Lecture 23:
Generalization and Theory

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

October 12, 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*, which means making *counterfactual predictions* for novel settings.

Two ways to think about generalization:

- Bare-foot empirical approach (meta-analysis, statistical extrapolation).
- Through the lens of theory.

Today we will focus on the second.

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).
- Ashworth et al. (2019) elaborate on connecting models to empirics:

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).
- Ashworth et al. (2019) elaborate on connecting models to empirics:
  - Causal identification for empirics ≈ "all else equal" comparative statics for models

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).
- Ashworth et al. (2019) elaborate on connecting models to empirics:
  - Causal identification for empirics $\approx$ "all else equal" comparative statics for models
  - Reinterpreting: use models to enumerate mechanisms that can give rise to an observed pattern; use empirics to evaluate these mechanisms.

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).
- Ashworth et al. (2019) elaborate on connecting models to empirics:
  - Causal identification for empirics ≈ "all else equal" comparative statics for models
  - Reinterpreting: use models to enumerate mechanisms that can give rise to an observed pattern; use empirics to evaluate these mechanisms.
  - Elaborating: use models to derive a variety of observable implications; use empirics to evaluate these various observable implications (≈ "elaborate hypotheses" of Fisher/Rosenbaum).

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).
- Ashworth et al. (2019) elaborate on connecting models to empirics:
  - Causal identification for empirics ≈ "all else equal" comparative statics for models
  - Reinterpreting: use models to enumerate mechanisms that can give rise to an observed pattern; use empirics to evaluate these mechanisms.
  - Elaborating: use models to derive a variety of observable implications; use empirics to evaluate these various observable implications (≈ "elaborate hypotheses" of Fisher/Rosenbaum).
  - Distinguishing: use models to derive distinguishing patterns; use empirics to assess whether these distinguishing patterns are present (relate to Nakamura & Steinsson 2018).

Motivation

- *Models* and generalization: models generate counterfactual predictions (Acemoglu 2010).
- Clarke and Primo (2010): empirics should guide us in making useful models (model:empirics as map:geography).
- Ashworth et al. (2019) elaborate on connecting models to empirics:
  - Causal identification for empirics ≈ "all else equal" comparative statics for models
  - Reinterpreting: use models to enumerate mechanisms that can give rise to an observed pattern; use empirics to evaluate these mechanisms.
  - Elaborating: use models to derive a variety of observable implications; use empirics to evaluate these various observable implications (≈ "elaborate hypotheses" of Fisher/Rosenbaum).
  - Distinguishing: use models to derive distinguishing patterns; use empirics to assess whether these distinguishing patterns are present (relate to Nakamura & Steinsson 2018).
  - Disentangling: use different empirical studies to measure importance of different mechanisms implied by different theories.

Motivation

Two examples:

- Reinterpreting to pinpoint indeterminacies and point the way for further research,
- Use a structural model to infer optimal policy from a randomized experiment.

Reinterpreting to pinpoint indeterminacies

Pinpointing indeterminacies

Following discussion in Wolpin (2013, Section 3.3.1-3.3.3):

- Project STAR: 1985-9 field experiment on effects of class size on learning.
- Classes within 80 schools randomly assigned to having class sizes that were small (13-17) vs regular size (22-26) in 1st-3rd grade.

Pinpointing indeterminacies

Following discussion in Wolpin (2013, Section 3.3.1-3.3.3):

- Project STAR: 1985-9 field experiment on effects of class size on learning.
- Classes within 80 schools randomly assigned to having class sizes that were small (13-17) vs regular size (22-26) in 1st-3rd grade.
- Analysis by Krueger (1999): "students in small classes score about five-seven percentage points higher than those assigned to regular-size classes" (ca. .2-.3 sd effect).
- Formally, we can express this reduced form effect for student i in class g with class size S_{ig} and achievement A_{ig} as

$$A_{ig} = \eta + \alpha S_{ig} + u_{ig}.$$

- How to interpet α?

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).
- Achievement levels in $t = 0$ are $A_0 = g_0(I_0, \mu)$.

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).
- Achievement levels in $t = 0$ are $A_0 = g_0(I_0, \mu)$.
- Parental resources are W and their pref for achievement is ε.

