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

Topic 3: Other Quasi-Experimental Methods

Tim Maloney
Chief Economist
19-21 February 2019

MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Introduction

- Previous topic showed that ***Matching*** was a quasi-experimental method that imitates an ***RCT***. It has a number of advantages and disadvantages over regression analysis.

- Both methods seek to estimate causal effects of interventions. Both methods fail to produce unbiased estimators if unobserved differences exist between the treated and controls (i.e., matching isn't complete, and regression estimators suffer from omitted variable bias).

- In this topic, we explore three other quasi-experimental approaches that can be used to overcome this potential problem.

Instrumental Variables

- Suppose we have the following simple regression model:

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

- Holding X_i constant, we want to isolate the treatment effect γ. The problem is that D_i and u_i are correlated. OLS estimators are biased.

- One solution is to use ***Instrumental Variables (IV)***.

- A valid instrument Z_i is both ***relevant*** (i.e., $corr(Z_i, D_i) \neq 0$) and ***exogenous*** (i.e., $corr(Z_i, u_i) = 0$).

Two Stage Least Squares (TSLS)

- ***First Stage*** involves decomposing the variation in D_i that is correlated with u_i and the part that isn't. Consider the linear regression below:

$$D_i = \pi_0 + \pi_1 Z_i + v_i$$

The values of π_0 and π_1 are unknown and need to be estimated.

- ***Second Stage*** involves replacing D_i with its fitted value $\widehat{D}_i$ from the first stage and estimating this regression:

$$Y_i = \alpha + \beta X_i + \gamma \widehat{D}_i + u_i$$

Conditional on certain assumptions, this TSLS estimator $\hat{\gamma}^{TSLS}$ is consistent.

Examples of IVs

- Commodity demand functions:

$$Q_i = \alpha + \beta P_i + u_i$$

Possible instrument: Rainfall

- Impact of market work on academic achievement:

$$GPA_i = \alpha + \beta H_i + u_i$$

Possible instrument: Roommate's employment

- Influence of military service on subsequent earnings:

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

Possible instrument: Vietnam War draft lottery

IV Application: Encouragement Design

- In many cases, noncompliance in treatment is inevitable. We still want to estimate the causal effects of treatment. This is where ***Encouragement Design Studies (EDS)*** can be used – randomisation meets instrumental variables!
- ***Example***: We offer budgeting and financial planning advice to clients. Individuals are randomly assigned to treatment $(T_i = 1)$and control $(T_i = 0)$ groups. Only 75% of those assigned to the treated group take up the service (i.e., ***compliers***). Remaining 25% known as ***defiers***.
- ***The Problem***: If we regressed some subsequent outcome on this take-up variable D_i, the estimated coefficient would potentially be biased by the self-selection:

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

IV Application: Encouragement Design

- Three choices over what to do with the 'defiers':
  1. Exclude from study ('Per Protocol' approach)
  2. Shift to controls ('As Treated' approach)
  3. Keep in the treated ('Intention-to-Treat' (ITT) Approach)

- With ITT, we could substitute random assignment for actual treatment in this regression:

$$Y_i = \alpha + \gamma^{ITT} T_i + u_i$$

- One simple thing to do is to divide this estimated coefficient by the probability of compliance to get ***Local Average Treatment Effect (LATE)***:

$$\hat{\gamma}^{LATE} = \hat{\gamma}^{ITT} \Big/ Prob(compliance)$$

IV Application: Encouragement Design

- ***The TSLS Solution***: Randomised assignment is the instrument. It's uncorrelated by design with the disturbance term (*the exogeneity condition*) and correlated with the take up (*the relevance condition).*

$$\text{First Stage:} \quad D_i = \pi_0 + \pi_1 T_i + v_i$$

$$\text{Second Stage:} \quad Y_i = \alpha + \gamma \widehat{D}_i + u_i$$

- This TSLS estimator $\hat{\gamma}^{TSLS}$ is as another measure of ***LATE***.

