Lecture notes XII

Experiments and Quasi-Experiments

SGW Ch. 13

Randomized controlled experiments randomly selects subjects from a population and then randomly assigns them either to a treatment group or control group

E.g. a drug is given to randomly selected individuals, while others get a "placebo" (harmless substitute)

Controlled experiments rare in economics. But similar methods used in program evaluation.

E.g. job training programs, R&D tax refund schemes, etc.

Let X_i be the treatment level, e.g. a dummy variable: $Y_i = 1$ if treated, 0 else.

The effect of the treatment on the outcome variable (Y_i) is given by a regression equation:

(1) $$Y_i = \beta_0 + \beta_1 X_i + u_i$$

i) If treatment level, X_i, is assigned completely at random $\Rightarrow$ independent of all other determinants of $Y_i \Rightarrow E(u_i | X_i) = 0$
$\Rightarrow$ no omitted variable bias

ii) The causal effect of the treatment is $\Delta = E(Y_i | X_i = x) - E(Y_i | X_i = 0)$

$\Rightarrow \Delta = \beta_1 x$. Δ is called the treatment effect.

Even in controlled experiments real-world experiments deviate from the idealized experiment, which may cause

$$E(u_i \mid X_i) \neq 0$$

Reasons for failure of complete randomization:

i) The assignment and participation depend on the subject's preferences or motivation

ii) Partial compliance $\rightarrow$ a subject may be assigned to treatment but does not follow the instructions (e.g. does not take the drug)

iii) Subjects may drop out after initially participating $\rightarrow$ <u>Attrition</u>.

Attrition may be endogenous $\Rightarrow$ bias estimates of β_1 using OLS

Regression based estimates of treatment effects

Let $W_{1i}, \ldots, W_{ri}$ denote r ~~[illegible]~~ variables that characterize the subject i prior to the treatment, i.e. not affected by the treatment →

(2) $$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i$$

- X_i is the treatment variable
- W_{ki} are the control variables

Control variables are not variables of interest per se, but are included to avoid omitted variable bias in the estimate of β_1.
May also lead to a more efficient estimator of β_1.

Standard OLS assumption: $E(u_i | X_i, W_{1i}, \ldots, W_{ri}) = 0$, i.e. conditional mean-zero assumption

Weaker assumption: $E(u_i | X_i, W_{1i}, \ldots, W_{ri})$
$= E(u_i | W_{1i}, \ldots, W_{ri})$, i.e. conditional mean independence

Example: Conditional random assignment. That is, X_i may depend on the control variables, W_{ki}. However, given $W_{1i}, \ldots, W_{ri}$, u_i is independent of X_i.

- E.g. job training programs distributed to persons with low education.

Differences-in-Difference Estimator

Experimental data are often panel data: The same unit i is observed before and after the experiment.

Let ΔY_i be the change in the value of Y_i after the treatment compared to before the treatment

(7) $$\Delta Y_i = \beta_0 + \beta_1 X_i + u_i$$

The β_1-estimate using OLS is the Differences-in-difference estimator (DID)

Can also include control variables $W_{1i}, \ldots, W_{ri}$ in (7)

Main motivation for DID estimator is that the treatment may be correlated with pre-treatment value of Y_i.

The DID estimator removes fixed effects, d_i, that may be correlated with X_i. : $Cov(d_i, X_i) \neq 0$

$$Y_{1i} = \beta_{10} + d_i + U_{1i}$$

$$Y_{2i} = \beta_{20} + \beta_1 X_i + d_i + V_{2i}$$

$$\Delta Y_i = \beta_{20} - \beta_{10} + \beta_1 X_i + U_{2i} - U_{1i}$$

$$\equiv \beta_0 + \beta_1 X_i + U_i$$

Quasi-experiments

- randomness is introduced by variations in individual circumstances that make it appear <u>as if</u> the treatment is randomly assigned.

If the "as if" random variation only partly determines X_i, so that $Cov(X_i, U_i) \neq 0$, the exogenous part of the variation in X_i may be identified using instrumental variables

Quasi-experimental estimates of the effect of class size on achievement in Norway

Edwin Leuven Hessel Oosterbeek Marte Rønning

Introduction

- The effect of class size on achievement is still an unresolved issue in education research.
- Endogeneity problems may severely bias naive OLS-estimates of the class size effect
- Exogenous sources of variation in class size are key for a credible identification of the class size effect.
- Various recent studies acknowledge this and apply ingenious identification methods.
- Still no definite conclusion about the magnitude and even the sign of the class-size effect:
  - Recent experimental and quasi-experimental studies find negative effects (Angrist & Lavy 1999, Krueger 1999).
  - Some, which also exploit credible identification strategies, fail to confirm these findings (f.e. Hoxby 2000, Hanushek et al. 1996).

