Randomized trials for policy

A review of the external validity of treatment effects

Seán M. Muller

Development Policy Research Unit, University of Cape Town

Annual Bank Conference on Development Economics, 2-3 June 2014

Table of contents

Paper overview

1. Review of critiques of RCTs and responses

2. Review of literature(s) on external validity: programme evaluation; experimental economics; medicine; philosophy; structural econometrics; time-series econometrics

3. External validity as a problem of interaction: framework, possible solutions and implications

Background

- RCTs have become very important in (development) economics but also controversial:
  - Internal validity: (when) do RCTs identify a causal effect of interest?
  - External validity: what can we infer about causal relationships in non-experimental populations from RCT results?
  - Can RCTs address the 'big' questions of development?

- Overlap obscures the fundamental problem of external validity, so I focus on extrapolation from an *ideal experiment*.

Interaction: the fundamental challenge to external validity

Interaction: the fundamental challenge to external validity

Definition Simple external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1] = E[Y_i(1) - Y_i(0)|D_i = 0] \quad (1)$$

Interaction: the fundamental challenge to external validity

Definition Simple external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1] = E[Y_i(1) - Y_i(0)|D_i = 0] \quad (1)$$

- "all...threats to external validity [can be described] in terms of statistical interaction effects" (Cook and Campbell, 1979)

Interaction: the fundamental challenge to external validity

Definition Simple external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1] = E[Y_i(1) - Y_i(0)|D_i = 0] \quad (1)$$

- "all...threats to external validity [can be described] in terms of statistical interaction effects" (Cook and Campbell, 1979)
- Straightforward to show that if treatment variable interacts with some covariate(s) (W) then simple external validity fails where $E[W|D = 1] \neq E[W|D = 0]$

Illustrative example: class size and test scores

- Class size has been important example in EV debates (Angrist and Pischke, 2010), but empirical studies are based on an *additive* educational production function.

Illustrative example: class size and test scores

- Class size has been important example in EV debates (Angrist and Pischke, 2010), but empirical studies are based on an *additive* educational production function.
- Alternative (simple) theory: class size matters *because of what happens in the classroom*

Illustrative example: class size and test scores

- Class size has been important example in EV debates (Angrist and Pischke, 2010), but empirical studies are based on an *additive* educational production function.
- Alternative (simple) theory: class size matters *because of what happens in the classroom*
- Formally:

$$\triangle A_{ijgk} = \triangle\alpha_{0ig} + \alpha_1 H_{ig} + \beta(1 - \delta C_{gj}) f(q^*_{gj}, R_{gj}, \alpha_{0\cdot jg}) + \alpha_2 G_{gk} + \epsilon_{igjk}$$

Sampling and replication

- Cook and Campbell (1979) frame *problem* in terms of interaction and *solution* in terms of sampling or replication. (More recently, see Allcott and Mullainathan (2012)).

Sampling and replication

- Cook and Campbell (1979) frame *problem* in terms of interaction and *solution* in terms of sampling or replication. (More recently, see Allcott and Mullainathan (2012)).

- First option: 'random sampling for representativeness'

Sampling and replication

- Cook and Campbell (1979) frame *problem* in terms of interaction and *solution* in terms of sampling or replication. (More recently, see Allcott and Mullainathan (2012)).
- First option: 'random sampling for representativeness'
- Second option: 'deliberate sampling for heterogeneity'

Conditional external validity

Conditional external validity

Definition Conditional external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1]$$
$$= E_W[E[Y_i|T_1, D_i = 0, W_i] - E[Y_i|T_0, D_i = 0, W_i]|D_i = 1]$$

Conditional external validity

Definition Conditional external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1]$$
$$= E_W[E[Y_i|T_1, D_i = 0, W_i] - E[Y_i|T_0, D_i = 0, W_i]|D_i = 1]$$

Hotz et al. (2005) show three conditions are sufficient. Successful randomization plus:

Conditional external validity

Definition Conditional external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1]$$
$$= E_W[E[Y_i|T_1, D_i = 0, W_i] - E[Y_i|T_0, D_i = 0, W_i]|D_i = 1]$$

Hotz et al. (2005) show three conditions are sufficient. Successful randomization plus:

Location independence

$$D_i \perp\!\!\!\perp (Y_i(0), Y_i(1))|W_i \tag{2}$$

Conditional external validity

Definition Conditional external validity

$$E[Y_i(1) - Y_i(0)|D_i = 1]$$
$$= E_W[E[Y_i|T_1, D_i = 0, W_i] - E[Y_i|T_0, D_i = 0, W_i]|D_i = 1]$$