- ***Caution***: This is a highly condensed discussion of Encouragement Design. Lots on further complications. Encouragement might not exclude anyone from the service. Encouragement can come in multiple types and forms.

Regression Discontinuity

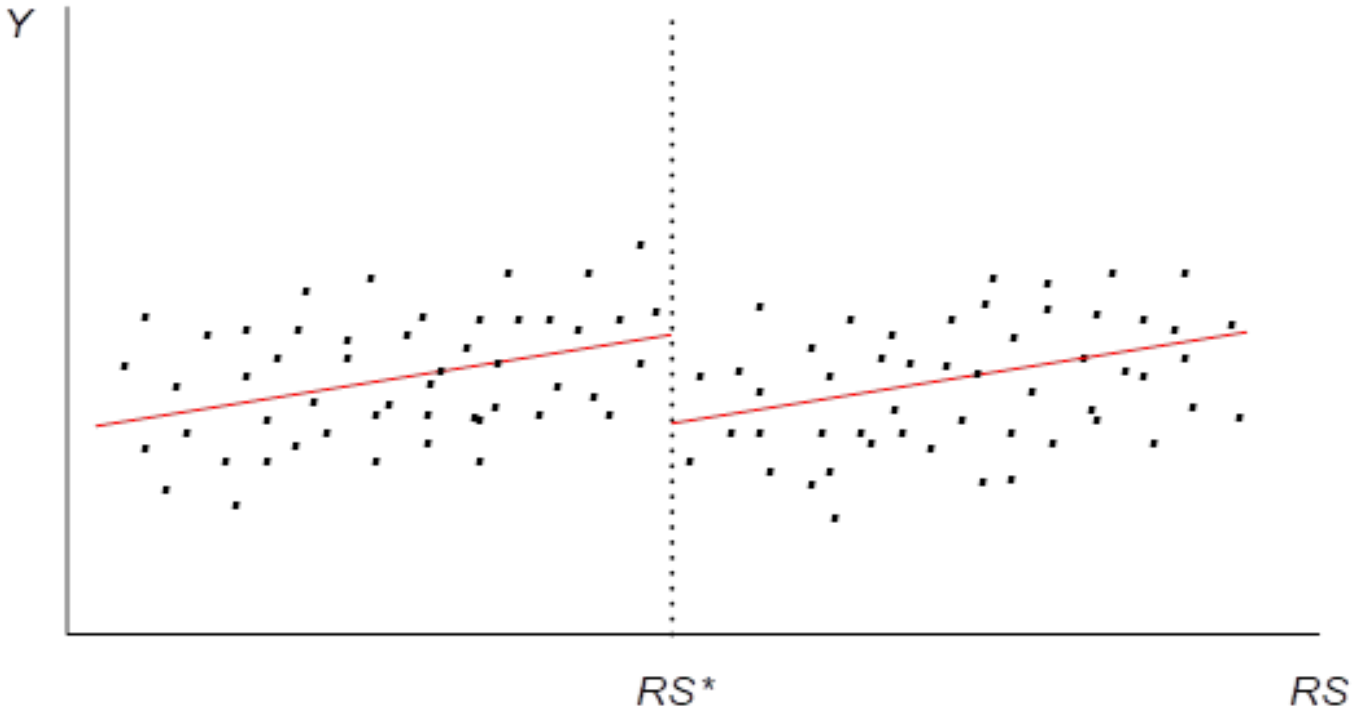
- Another quasi-experimental method in the toolkit for researchers seeking to understand causal impacts from observational data. Seen as having strong research design that imitates random assignment.

- ***Example***: Suppose we want to estimate the causal impact of Intensive Case Management (ICM) on some outcome.

- Imagine that treatment is based on a predicted risk score (i.e., clients are allocated to this programme if they a score above some threshold). Clients with a ***Risk Score (RS)*** greater than or equal to ***RS**** always get ICM, while those with a lower only get baseline services.