Table 2: Regression of actual class size on observables

	(1)	(2)
Girl	0.019	0.019
	(0.029)	(0.026)
Age	−0.064	0.071
	(0.051)	(0.046)
ln(family income)	0.430	0.168
	(0.042)∗∗∗	(0.025)∗∗∗
Education mother (years)	0.047	0.012
	(0.008)∗∗∗	(0.007)∗
Education father (years)	0.104	0.036
	(0.009)∗∗∗	(0.006)∗∗∗
1st or 2nd generation immigrants	1.313	−0.223
	(0.155)∗∗∗	(0.138)
Parents non-cohabiting	0.244	−0.017
	(0.045)∗∗∗	(0.040)
ln(pop. size school district)		1.211
		(0.068)∗∗∗
ln(rural pop. size school district)		0.434
		(0.148)∗∗∗
Year=2002		0.097
		(0.130)
adj. R-squared	0.0192	0.2103
N	115,084	115,084
N schools	1,000	1,000

Estimating class size effects

Assume that achievement of student i (y_i) is generated by the following equation:

$$y_i = x_i'\beta + w_{s(i)}'\alpha + \delta \cdot \overline{cs}_{s(i)} + \eta_{s(i)} + \psi_{t(i)} + \epsilon_i \qquad (1)$$

- x_i is a vector of observable attributes of the student and his parents
- $w_{s(i)}$ a vector of observable school and teacher characteristics
- $\overline{cs}_{s(i)}$ is student $i's$ class size
- $\eta_{s(i)}$ a school effect
- $s(i)$ identifies the school of pupil i
- $\psi_{t(i)}$ is an effect for the year in which student i is in her final year of lower secondary school (2001/2 or 2002/3)
- ϵ_i is all other determinants of achievement such as unobserved attributes of the student, parents and community.
- The coefficient of interest is δ, the class size effect.

Test score

- Exam results from national tests measured in the end of lower secondary school, grade 9.
  - 2001/2002 and 2002/2003.
  - Mathematics, English, Norwegian1 (Hovedmål) and Norwegian2 (Sidemål)
- Each student is only examined in one of the subjects, decided centrally shortly before the exam
  - the students and their teachers cannot prepare the whole school year for a particular exam
- The grading scale goes from 1 to 6 where 1 is fail and 6 is top score.
- Average scores for each of the four subjects are around 3.5 with standard deviations almost equal to 1.

Table 3: OLS. Dependent variable is exam results in 9th grade, unit of observations is the individual student's exam result