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).
- Achievement levels in $t = 0$ are $A_0 = g_0(I_0, \mu)$.
- Parental resources are W and their pref for achievement is ε.
- Inputs at $t = 0$ are utility-maximizing as $I_0 = f_0(W, \mu, \varepsilon)$.

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).
- Achievement levels in $t = 0$ are $A_0 = g_0(I_0, \mu)$.
- Parental resources are W and their pref for achievement is ε.
- Inputs at $t = 0$ are utility-maximizing as $I_0 = f_0(W, \mu, \varepsilon)$.
- Families choose location given expected school inputs, $\bar{S}_1 = \phi_S(A_0, W, \mu, \varepsilon)$, which relate to actual school choices in inputs as $S_1 - \bar{S}_1 = \psi(A_0, \mu)$.

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).
- Achievement levels in $t = 0$ are $A_0 = g_0(I_0, \mu)$.
- Parental resources are W and their pref for achievement is ε.
- Inputs at $t = 0$ are utility-maximizing as $I_0 = f_0(W, \mu, \varepsilon)$.
- Families choose location given expected school inputs, $\bar{S}_1 = \phi_S(A_0, W, \mu, \varepsilon)$, which relate to actual school choices in inputs as $S_1 - \bar{S}_1 = \psi(A_0, \mu)$.
- Family inputs in kindergarten: $I_1 = f_1(A_0, W, S_1 - \bar{S}_1, \mu, \varepsilon)$.

Pinpointing indeterminacies

Wolpin has us consider a more complete structural model:

- $t = 0$ is period between conception and kindergarten, $t = 1$ is kindergarten period, $t = 2$ is first grade period.
- I_0 are family inputs in period 0 that affect knowledge, and μ is the child's innate ability (independent of inputs).
- Achievement levels in $t = 0$ are $A_0 = g_0(I_0, \mu)$.
- Parental resources are W and their pref for achievement is ε.
- Inputs at $t = 0$ are utility-maximizing as $I_0 = f_0(W, \mu, \varepsilon)$.
- Families choose location given expected school inputs, $\bar{S}_1 = \phi_S(A_0, W, \mu, \varepsilon)$, which relate to actual school choices in inputs as $S_1 - \bar{S}_1 = \psi(A_0, \mu)$.
- Family inputs in kindergarten: $I_1 = f_1(A_0, W, S_1 - \bar{S}_1, \mu, \varepsilon)$.
- Achievement after kindergarten: $A_1 = g_1(I_0, I_1, S_1, \mu)$.

Pinpointing indeterminacies

- Now, based on this model and taking location decisions as given,

$$\frac{dA_1}{dS_1} = \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}.$$

Pinpointing indeterminacies

- Now, based on this model and taking location decisions as given,

$$\frac{dA_1}{dS_1} = \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}.$$

- In Project STAR, the change in class size was unanticipated, and so location decisions were not affected.
- As such, α corresponds precisely to expected value of $\frac{dA_1}{dS_1}$:

$$\alpha = \mathrm{E}\left[\frac{dA_1}{dS_1}\right] = \int \left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}\right] dF(W, \mu, \varepsilon).$$

Pinpointing indeterminacies

- Now, based on this model and taking location decisions as given,

$$\frac{dA_1}{dS_1} = \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial(S_1 - \bar{S}_1)}.$$

- In Project STAR, the change in class size was unanticipated, and so location decisions were not affected.
- As such, α corresponds precisely to expected value of $\frac{dA_1}{dS_1}$:

$$\alpha = \mathrm{E}\left[\frac{dA_1}{dS_1}\right] = \int \left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_1}\frac{\partial I_1}{\partial(S_1 - \bar{S}_1)}\right] dF(W,\mu,\varepsilon).$$

- Equiv to "total effect of an exogenous change in class size on achievement, not holding other inputs constant," not effect of class size on the achievement production function ($\mathrm{E}[\partial g_1/\partial S_1]$).
- An average "policy effect" rather than a ceteris paribus effect.

Pinpointing indeterminacies

- Policy effects and ceteris paribus effects could relate in various ways:
  - If parents substitute inputs in S_1, the policy effect may be smaller than the ceteris paribus effect.
  - But if parents complement, then the opposite would be true.
- So, suppose $\alpha = 0$. Does this mean the program had “no effect”? Does it mean that it didn’t improve welfare?
- So there is an indeterminacy here in what, exactly, α means from a welfare perspective.