Regression Discontinuity

- ***Motivation for this Method***: Around the threshold, clients are assumed to be essentially identical. Only difference is this treatment cutoff. Outcome differences on either side of this threshold is the treatment effect.



MINISTRY OF SOCIAL DEVELOPMENT
TE MANATŪ WHAKAHIATO ORA

Regression Discontinuity

- ***Regression Specification***: This is a 'best case scenario' where the relationship between the variables is uniformly linear. All observations on risk scores identify the constant slope of the regression line β.

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

where

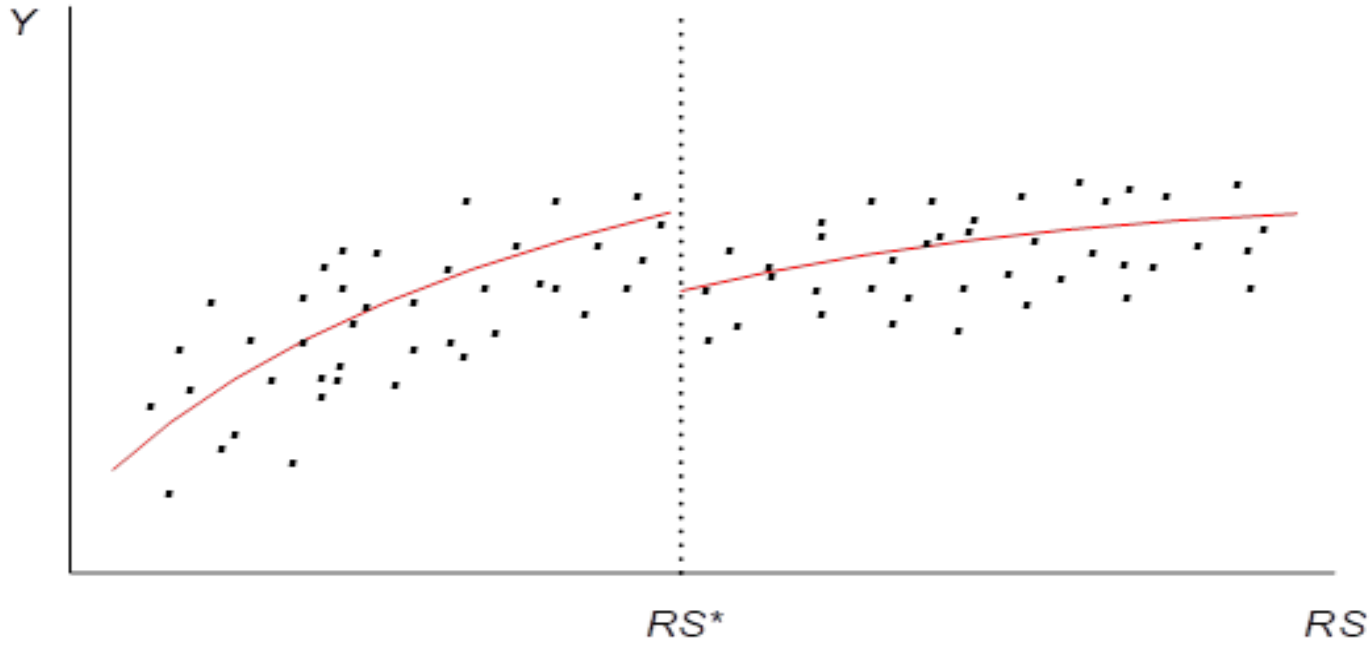
$$D_i = \begin{cases} 1 & if\ RS_i \geq RS^* \\ 0 & if\ RS_i < RS^* \end{cases}$$

The ***Average Treatment Effect*** is γ.

