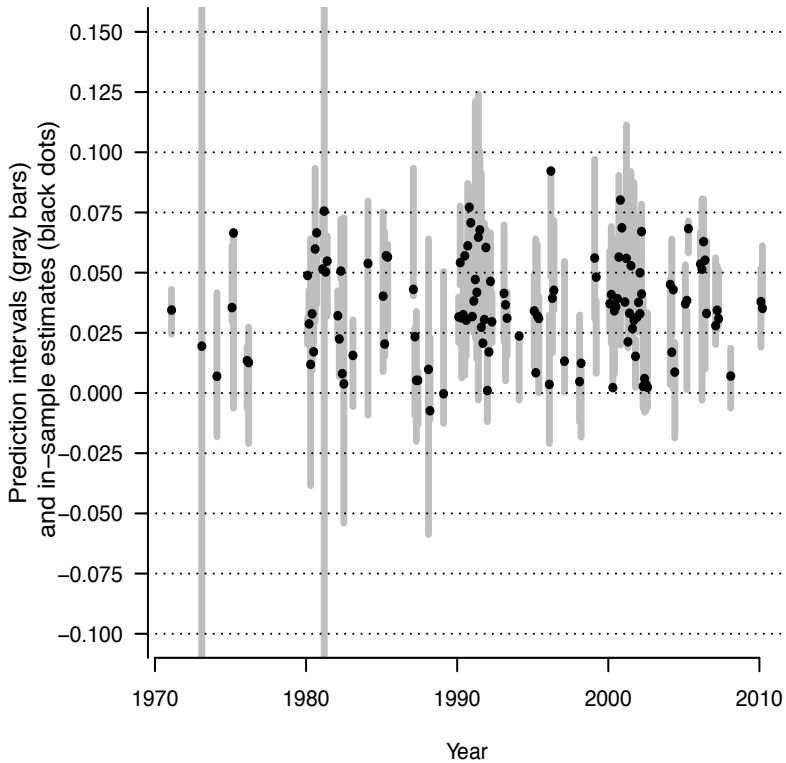
- Need to estimate $\hat{\tau}_1$ and $\text{Var}\left[\zeta_c | C(S_1)\right]$.
- Use series approximation with order determined by LASSO (Belloni, Chernozhukov, and Hansen, 2014).

Data:

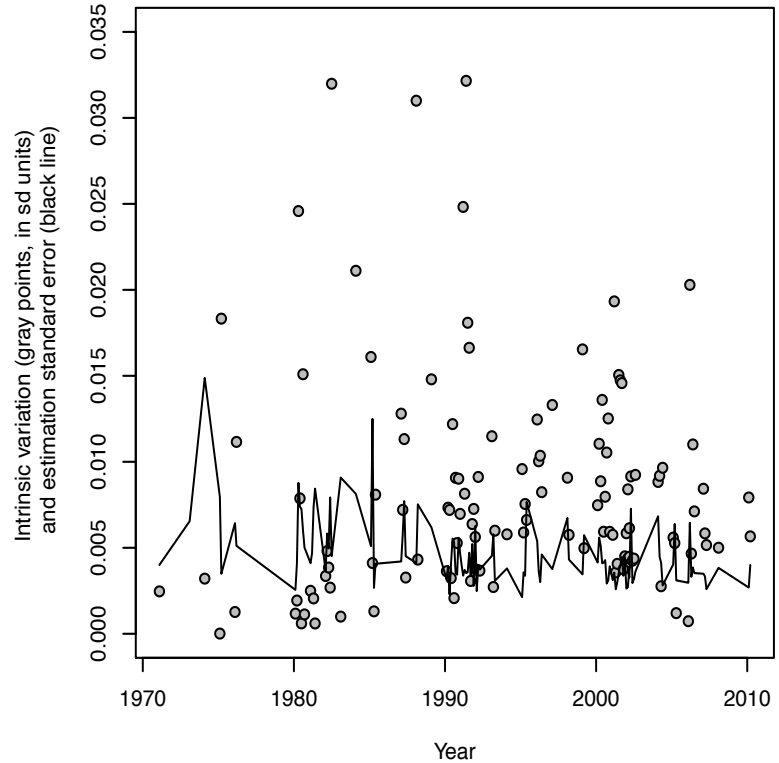
- "Naturalistic simulation" using replications of Angrist & Evans (1998) incremental fertility natural experiment with census data.

Extrapolate or Experiment?

Panel B: Prediction intervals and validation estimates



Panel C: Uncertainty estimates



Conclusions regarding empirical generalization

- Discussions over external validity are often based on theoretical speculation.
- But with data we can do more: *empirical* scrutiny of external validity questions (Imbens, 2010).
- This requires methods for characterizing effect distributions, conducting extrapolation exercises, and using results to inform decision problems that may only require partial or coarsened information.