Lecture notes to Stock and Watson chapter 13

Quasi experiments

Tore Schweder

November 2009

Outline

- "Political experiments", Quasi experiments, experiments
- The focus is on potential intervention/regulation: "treatment"
- Clinical trials are (double) blind and randomized to allow treatment effects to be inferred
- Quasi experiments are "as-if" experiments
- Example: experiment to evaluate the effect of reducing class size in schools
- Threats to quasi experiments: external and internal validity
- Heterogeneity in causal effects - IV regression
- Make a good experimental protocol, follow it, use it in inference and report
- Meta analysis: collect relevant studies, integrate their results
- Summary

Quasi experiments

Experimental and Quasi-Experimental Designs for Generalized Causal Inference
by William R. Shadish , Thomas D. Cook , Donald T. Campbell,

Project STAR (SW: 13.4)

- Intervention: small class, or regular class with aid-in teacher
- Protocol: keep pupils in their class for 5 years, but...
- Design: randomize intervention and teacher within schools, assign new pupils randomly
- Performance measure: SAT test score

TABLE 13.1 Project STAR: Differences Estimates of Effect on Standardized Test Scores of Class Size Treatment Group

	Grade			
Regressor	K	1	2	3
Small class	13.90** (2.45)	29.78** (2.83)	19.39** (2.71)	15.59** (2.40)
Regular size with aide	0.31 (2.27)	11.96** (2.65)	3.48 (2.54)	−0.29 (2.27)
Intercept	918.04** (1.63)	1039.39** (1.78)	1157.81** (1.82)	1228.51** (1.68)
Number of observations	5786	6379	6049	5967

The regressions were estimated using the Project STAR Public Access Data Set described in Appendix 13.1. The dependent variable is the student's combined score on the math and reading portions of the Stanford Achievement Test. Standard errors are given in parentheses under the coefficients. **The individual coefficient is statistically significant at the 1% significance level using a two-sided test.

Project STAR cont.

- Statistical vs substantial effect
- how large is the effect?
  - relative to sd by level
  - relative to other effects (gender)
  - relative to mean score by level

TABLE 13.3 Estimated Class Size Effects in Units of Standard Deviations of the Test Score Across Students

	Grade			
Treatment Group	K	1	2	3
Small class	0.19** (0.03)	0.33** (0.03)	0.23** (0.03)	0.21** (0.03)
Regular size with aide	0.00 (0.03)	0.13** (0.03)	0.04 (0.03)	0.00 (0.03)
Sample standard deviation of test scores (s_Y)	73.70	91.30	84.10	73.30

The estimates and standard errors in the first two rows are the estimated effects in Table 13.1, divided by the sample standard deviation of the Stanford Achievement Test for that grade (the final row in this table), computed using data on the students in the experiment. Standard errors are given in parentheses under coefficients. **The individual coefficient is statistically significant at the 1% significance level using a two-sided test.

Project STAR cont.

Control variables: for individual pupils, teacher, school (random effect for class?)

TABLE 13.2 Project STAR: Differences Estimates with Additional Regressors for Kindergarten

Regressor	(1)	(2)	(3)	(4)
Small class	13.90** (2.45)	14.00** (2.45)	15.93** (2.24)	15.89** (2.16)
Regular size with aide	0.31 (2.27)	−0.60 (2.25)	1.22 (2.04)	1.79 (1.96)
Teacher's years of experience		1.47** (0.17)	0.74** (0.17)	0.66** (0.17)
Boy				−12.09** (1.67)
Free lunch eligible				−34.70** (1.99)
Black				−25.43** (3.50)
Race other than black or white				−8.50 (12.52)
Intercept	918.04** (1.63)	904.72** (2.22)		
School indicator variables?	no	no	yes	yes
$\overline{R}^2$	0.01	0.02	0.22	0.28
Number of observations	5786	5766	5766	5748

The regressions were estimated using the Project STAR Public Access Data Set described in Appendix 13.1. The dependent variable is the combined test score on the math and reading portions of the Stanford Achievement Test. The number of observations

Project STAR cont., a meta analysis

- Comparison with other studies
  - like with like!
  - testing for equality in effect across studies

TABLE 13.4 Estimated Effects of Reducing the Student–Teacher Ratio by 7.5 Based on the STAR Data and the California and Massachusetts Observational Data

Study	$\hat{\beta}_1$	Change in Student–Teacher Ratio	Standard Deviation of Test Scores Across Students	Estimated Effect	95% Confidence Interval
STAR (grade K)	−13.90** (2.45)	Small class vs. regular class	73.8	0.19** (0.03)	(0.13, 0.25)
California	−0.73** (0.26)	−7.5	38.0	0.14** (0.05)	(0.04, 0.24)
Massachusetts	−0.64* (0.27)	−7.5	39.0	0.12* (0.05)	(0.02, 0.22)