Hotz et al. (2005) show three conditions are sufficient. Successful randomization plus:

Location independence

$$D_i \perp\!\!\!\perp (Y_i(0), Y_i(1))|W_i \tag{2}$$

Overlapping support

$$\text{For all } w,\ \delta < Pr(D_i = 1|W_i = w) < 1 - \delta, \tag{3}$$
$$\text{for some } \delta > 0 \text{ and for all } w \in W$$

Problem 1: Empirical requirements

Table: *Empirical requirements for external validity*
(assuming an ideal experiment, no specification of functional form)

R1	The interacting factors (W) must be known *ex ante*
R2	All elements of W must be observed in both populations
R3.1	Empirical measures of elements of W must be comparable across populations
R4.1	The researcher must be able to obtain unbiased estimates of the conditional average treatment effect ($E[\triangle \vert D = 0, W]$) for all values of W

Problem 2: Inconsistency

- Manski (2013a,b) has noted asymmetry in dealing with internal and external validity. Above framework elucidates one simple aspect of this.

Problem 2: Inconsistency

- Manski (2013a,b) has noted asymmetry in dealing with internal and external validity. Above framework elucidates one simple aspect of this.
- Assumptions required for non-experimental matching methods?

Problem 2: Inconsistency

- Manski (2013a,b) has noted asymmetry in dealing with internal and external validity. Above framework elucidates one simple aspect of this.
- Assumptions required for non-experimental matching methods?
  1. Unconfoundedness/selection on observables $(T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))|X)$

Problem 2: Inconsistency

- Manski (2013a,b) has noted asymmetry in dealing with internal and external validity. Above framework elucidates one simple aspect of this.
- Assumptions required for non-experimental matching methods?
  1. Unconfoundedness/selection on observables $(T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))|X)$
  2. Overlapping support (across $T = 0$ and $T = 1$)

Problem 2: Inconsistency

- Manski (2013a,b) has noted asymmetry in dealing with internal and external validity. Above framework elucidates one simple aspect of this.
- Assumptions required for non-experimental matching methods?
  1. Unconfoundedness/selection on observables $(T_i \perp\!\!\!\perp (Y_i(0), Y_i(1))|X)$
  2. Overlapping support (across $T = 0$ and $T = 1$)
- But these are equivalent in form to requirements for conditional external validity...using X and T, instead of W and D

Implications

- External validity problem is currently unresolved

Implications

- External validity problem is currently unresolved
- Therefore:

Implications

- External validity problem is currently unresolved
- Therefore:
  1. Either more caution in claiming policy relevance of randomised evaluations;

Implications

- External validity problem is currently unresolved
- Therefore:
  1. Either more caution in claiming policy relevance of randomised evaluations;
  2. *Or* acceptance that qualitative (subjective?) assessment of external validity is inconsistent with insisting on randomization for internal validity.

Implications

- External validity problem is currently unresolved
- Therefore:
  1. Either more caution in claiming policy relevance of randomised evaluations;
  2. *Or* acceptance that qualitative (subjective?) assessment of external validity is inconsistent with insisting on randomization for internal validity.
  3. Replication (maybe also random sampling) cannot answer external validity question without information on interacting variables.

Implications

- External validity problem is currently unresolved
- Therefore:
  1. Either more caution in claiming policy relevance of randomised evaluations;
  2. *Or* acceptance that qualitative (subjective?) assessment of external validity is inconsistent with insisting on randomization for internal validity.
  3. Replication (maybe also random sampling) cannot answer external validity question without information on interacting variables.
- 'Theory' may help by providing guidance on what the interacting factors *might* be, but empirical obstacles remain impressive and may be insurmountable in some (many?) cases

References

Allcott, H. and S. Mullainathan (2012). External validity and partner selection bias. *NBER Working Paper* (18373).

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Cook, T. D. and D. T. Campbell (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Wadsworth.

Hotz, V. J., G. W. Imbens, and J. H. Mortimer (2005). Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics 125*, 241–270.

Manski, C. F. (2013a). *Public policy in an uncertain world: analysis and decisions*. Cambridge (MA): Harvard University Press.

Manski, C. F. (2013b). Response to the review of 'public policy in an uncertain world'. *Economic Journal 123*, F412–F415.