Pinpointing indeterminacies

- Another type of indeterminacy regarding α has to do with the way that it would relate to kids coming into kindergarten in years after the experiment *in the same schools that were part of the treatment group*.

Pinpointing indeterminacies

- Another type of indeterminacy regarding α has to do with the way that it would relate to kids coming into kindergarten in years after the experiment *in the same schools that were part of the treatment group*.
- Suppose that parents don't relocate. Nonetheless, once they've learned about the change they have a new $\bar{S}_1$, and so may adjust I_0. So for this second wave of kids

$$\mathrm{E}\left[\frac{dA_1}{dS_1}\right] = \int \left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_0}\frac{\partial I_0}{\partial \bar{S}_1} + \frac{\partial g_1}{\partial I_1}\left(\frac{\partial I_1}{\partial \bar{S}_1} + \frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}\right)\right] dF(W, \mu, \varepsilon)$$

which generally is not equal to α.

Pinpointing indeterminacies

- Another type of indeterminacy regarding α has to do with the way that it would relate to kids coming into kindergarten in years after the experiment *in the same schools that were part of the treatment group*.
- Suppose that parents don't relocate. Nonetheless, once they've learned about the change they have a new $\bar{S}_1$, and so may adjust I_0. So for this second wave of kids

$$\mathrm{E}\left[\frac{dA_1}{dS_1}\right] = \int\left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial I_0}\frac{\partial I_0}{\partial \bar{S}_1} + \frac{\partial g_1}{\partial I_1}\left(\frac{\partial I_1}{\partial \bar{S}_1} + \frac{\partial I_1}{\partial (S_1 - \bar{S}_1)}\right)\right] dF(W, \mu, \varepsilon)$$

which generally is not equal to α.
- So we are already forced to speculate in making inferences to these kids – the little brothers and sisters of experimental first cohort!
- Now allow for relocation...

Pinpointing indeterminacies

- The structural model allows us to identify two types of indeterminacies:
  - Welfare impact given an estimate of α.
  - Implications for keeping the treatment in place.
- What are the implications for how a research program should proceed?
  - Would have been nice for this analysis to *precede* the implementation of the experiment to define additional outcomes and effects to measure: *theory-informed pre-analysis plan*.
  - Offers suggestions on what kind of research should follow...

Pinpointing indeterminacies

Parental Responses to Information About School Quality: Evidence from Linked Survey and Administrative Data*

Ellen Greaves Iftikhar Hussain Birgitta Rabe Imran Rasul†

January 2021

Abstract

Multiple inputs determine children's academic achievement. We study the interaction between family and school inputs by identifying the causal impact of information about school quality on parental time investment into children. Our setting is England, where credible information on school quality is provided by a nationwide school inspection regime. Schools are inspected at short notice, with school ratings using hard and soft information. As such soft information is not necessarily known to parents *ex ante*, inspection ratings provide news to parents that shifts parental beliefs about school quality, and hence their investment into their children. We study this using household panel data linked to administrative records on school performance and inspection ratings. Within the same academic year, we observe some households being interviewed pre school inspection, and others being interviewed post inspection. Treatment assignment is determined by a household's survey date relative to the school inspection date, and shown to be as good as random. We find that parents receiving good news over school quality significantly decrease time investment into their children (relative to parents that will later receive such good news). Our data and design allow us to provide insights on the distributional and test score impacts of the nationwide inspections regime, through multiple margins of endogenous response of parents and children. Our findings highlight the importance of accounting for interlinked private responses by families to new public information on school quality. *JEL Classification: I20, I24.*

Pinpointing indeterminacies

- A second example comes from Bueno de Mesquita & Tyson (2020).
- "What is the effect of protest on government policy?"
- A common research design is to use, e.g., rainfall as an instrument for whether protest happens. What does this tell us?