Regression Discontinuity

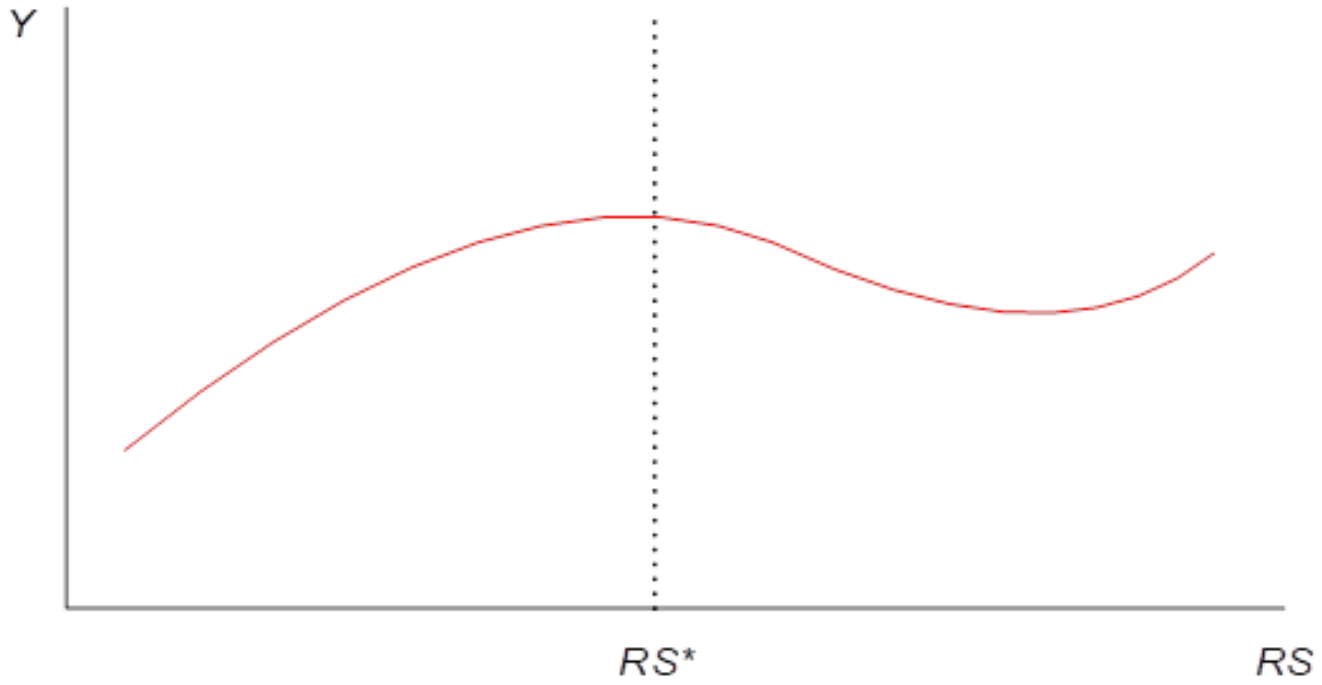
- In principle, nonlinear specifications are easily handled (e.g., a quadratic).

