External Validity

Nandan Rao

September, 2019

Introduction

External Validity and Scientific Inference

The Danger of Informal Inference

Current State of the Art in Economics

The Origins of Structure

Introduction

The Randomista Fight

- Experimental and quasi-experimental methods have become the standard, with RCTs as the gold standard.

- Indentification police!

- Many prominent econometricians have rasied the alarm about this obsession. Heckman (2008), Manski (2013), Deaton (2010), Deaton and Cartwright (2018)

Applied Research

- If applied research is to be applied, it needs to be applicable.

- This means external validity. Manski (2013)

- Experiments, identification, etc. focus on internal validity (we will define this well later).

- A plethory of new studies show that experiments don't generalize. (**???**), Gechter (2015), Allcott (2015), Bisbee et al. (2017), Rosenzweig and Udry (2019)

This work

How did we get here and where can we go?

External Validity and Scientific Inference

Taxonomy of validities[1]

Statistical Conclusion Validity: The validity of inferences about the correlation (covariation) between treatment and outcome.

Internal Validity: The validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured.

Construct Validity: The validity of inferences about the higher order constructs that represent sampling particulars.

External Validity: The validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.

[1]Shadish, Cook, and Campbell (2002)

What is inference?

Inference is the process of reasoning and can be broken into two parts:

1. Reasoning from particulars to generals (induction)

2. Reasoning from generals to particulars (deduction).

Syllogisms

All men are mortal

Socrates is a man

Therefore, Socrates is mortal

Hume's Problem of Induction

"As to past Experience, it can be allowed to give direct and certain information of those precise objects only, and that precise period of time, which fell under its cognizance: but why this experience should be extended to future times, and to other objects, which for aught we know, may be only in appearance similar, this is the main question on which I would insist"

The Design of Experiments

"it is possible to draw valid inferences from the results of experimentation. . . as a statistician would say, from a sample to the population from which the sample was drawn, or, as a logician might put it, from the particular to the general." – (???)

Statistics as Induction

- Allows us to extend conclusions not just to objects similar in appearance, but to a population from a sample.

- Fisher's framework for significance testing falls under the validity of "statistical conclusion validity" in Shadish, Cook, and Campbell's validity taxonomy.

- His theory of experiments randomization fall under "internal validity."

- The causal discovery comes straight from John Stuart Mill's Method of Difference.

11

Statistics as Induction

"Others are concerned with deducing the causes of a given effect. Still others are interested in understanding the details of causal mechanisms. The emphasis here will be on measuring the effects of causes because this seems to be a place where statistics, which is concerned with measurement, has contributions to make." – **(???)**

Counterfactual Models

- The effects of causes is not the same as the causes of a given effect or the details of the causal mechanism.

- Also does not distinguish "cause" from "necessary part of the cause".

- This focus on the effects of causes has led to the success of this framework: it operationalizes Mill's method of difference!

- The effects of causes is a context-specific question. It does not ask the question "why". It is a black-box method for causal identification[2].

---

[2]Heckman (2008)

Agriculture and Medecine

- When contexts can be reproduced, when the population that is sampled from is the same population on which one will infer conclusions, then the "effects of a cause" is exactly the right question.

- In Fisher's context, this was the case. The context for growing corn can be replicated, because one knows the system is encapsulated: sunlight and soil.

- Economics is not like this. Social sciences are different. This is why Shadish, Cook, and Campbell go to such lengths to enumerat threats to external validity.

The Danger of Informal Inference

Policy Makers

Policy makers are experts of their target domain. Should they not be the most capable of deciding if an empirical study can extrapolate to their context? Are they not better suited to the job than a researcher?

In the framework of scientific inference, this amounts to reasoning from particulars to particulars.

The Warning of John Stuart Mill

"In reasoning from a course of individual observations to some new and unobserved case, which we are but imperfectly acquainted with (or we should not be inquiring into it), and in which, since we are inquiring into it, we probably feel a peculiar interest; there is very little to prevent us from giving way to negligence, or to any bias which may affect our wishes or our imagination, and, under that influence, accepting insufficient evidence as sufficient." – John Stuart Mill

Reasoning from Individual Studies

- Individual studies, performed in the counterfactual model, amount to particular experiences.

- Making policy decisions directly on the basis of such particulars is the exact situation that John Stuart Mill warned is fertile ground for bias and imagination.

- The defense against this is a "formal framework."

- Does such a framework exist?

Current State of the Art in Economics

Shadish, Cook, and Campbell

Replication Studies

The State of Applied Econometrics[3]

Structural Models

This is an obvious choice. What problems exist?

The Origins of Structure

foo

Invariant Conditionals

Invariant Conditionals

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation," 1117–65. https://doi.org/10.1093/qje/qjv015.Advance.

Athey, Susan, and Guido W Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation" 31 (2): 3–32.

Bisbee, James, Rajeev Dehejia, Cristian Pop-Eleches, and Cyrus Samii. 2017. "Local Instruments, Global Extrapolation: External Validity of the Labor Supply–Fertility Local Average Treatment Effect." *Journal of Labor Economics* 35 (S1): S99–S147. https://doi.org/10.1086/691280.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48 (2): 424–55. https://doi.org/10.1257/jel.48.2.424.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and 23
using instrumental variables in controlled trials." *Social Science &*