Pinpointing indeterminacies

Bueno de Mesquita & Tyson (2020) consider the following example:

- Protestors are either high benefit, $\overline{\theta}$, or low benefit, $\underline{\theta}$. High types exist with probability p, known to govt.
- Cost of protesting may be high, $\underline{c}$, or low, $\underline{c}$.
- Protest only occurs if benefit exceeds cost. Let $\overline{\theta} > \overline{c} > \underline{\theta} > \underline{c}$. High types always protest, low type only when costs are low.
- Govt. observes whether protest occurs, $\mathbb{I}$, thens form posterior belief $\hat{p}$ about whether observing $\overline{\theta}$ type.
- Govt. formulates a response as

$$r(\mathbb{I},\hat{p}) = \begin{cases} \overline{\alpha} + \overline{\beta}\hat{p} & \text{if } \mathbb{I} = 1 \\ \underline{\alpha} + \underline{\beta}\hat{p} & \text{if } \mathbb{I} = 0 \end{cases}$$

- α terms are "direct effect of protest" and β terms are "informational effect."

Pinpointing indeterminacies

- Suppose we have random coin-flip shocks that affect costs, such that shocks make it so that both types are willing to protest.
- If these shocks *are hidden to the govt.*, then upon observing protest, govt. can only infer $\hat{p} = 2p/(1+p)$, and upon no protest, $\hat{p} = 0$.
- The reduced form effect of a such a change in costs is thus,

$$
\begin{aligned}
& r\left(1, \frac{2p}{1+p}\right) - \left[pr\left(1, \frac{2p}{1+p}\right) + (1-p)r(0,0)\right] \\
& = (1-p)\left[\overline{\alpha} - \underline{\alpha} + \overline{\beta}\left(\frac{2p}{1+p}\right)\right]
\end{aligned}
$$

Pinpointing indeterminacies

- But if shocks are *observable* to the govt.. Upon protest, govt. will also look to whether there is a cost shock. If there is, govt. can only infer $\hat{p} = p$, if there is not, then govt. can infer $\hat{p} = 1$. No protest only happens in the absence of a shock, and in that case $\hat{p} = 0$.
- The reduced form effect of this change in observable costs is

$$r(1,p) - [pr(1,1) + (1-p)r(0,0)] = (1-p)(\overline{\alpha} - \underline{\alpha}).$$

- Unlike before, the *informational effect* is no longer present.

Pinpointing indeterminacies

Implications of the analysis:

- The analysis here applies to situations where you have "treatments" that are the product of choices that may signal information to those whose responses serve as the effects of interest.
- Generally, it is important to assess whether the research design targets a causal quantity that you really care about.

Use a structural model to infer optimal policy

Use a structural model to infer optimal policy

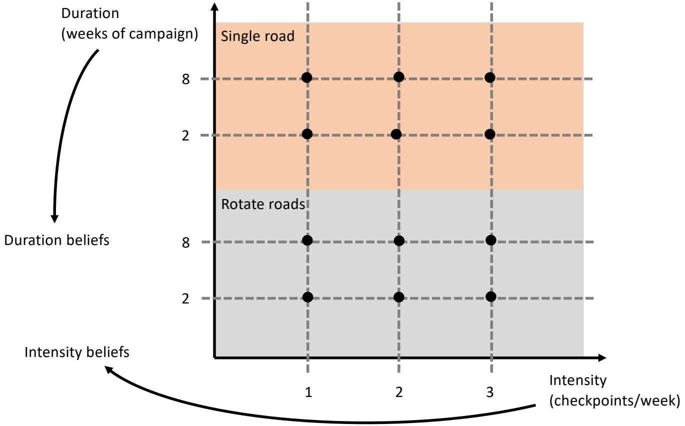
- Models can be excellent tools for *policy design*.
- Model mechanisms that sustain a problem, use it to define interventions to disrupt the mechanism.
- Experimental (or natural experimental) evidence can be used to calibrate the model to infer optimal policies.

Use a structural model to infer optimal policy

- Banerjee et al. (2019) study strategies to combat drunk driving.
- Analysis is informed by a model of *optimal deterrence*, where the key mechanism is drunk drivers' beliefs about likelihood and intensity of punishment.
- Goal is *efficient* deployment of resources: how to achieve target levels of deterrence at minimal cost.
- Test with an RCT: "sobriety checkpoints were either rotated among 3 locations or fixed in the best location, and the intensity [(checkpoints/week, duration of campaign, incentives)] of the crackdown was cross-randomized."
- Assessed outcomes on traffic volume, number of drunk drivers apprehended, police movement, accidents, and deaths.

