Identifying Causal Effects in Research and Evaluation: The Experimental and Quasi-Experimental Toolkit

Topic 1: The Experimental Ideal

Tim Maloney
Chief Economist
19-21 February 2019

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Introduction

- My background
  - Economics PhD, University of Wisconsin-Madison, 1983
  - University of Auckland (1991-2010), AUT (2010 +)
  - Currently Professor of Economics and Co-Director of the Centre for Social Data Analytics (CSDA) at AUT. Primary fields of interest are Labour Economics, Econometrics and Public Policy
- Purpose of this Workshop
  - To provide an overview of modern methods for identifying causal effects of interventions
  - A review of experimental and quasi-experimental methods – their potential and drawbacks

From Models to Methods

- Recommended Readings
  - Angrist and Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton University Press
  - Panhans and Singleton (2015) 'The Empirical Economist's Toolkit: From Models to Methods', Working Paper, Duke University
  - King and Nielsen (2019) 'Why Propensity Scores Should Not Be Used for Matching', Working Paper, Harvard and MIT

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

The Basic Multiple Regression Model

- Consider the Following Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + \ \beta_2 Z_i + u_i \qquad i = 1, \ldots, n$$

  - Causality is assumed to run from right to left.
  - Take a cross section of workers. We observe their wage rates Y_i, and assume that this wage variation is partly explained by education X_i and labour market experience Z_i. Everything else is relegated to the disturbance term u_i.
  - Suppose coefficient β_1 is taken as a measure of the return on this investment in education.
  - *Q*: What's potentially wrong with this interpretation?

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

The Selection Problem

- Start with a research question: *Do social welfare programmes improve wellbeing for the unemployed?*
- Suppose we have the following data from a random sample on overall life satisfaction for those on and off benefits:

Group	Sample Size	Mean Wellbeing	Standard Deviation
On Benefit	8,000	6.4	1.183
Off Benefit	120,000	8.0	1.265

Difference in means is 1.6. Reject the null hypothesis that the population means are the same at better than a 1% level (*t*-statistic > 100).

- *Q*: What's the problem with a causal interpretation that being on benefit makes you worse off?

A Formal Approach

- We have a binary variable for benefit receipt ($D_i = \{0,1\}$). Our outcome of interest is Y_i. The relevant question is how is Y_i *causally affected* by being on a benefit?

- In an ideal world, we'd observe wellbeing of an individual both on and off the benefit under the same circumstances.

$$\text{Potential Outcomes} \begin{cases} Y_{0i} & \text{if } D_i = 0 \\ Y_{1i} & \text{if } D_i = 1 \end{cases}$$

- Difference between the two outcomes is the *causal effect* of being on a benefit. Problem is that we never see both outcomes for a given individual!

A Formal Approach

- All we see is the observed difference in wellbeing in our sample:

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$$

- The observed difference in outcomes can be broken into two components:

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)}_{\text{Average Treatment Effect on the Treated}}$$

$$+ \underbrace{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{\text{Selection Bias}}$$

- Problem is that we observe the LHS of this expression, but can’t decompose into the ATE and Selection Bias.

A Formal Approach

- ***Random Assignment*** eliminates this selection problem. Suppose those eligible were randomly allocated to a benefit. By design D_i and Y_{0i} are independent. Two effects. ***First***, selection bias term disappears:

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)}_{\text{Average Treated Effect on the Treated}} + \underbrace{0}_{\text{Selection Bias}}$$

- ***Second***, latent term can be replaced with its observable counterpart:

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{\text{Average Treated Effect on the Treated}}$$

CASE STUDY: The Tennessee STAR Experiment

- Do smaller classes result in better academic achievement for students?
- Controls randomly allocated to classes with 22+/-, treated assigned to classes with 15+/-. Involved just under 12,000 first-year students (5 to 6 years old). Outcome of interest literacy and numeracy test results following random assignment (up to ages 8 or 9).
- Did the random assignment practice follow protocol?
- Was there evidence of 'balance' in the control and treated groups?