$$Y_i = \alpha + \beta_1 RS_i + \beta_2 RS_i^2 + \gamma D_i + u_i$$



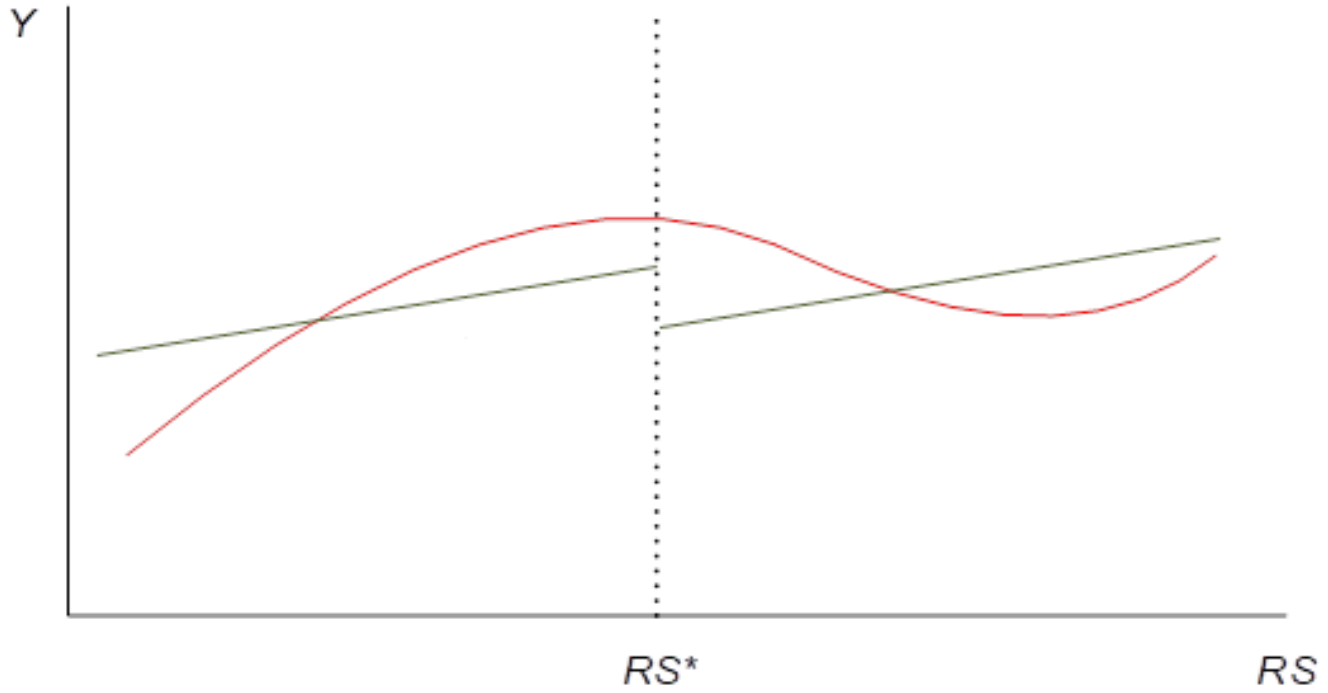
Regression Discontinuity

- Problems can arise if the unknown functional form has a complex shape like the one below.



Regression Discontinuity

- Imposing a linear specification with a constant slope and possible break at this threshold would give us a false indication of a treatment effect:



Regression Discontinuity

- ***Primary Threats*** to RD analysis include:

  1. Other interventions influencing the outcome might be triggered by this same threshold

  2. The underlying relationship isn't smooth

  3. Individuals (or programme administrators) can manipulate the risk score to push above (or below) the threshold

- Finally, quick word on ***Fuzzy Discontinuities***. Treatment doesn't hinge on the risk scores alone. Having a score above the threshold makes it *more likely* that you'll be treated. Brings us full-circle to encouragement design and instrumental variables, where threshold is the instrument.

Difference in Differences

- ***Example***: Consider a programme implemented in 2012 to bring teacher aides into the Auckland primary schools. Want to know the impact of this policy on academic performance.

- Assume all students take a standardised achievement test in literacy and numeracy at the end of every academic year. Combined scores range from 0 to 100.

- Could compare mean test scores in these schools in 2011 (pre-programme) and 2013 (post-programme).

- *Q*: What's the problem with this estimate of the programme's effect?

Difference in Differences

- Suppose we also have similar test scores for kids in Hamilton and Whangarei. These are possible controls. They didn't get the treatment but may have experienced similar other factors might influence test scores over this period.

  This estimator could be written:

$$DID = \left(\bar{Y}_{2013}^{AUCKLAND} - \bar{Y}_{2011}^{AUCKLAND}\right) - \left(\bar{Y}_{2013}^{OTHER} - \bar{Y}_{2011}^{OTHER}\right)$$

  Change in mean test scores in treated Auckland schools minus change in mean test scores in non-treated Hamilton and Whangarei schools.