Use a structural model to infer optimal policy

Design space:



Use a structural model to infer optimal policy

They first carry out reduced form analysis:

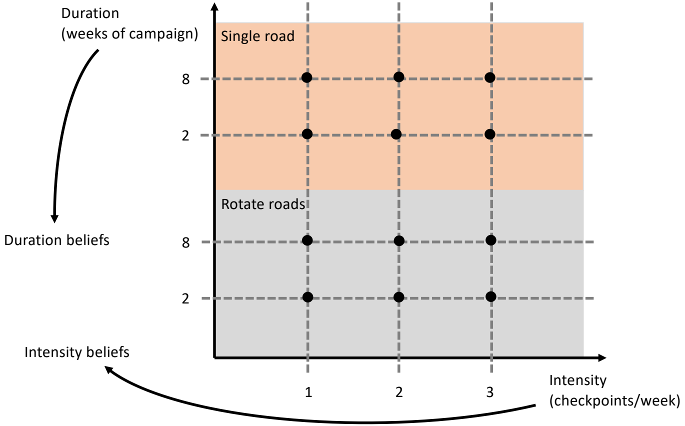
- "Rotating checkpoints reduced night accidents by 17%, and night deaths by 25%, while fixed checkpoints had no significant effects."

Table 4: Fixed vs. Rotating Pooled Results

	Daylight		Darkness		Day & Night
	(1) Accidents	(2) Deaths	(3) Accidents	(4) Deaths	(5) Deaths
Fixed checkpoints during & post intervention	-0.0021	-0.0041	-0.0016	-0.0029	-0.0046
	(0.0045)	(0.0033)	(0.0029)	(0.0023)	(0.0035)
Rotating checkpoints during & post intervention	0.0084*	-0.0005	-0.0096***	-0.0050*	-0.0026
	(0.0045)	(0.0036)	(0.0028)	(0.0029)	(0.0041)
Month fixed effects	Yes	Yes	Yes	Yes	Yes
Police Station FE	Yes	Yes	Yes	Yes	Yes
Mean of dep. variable	0.085	0.029	0.033	0.016	0.045
P-value of test fixed = rotating effect	0.0398	0.378	0.0130	0.473	0.660
N	5090	4724	5090	4724	5090

Use a structural model to infer optimal policy

Design space:



Use a structural model to infer optimal policy

Then use RCT data to calibrate a structural model and determine "optimal" strategy:

- Accounts for driver learning and strategic response.
- Allows them to evaluate counterfactual policies.
- Rotating checkpoints dominate in experiment, so focus on them.
- Optimize wrt duration, intensity, and distribution over primary vs secondary roads.

Table 11: Optimal Crackdown Strategy and Duration

	Prior Warning	Implementation Compliance: Full			Implementation Compliance: Partial		
		Duration	Pct. on road 1	Decrease in drunken driving	Duration	Pct. on road 1	Decrease in drunken driving
		(1)	(2)	(3)	(4)	(5)	(6)
Estimated beliefs	No	82 days	35%	36.1%	83 days	35%	25.0%
	Yes	77 days	20%	46.6%	81 days	35%	36.5%
Equilibrium beliefs	No	21 days	75%	89.9%	43 days	95%	71.7%
	Yes	>31 days	35%	100%	>31 days	35%	100%

Conclusions

Conclusions

- Theoretical models *are* a tool for counterfactual prediction, but the problem is you can make a model to motivate almost any prediction.

Conclusions

- Theoretical models *are* a tool for counterfactual prediction, but the problem is you can make a model to motivate almost any prediction.
- Empirics are a way of gaining insight on how the world works, but the problem is that without the discipline of a formal model you might overstep (e.g., Project STAR example).

Conclusions

- Theoretical models *are* a tool for counterfactual prediction, but the problem is you can make a model to motivate almost any prediction.
- Empirics are a way of gaining insight on how the world works, but the problem is that without the discipline of a formal model you might overstep (e.g., Project STAR example).
- A goal, then, is to find productive ways to get models and empirics to interact.