Evaluating *Ex Ante* Counterfactual Predictions with *Ex Post* Causal Inference

Cyrus Samii, joint with Michael Gechter, Rajeev Dehejia, and Cristian Pop-Eleches

2019 DRI Annual Conference
Motivation

• What **method** to use to inform a policy choice in a given context?
  • External experiments? Internal observational evidence? (cf. Pritchett & Sandefur 2015)
  • Highly nuanced model? Simple model? Agnostic “model-free” methods?

• To answer the question, need a way to **judge** methods.
  • Methods make *ex ante* proposals.
  • We leverage subsequent experiments to do *ex post* evaluation.
Contribution

• Define criterion for methods to target *ex ante*
  • Consider a policy maker’s utility function
  • Ask methods to recommend policies to maximize *ex ante* this utility

• Define a method for *ex post* judgement based on this criterion
  • We use *ex post* experiments to judge
  • Experiments offer design-unbiased ways to compare methods

• Consider a *variety of methods* to form policy recommendations
  • Structural models of varying complexity
  • Reduced form methods of varying complexity
Application

• Promoting school enrollment in Morocco.
  • Policy maker is considering conditional cash transfers.
  • Wants to know how to optimally target the transfers.
  • Maximizes
    \[
    \text{treated} \times (\text{gains} - \text{alternative uses of funds}) \\
    + \text{untreated} \times (\text{alternative uses of funds} – \text{foregone gains})
    \]

• Information available *ex ante* to derive recommendations
  • Survey data from Morocco
  • Experiments from elsewhere (Mexico Progresa)
Methods for Ex Ante Recommendations

• Extrapolating from Progresa RCT in Mexico
  • Generalized random forest
    (cf. Athey et al. 2019)
  • Simple regression interacting gender and years-of-ed. dummies

• Analyzing observational data in Morocco
  • Semiparametric static structural model
    (cf. Todd & Wolpin 2006)
  • Parametric dynamic structural model
    (cf. Attanasio et al. 2012)
  • Exogeneity and (semi-)parametric assumptions
Strategy for Ex Post Evaluation

• Each method tries to maximize policy-maker utility
• How did they do?
• Make ex post judgment using Tayssir RCT in Morocco
• RCT allows for design-unbiased ex post estimation of utility from each method

\[
\text{treated} \times (\text{gains - alternative uses of funds}) \\
+ \text{untreated} \times (\text{alternative uses of funds – foregone gains})
\]
Results (preliminary)

• First, sanity check: optimize utility in Mexico using data from Mexico
• GRF significantly outperforms simple regression

|                        | w/ CE |
|------------------------|-------|
| Share treated (GRF)    | 0.169 |
| Share treated (yrs. educ-sex) | 0.107 |
| Enrollment difference  | 0.140 |
| SE enroll. diff.       | 0.004 |
| Welfare difference     | 0.127 |
| SE welfare diff.       | 0.004 |

Welfare comparison for GRF vs. yrs.educ.-sex extrapolation

• For structural models, original papers showed similar
Results (preliminary)

• Now move on to Morocco

• Simple regression *from Mexico* outperforms blanket rule of treating everyone:

|                                            | w/ CE |
|-------------------------------------------|------|
| Share treated (yrs. educ-sex)             | 0.020|
| Share treated (all)                      | 1    |
| Enrollment difference                    | -0.054|
| SE enroll. diff.                         | 0.010|
| Welfare difference                       | 0.028|
| SE welfare diff.                         | 0.009|

Welfare comparison for yrs.educ.-sex extrapolation vs. treat all.
Results (preliminary)

• But we *do not* see additional gains from fancier GRF extrapolation:

|                                | w/ CE  |
|--------------------------------|--------|
| Share treated (GRF)            | 0.172  |
| Share treated (yrs. educ-sex)  | 0.020  |
| Enrollment difference          | 0.013  |
| SE enroll. diff.               | 0.007  |
| Welfare difference             | 0.005  |
| SE welfare diff.               | 0.006  |

Welfare comparison for GRF vs. yrs.educ.-sex extrapolation
Results (preliminary)

- Nor from structural models fit to Morocco observational data:

|                      | w/ CE |                      | w/ CE |
|----------------------|-------|----------------------|-------|
| Share treated (SPS)  | 0.269 | Share treated (SPS)  | 0     |
| Share treated (yrs. educ-sex) | 0.020 | Share treated (yrs. educ-sex) | 0.020 |
| Enrollment difference | 0.010 | Enrollment difference | 0.002 |
| SE enroll. diff.     | 0.009 | SE enroll. diff.     | 0.003 |
| Welfare difference   | -0.010| Welfare difference   | 0.003 |
| SE welfare diff.     | 0.008 | SE welfare diff.     | 0.003 |

Welfare comparison for SPS vs. yrs.educ.-sex extrapolation

Welfare comparison for DPS vs. yrs.educ.-sex extrapolation
Implications and Extensions

• Simplicity travels, nuance does not?
• What if we add more external experiments?
  • Ongoing work with RCTs from 7 contexts
• What if effects are more complex?
  • Ongoing work considers spillovers and equilibrium effects.