Designing RCTs with External Validity in Mind

Sylvain Chassang
NYU & NBER
chassang@nyu.edu

preliminary, updates will be made available at
www.sylvainchassang.org

November 12, 2019
What’s hard about external validity?

Setup

- Contexts $C_0$ and $C_1$
- Can measure $\mathbb{E}[\Delta Y | C_0]$, would like to measure $\mathbb{E}[\Delta Y | C_1]$
- Effects mediated by covariates $X$

The good case: consistent extrapolation

1. $X$ observable (or revealed)
2. $\text{supp } X|C_1 \subseteq \text{supp } X|C_0$, and
3. $\mathbb{E}[\Delta Y | X, C] = \mathbb{E}[\Delta Y | X]$

What can we do when these fail?

- get as close as possible
- view experimentation as a dynamic process
Rich and diverse covariates

**Enriching covariates**
- goal $\mathbb{E}[\Delta Y|X, C] \approx \mathbb{E}[\Delta Y|X]$
- include more observables
  e.g. network structure (Banerjee et al., 2019)
- include elicited/structurally identified unobservables
  e.g. value and beliefs over treatment (Chassang et al., 2012), trust metrics, structural params

**Diverse covariates**
- goal $\text{supp } X|C_1 \subset \text{supp } X|C_0$
- use more heterogeneous samples, across space, time, socio economic status
- similar to the point that IV and RD reflect treatment effects for selected population
A note on balance

More diverse covariates have a cost in terms of power.

Imbalances from random assignment get worse if covariates are heterogenous.

Rerandomization and balance in RCT design

- Bruhn and McKenzie (2009) highlights that rerandomization frequent, but rarely accounted for.
- Banerjee et al. (forthcoming) clarifies trade off between balance and robustness → it’s small.

Provides guidelines: pick a random assignment within a fixed quantile of balance (e.g. top 5, 10%).

- Python package rct, simplify such draws, ready for use, if interested please reach out.

\[\text{pip install rct}\]
Testing for consistent extrapolation

Valuable side effect of diverse covariates: extrapolate within experiment

\[ C_0 = C_{0,A} \cup C_{0,B} \]

Test of consistent extrapolation (related to Kowalski (2018))

\[ \mathbb{E}[\Delta Y|X, C_{0,A}] = \mathbb{E}[\Delta|X, C_{0,B}] \]

OK to use more representative weights when estimating conditional treatment effects

Might be a good case for boosting (Freund et al. (1999))
What if consistent extrapolation fails?

View experimentation as a process
Experiment today, experiment tomorrow . . .
Structured Speculation

When consistent extrapolation fails, any out-of-context estimate is a subjective belief

Because experimentation is dynamic, beliefs are not cheap

► today’s predictions about how treatment effects vary with context, or how a different treatment will perform, can be confronted against tomorrow’s experiment

► Banerjee et al. (2017) advocate dedicated section in papers where authors freely engage in speculation with no burden of proof

► potentially helps get valuable information out, collect different data
Evaluating Experts

► Shift the question from

“what is the treatment effect in a single new context” to
“can we select the best extrapolation method over successive implementations”

► Candidate experts: experimenter (potentially from structured speculation), group predictions, prominent ML frameworks ...

Online learning litt (Hannan (1957), Blackwell (1956), Foster and Vohra (1999)) suggests doable

► Difficulty – hard to match studies
treatments change, covariates change, outcomes change

Institutional suggestion: create standards for what covariates and outcome to measure and how
Will facilitate meta-analyses
Build in adaptation

Why would a policy maker rely on non-context relevant evidence to implement policy?

Once a policy looks interesting, why not run a rapid local experimentation to validate context predictions?

Can be an RCT, a staggered implementation design . . .

Suggests two new questions:

- How well can we extrapolate using a seed of context relevant data?
- Becomes interesting to predict the **option value of treatment** rather than just the realized treatment effect
Takeaways

▶ use diversified and balanced samples
▶ engage in structured speculation
▶ robustly learn the best extrapolating expert
  standardize $X$ and $Y$ to facilitate linking studies
▶ build in adaptation in analysis, and start thinking about
  option value of treatments


References II


References III