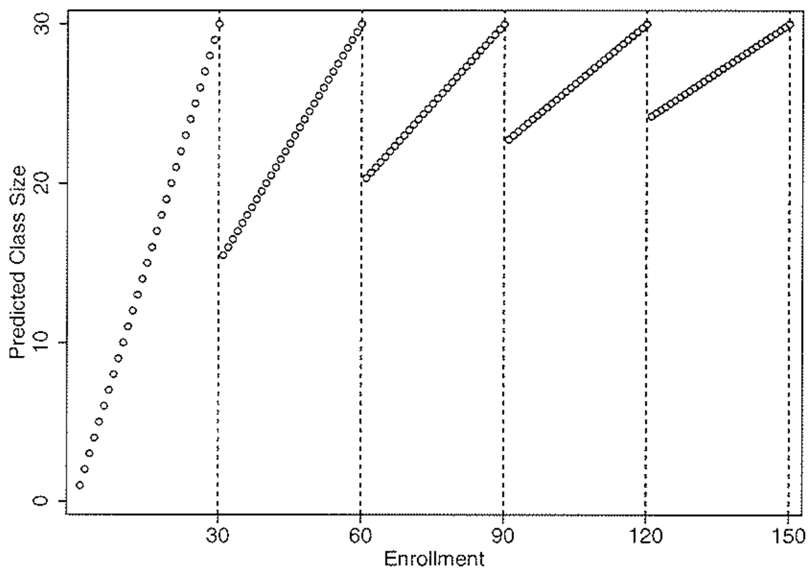
	Mathematics			Languages (pooled)		
	(1)	(2)	(3)	(4)	(5)	(6)
Average class size grade 7-9	0.005	−0.006	−0.005	0.004	−0.006	−0.005
	(0.003)*	(0.002)**	(0.002)**	(0.002)**	(0.002)***	(0.002)***
Individual characteristics						
Girl		0.018	0.018		0.506	−0.506
		(0.012)	(0.012)		(0.010)***	(0.0010)***
Age (years)		0.071	0.072		0.037	0.036
		(0.019)***	(0.019)***		(0.014)**	(0.014)**
ln(family income)		0.180	0.181		0.144	0.144
		(0.009)***	(0.009)***		(0.008)***	(0.008)***
Education mother (years)		0.075	0.075		0.057	0.057
		(0.002)***	(0.002)***		(0.002)***	(0.002)***
Education father (years)		0.066	0.066		0.049	0.049
		(0.002)***	(0.002)***		(0.002)***	(0.002)***
1st or 2nd generation immigrants		−0.446	−0.437		−0.232	−0.225
		(0.032)***	(0.032)***		(0.025)***	(0.025)***
Parents have different address		−0.345	−0.342		−0.189	−0.188
		(0.013)***	(0.013)***		(0.009)***	(0.009)***
ln(pop. size school district)		0.017	0.028		0.011	0.015
		(0.008)**	(0.008)***		(0.005)**	(0.006)**
ln(rural pop. size school district)		−0.018	−0.024		−0.011	−0.016
		(0.016)	(0.016)		(0.010)	(0.010)
School characteristics:						
Average teacher education (years)			0.042			0.070
			(0.055)			(0.042)*
Average teacher experience			0.016			0.011
			(0.003)***			(0.002)***
Fract. of female teachers			−0.097			0.059
			(0.110)			(0.077)
Fract. teachers with a temp. contract			0.067			0.022
			(0.088)			(0.065)
Combined school			0.052			0.028
			(0.027)**			(0.022)
R-squared	0.001	0.164	0.167	0.011	0.171	0.173
N pupils	38,045	38,045	38,045	77,039	77,039	77,039
N schools	755	755	755	958	958	958

Note: Standard errors are heteroscedasticity robust and corrected for school level clustering. All regressions include a control for year of observation; the laguage regressions also inlcude dummies for the language the student took the exam in.

1st approach: maximum class size rule as predictor for endogenous actual class size (AL, 1999)

- Lower secondary schools in Norway are subject to maximum class size rules of 30 students per class.
- This rule creates a discontinuous relation between enrollment and class size.
  - Just above multiples of 30 class size drops substantially.

Rule



1st approach (cont'd)

- Predicted class size in grade 7 is used to instrument average class size in lower secondary school.
  - Enrollment (and therefore predicted class size) in subsequent grades is potentially endogenous

Results: teacher hours

	All	DS±5
	First Stage[1]	
Predicted Class size	-0.671	-0.413
	(0.080)***	(0.106)***
	Second Stage - Mathematics	
Teacher Hours/Pupil	0.005	-0.012
	(0.006)	(0.020)
	Second Stage - Language	
Teacher Hours/Pupil	-0.002	-0.010
	(0.003)	(0.007)

IV Regression wit heterogeneous causal effects

Assume:

$$Y_i = \beta_{0i} + \beta_{1i} X_i + u_i$$

$$= \beta_0 + \beta_1 X_i + (\beta_{0i} - \beta_0) + (\beta_{1i} - \beta_1) X_i + u_i$$

$$= \beta_0 + \beta_1 X_i + \omega_i, \quad \omega_i = \beta_{0i} - \beta_0 + (\beta_{1i} - \beta_1) X_i + u_i$$

- $E(\omega_i) = 0$
- $E(\omega_i | X_i) \neq 0$

Assume Z_i is a valid instrument for X_i:

$$X_i = \Pi_{0i} + \Pi_{1i} Z_i + V_i$$

The TSLS estimator of β_1 is then

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, X)} = \frac{E(\beta_{1i} \Pi_{1i})}{E(\Pi_{1i})}$$

is called local average treatment effect (LATE), while $E(\beta_{1i})$ is called average treatment effect.