The estimated coefficient $\hat{\beta}_1$ for the STAR study is taken from column (1) of Table 13.2. The estimated coefficients for the California and Massachusetts studies are taken from the first column of Table 9.3. The estimated effect is the effect of being in a small class versus a regular class (for STAR) or the effect of reducing the student–teacher ratio by 7.5 (for the California and Massachusetts studies). The 95% confidence interval for the reduction in the student–teacher ratio is this estimated effect ± 1.96 standard errors. Standard errors are given in parentheses under estimated effects. The estimated effects are statistically significantly different from zero at the *5% level or **1% significance level using a two-sided test.

Threats to statistical conclusion validity

Rune Elvik (1998) Assessing the validity of evaluation research by means of meta-analysis; case illustrations from road safety research, Doctoral thesis, UiO

- lack of statistical power (small sample)
- Violation of statistical assumptions
  - units are dependent, distribution of response long-tailed,...
- Fishing
  - hypotheses not pre-specified, fishing out tests with small p-values
    - P-value plot Schweder and Spjøtvol (1982)
- Unreliable treatment implementation
  - Is the intervention policy under study really implemented in the quasi experiment?

Threats to internal validity

- Maturation
  - a genral time trend is mixed up with intervention effect in a before-after design
- Statistical regression
  - if intervention (new road lights where and when many accidents) is taken when the response variable by chance is bad, chance variation might be confused with random variation
- Self-selection
  - people prone to improve sign up for the experiment
  - firms positive to 6 hour day want to participate in the experiment
- Ambiguity of causal direction
  - endogeneity, omitted variables, ...

Hawthorne effect, experimental effect also in routine administration?

Being in an experiment affects the individuals studied.

- Placebo effect = Hawthorne effect
  - up to 75% of the effectiveness of anti-depressant medication is due to the placebo-effect rather than the treatment itself.
  - A May 7, 2002 article in The Washington Post titled "Against Depression, a Sugar Pill Is Hard to Beat" stated, "A new analysis has found that in the majority of trials conducted by drug companies in recent decades, sugar pills have done as well as – or better than – antidepressants.
  - a meta-analysis in the Journal of Psychiatric Research evaluated the effect of placebos for 12 weeks after an initial 6-8 weeks of successful therapy. They found that 79% of depressed patients receiving placebo remained well compared to 93% of those receiving antidepressants

    Wikipedia

Hawthorne effect?

- In the STAR study?
- "Forsøk i skolen" ("experiments" in Norwegian schools) are often driven by enthusiasts ....

Threats to external validity

are intervention effects stable in time and place?

- Interaction between selection and intervention
  - will the effects seen in the quasi experiment work on routine basis, or are the selected units in the experiment "special"?
  - enthusiastic teacher; placebo effect; ...
- Interaction of setting and intervention
  - will the regulation of the financial market tried out in "normal" times also work in troubled times?
- Interaction of history and intervention
  - will effects seen in one historic period also persist?

Causal effects - for the individual unit, or the collective?

- Homogeneous causal effect
  - the individual unit will have benefit β_{interv} of intervention
  - before-after design would have no variability in the difference!
- Heterogeneity in causal effect
  - the individual unit will have expected benefit β_{interv} of intervention
  - before-after design would have variability in the difference
  - $\widehat{\beta}_{interv}$ is unbiased
  - IV-regression consistent if treatment effects and instrument effects are uncorrelated

$$(13.11) : \widehat{\beta}_{interv}^{TSLS} \rightarrow \frac{\sigma_{ZY}}{\sigma_{ZX}} = \frac{E\left(\beta_{interv}\pi_1\right)}{E\left(\pi_1\right)}$$

Explicit modeling for quasi-experiments

and for observational studies in general

- Random variation in causal effect
  - β_{interv} should be modeled as a random variable. Stata and other software can handle.
- Selection processes
  - Model the selection process. Estimate the parameters of the selection process jointly with the effect process
- Endogeneous causality with instrument Z

$$\begin{array}{ccc} X & \leftrightarrow & Y \\ \uparrow & & \\ Z & & \end{array}$$

  - Model the complete system, and estimate jointly (ML)
  - If X is binary, use logistic regression for $Z \rightarrow X$

Summary

- Social experiments are not experiments - hard to learn!
- Quasi experiments helps!
- Be clear on protocol, design and performance measure!
- Quasi experiments are costly - do a pilot study!
- Be aware of treats to validity: internal, statistical conclusion, and external
- Use meta-analysis to summarize results from various studies
- Do SW: 13.3,7,8