Difference in Differences

- Let's use some numbers to drive this point home:

$\bar{Y}$	2011	2013
Treated Group of Schools in Auckland	74	83
Control Group of Schools in Hamilton and Whangarei	71	75

- Using just the change in test scores over time in Auckland schools, we might conclude that teacher aides caused a 9-point jump in test scores. Yet, test scores also improved over the same period by 4 points outside of Auckland. The DID estimator is a 5-point impact:

$$DID = (83 - 74) - (75 - 71) = 9 - 4 = 5$$

Difference in Differences

- Only aggregated data needed on mean test scores over the two years from these regions. DID methods could also use school-level or even individual-level data.
- ***Example***: Suppose we have data from 300 schools in Auckland and 140 schools outside of Auckland:

Y_i	2011	2013
Treated Group of Schools in Auckland		
School 1	66	75
School 2	76	87
…	…	…
School 300	69	82
Control Group of Schools in Hamilton and Whangarei		
School 1	64	70
School 2	68	71
…	…	…
School 140	73	80

Difference in Differences

- With school-level data, we can switch to regression analysis:

$$Y_{it} = \alpha + \beta P_{it} + \delta D_{it} + \gamma P_{it} D_{it} + u_{it}$$

where $$P_{it} = \begin{cases} 1 & if\ t = 2013\ (post\ period) \\ 0 & if\ t = 2011\ (pre\ period) \end{cases}$$

$$D_{it} = \begin{cases} 1 & if\ treated\ (Auckland\ school) \\ 0 & if\ control\ (non - Auckland\ school) \end{cases}$$

The DID estimator (i.e., the average treatment effect) is γ.

Difference in Differences

- Advantages of regression analysis in producing DID estimators:

1. Gives us standard error estimates

2. Allows us to control for other measurable differences across units

$$Y_{it} = \alpha + \beta P_{it} + \delta D_{it} + \gamma P_{it} D_{it} + \eta_1 X_{1it} + \cdots + \eta_k X_{kit} + u_{it}$$

3. Increases the precision of our DID estimator

Difference in Differences

- One final advantage of panel data for DID estimators. Suppose the model looks like this (where we could easily add time-varying covariates):

$$Y_{it} = \alpha + \beta P_{it} + \delta D_{it} + \gamma P_{it} D_{it} + \theta_i + u_{it}$$

where θ_i in a school-specific, time-invariant factor (AKA school fixed effects). There would be 440 of these school-specific intercepts to estimate in this specification.

Difference in Differences

- Start with the expression in 2013:

$$Y_{i2013} = \alpha + \beta P_{i2013} + \delta D_{i2013} + \gamma P_{i2013} D_{i2013} + \theta_i + u_{i2013}$$

and subtract the expression in 2011:

$$Y_{i2011} = \alpha + \beta P_{i2011} + \delta D_{i2011} + \gamma P_{i2011} D_{i2011} + \theta_i + u_{i2011}$$

We get:

$$Y_{i2013} - Y_{i2011} = (\alpha - \alpha) + \beta(P_{i2013} - P_{i2011}) + \delta(D_{i2013} - D_{i2011}) + \gamma(P_{i2013} D_{i2013} - P_{i2011} D_{i2011}) + (\theta_i - \theta_i) + (u_{i2011} - u_{i2011})$$

Difference in Differences

- First-differencing eliminates school fixed effects (and lots of other things!):

$$\Delta Y_i = \beta + \gamma D_i + \Delta u_i$$

End up with a simple expression, where γ is the DID estimator of the treatment effect that controls for:

- common trends
- school-specific, time-invariant factors
- and potentially observable time-varying factors about students and these schools.

Conclusions

- Tried to provide an overview of the basic methodological toolkit for formal programme evaluation.
- Random assignment is the 'gold standard' because it eliminates selection bias on both observables and unobservables, allowing us to produce the causal effects we want.
- RCTs aren't always feasible, ethical or easy to administer. As a result, quasi-experimental methods need to be considered. All of them have pros and cons. In the end, choosing the right tool for a particular application is critical. Multiple methods might help triangulate findings.