CASE STUDY: The Tennessee STAR Experiment

- The regression counterpart to a Randomised Control Trial could be written as a two-variable specification:

$$Y_i = \alpha + \beta D_i + u_i$$

where D_i is a dummy variable equal to one if treated; zero otherwise. The slope coefficient β is the Average Treatment Effect (ATE). Think of Y_i in this case as the percentile score on a cognitive achievement test.

- If we suspect random assignment may have been compromised, we can explicitly control for *k* factors through a multiple regression specification:

$$Y_i = \alpha + \beta D_i + \gamma_1 X_{1i} + \cdots + \gamma_k X_{ki} + u_i$$

Estimated Class-Size Effects on Test Scores
Tennessee STAR Experiment

CASE STUDY: The Tennessee STAR Experiment

Krueger (*QJE*, 1999, 114: 497-532)

Explanatory Variables	(1)	(2)	(3)	(4)
Small Class Dummy	**4.82 (2.19)**	**5.37 (1.26)**	**5.36 (1.21)**	**5.37 (1.19)**
White/Asian	--	--	0.53 (1.09)	0.31 (1.07)
Female	--	--	8.35 (1.35)	8.44 (1.36)
Free Lunch	--	--	4.48 (0.63)	4.39 (0.63)
White Teacher	--	--	--	-0.57 (2.10)
Teacher Experience	--	--	--	0.26 (0.10)
Master’s Degree	--	--	--	-0.51 (1.06)
School Fixed Effects	No	Yes	Yes	Yes
R^2	0.01	0.25	0.31	0.31

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Case Study: Tennessee STAR Experiment

Well-known and often-cited example of a successful RCT in policy evaluation.

What problems were raised about this particular study and RCTs in general?

- ***Costs***: Not trivial. $(US)12 million in mid-1980s. Equivalent of about $(NZ)45 million today.
- ***Delay in Findings***: Conducted after several years of design and contracting. Four-year follow-up period. Analysis took several years more. Ten years from design to published results!
- ***Impracticality***: In most situations, RCTs are impractical for a number of reasons. This could be universal rollout, or ethical issues about denying treatment to those who would almost surely benefit.

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Omitted Variable Bias and Conditional Independence Assumption (CIA)

- One of the most fundamental problems in regression analysis is ***Omitted Variable Bias***, and this has ramifications for our experimental design.

- Consider the following stylised regression model:

$$lnW_i = \alpha + \beta S_i + \gamma A_i + u_i$$

  Our goal is to estimate the rate of return β to years of education S_i. To do this, we need to hold constant innate ability A_i.

- Invoke the ***Conditional Independence Assumption (CIA)***. Allows a causal interpretation to our regression models. Says that, conditional on ability, selection bias disappears. In other words, once we control for ability, years of schooling are '*as good as randomly assigned*'.

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Omitted Variable Bias and Conditional Independence Assumption (CIA)

- Suppose we have no data on individual ability, and estimate the following restricted (short) regression model:

$$lnW_i = \alpha' + \beta' S_i + v_i$$

- Expected value of the estimated slope coefficient can be written:

$$E(b') = \beta + \underbrace{\gamma b_{32}}_{\text{bias term}}$$

- Equal to what we want (the true rate of return on education), plus a 'bias term' that depends on the return to ability and something close to the correlation between these explanatory variables:

$$\text{where } b_{32} = \frac{cov(S_i, A_i)}{var(S_i)} \text{ or } \hat{A}_i = c + b_{32} S_i$$

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Omitted Variable Bias and Conditional Independence Assumption (CIA)

- If $\gamma, b_{32} > 0$, then estimator for the return to education is *biased upward* in the restricted model (i.e., $E(b') > \beta$).

- *Q*: Under what circumstances would this bias term disappear?

- Formula for omitted-variable bias is one of the most important things to remember about regression analysis. If one claims that this bias doesn't exist (i.e., the equation is 'correctly specified'), this means that the model *may* have a causal interpretation. Sometimes unstated (but implicit), the researcher is invoking CIA. This is a very high bar for any analysis!

- SIDEBAR: Avoid 'Bad Controls.' They can make things worse!