ECON4150 - Introductory Econometrics

Lecture 14: Quasi-Experiments

Monique de Haan
(moniqued@econ.uio.no)

Stock and Watson Chapter 13

Lecture outline

- What are quasi-experiments?
- Difference-in-differences
- Using quasi-experimental variation as instrument
- Heterogeneous effects in (quasi-)experiments
    - Heterogeneous effects & OLS
    - Heterogeneous effects & 2SLS
- Regression discontinuity design

Terminology: experiments and quasi-experiments

Previous lecture we discussed:

Experiments: designed and implemented consciously by human researchers.

- An **experiment** randomly assigns subjects to treatment and control groups (think of clinical drug trials)

This week we will discuss:

Quasi-experiments or natural experiments have a source of randomization that is “as if” randomly assigned.

- This variation was however not the result of an explicit randomized treatment and control design.

Different Types of Quasi Experiments

There are 2 types of quasi experiments

1. Whether an an individual (entity) receives treatment is “as if” randomly assigned, possible conditional on certain characteristics
   - For example a new policy measure that is implemented in one but not in another area, whereby the implementation is “as” if randomly assigned.
2. Whether an an individual (entity) receives treatment is partially determined by another variable that is “as if” randomly assigned.
   - The variable that is “as if” randomly assigned can then be used as an instrumental variable in a 2SLS regression analysis.

Quasi experiment with conditional “as if” randomization

- If the treatment in a quasi-experiment is “as if” randomly assigned, conditional on observed characteristics W....
- ....we can estimate the treatment effect by OLS while including W as control variable.
- This is similar as with an experiment with conditional randomization.
- We can obtain an unbiased effect of the treatment based on the conditional mean independence assumption...
- ...by estimating the following equation by OLS

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i \qquad \textit{with} \qquad E\left[u_i | X_i, W_i\right] = E\left[u_i | W_i\right]$$

Difference-in-Differences (DiD)

- What if the treatment in a quasi-experiment is “as if” randomly assigned, conditional on unobserved characteristics?
- If these differences in unobserved characteristics are time-invariant,...
- ...and we observe outcomes for the treatment & control group before & after the treatment ...
- ... we can use a method called difference-in-differences

DID: two groups & two time periods

- Two groups: treatment group ($g = Tr$) and control group ($g = C$)
- Two time periods: Before ($t = 0$) and after ($t = 1$)
- Potential outcomes:

$Y_{igt}(1)$ outcome for entity i in group g in period t in case of **treatment**

$Y_{igt}(0)$ outcome for entity i in group g in period t in case of **no treatment**

- We assume additive structure for mean potential outcome in case of no treatment (heart of dif-in-dif set-up):

$$E\left[Y_{igt}(0)\right] = \alpha_g + \lambda_t$$

- α_g= time-invariant group effect
- λ_t= time effect which is constant across groups

DID: two groups & two time periods

- A treatment takes place in treatment group but not in control group
- Suppose we observe outcomes before ($t = 0$) and after ($t = 1$) the treatment (panel data)
- Let the treatment indicator X_{gt}:
  - equal 1 for treatment group ($g = Tr$) in the second period ($t = 1$)
  - equal 0 otherwise
- We can write the observed outcome as a function of the potential outcomes

$$Y_{igt} = Y_{igt}(1) \cdot X_{gt} + Y_{igt}(0) \cdot (1 - X_{gt})$$

- Taking expectations and rewriting gives

$$\begin{aligned} E[Y_{igt}] &= E[Y_{igt}(1) - Y_{igt}(0)] \cdot X_{gt} + E[Y_{igt}(0)] \\ &= \beta X_{gt} + \alpha_g + \lambda_t \end{aligned}$$

With the average causal effect of the treatment: $E[Y_{igt}(1) - Y_{igt}(0)] = \beta$

DID: two groups & two time periods

$$E[Y_{igt}] = \beta \cdot X_{gt} + \alpha_g + \lambda_t$$

	Before ($t = 0$)	**After ($t = 1$)**
Treatment group ($g = Tr$)	$E[Y_{i\,Tr\,0}] = \alpha_{Tr} + \lambda_0$	$E[Y_{i\,Tr\,1}] = \beta + \alpha_{Tr} + \lambda_1$
control group ($g = C$)	$E[Y_{i\,C\,0}] = \alpha_C + \lambda_0$	$E[Y_{i\,C\,1}] = \alpha_C + \lambda_1$

- Comparing outcomes for treated and controls after intervention:

$$E[Y_{i\,Tr\,1}] - E[Y_{i\,C\,1}] = \beta + (\alpha_{Tr} - \alpha_C)$$

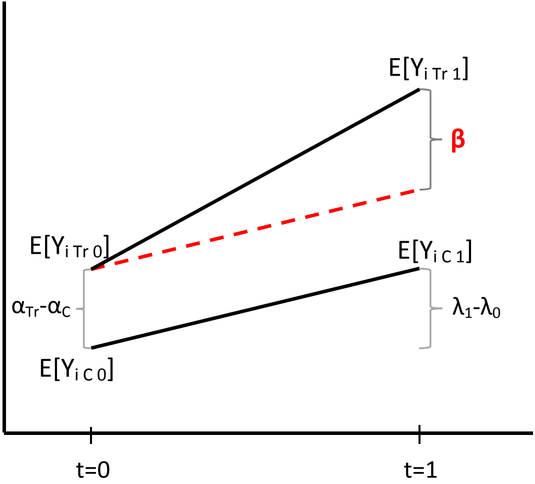
- Comparing outcomes for treated before and after treatment:

$$E[Y_{i\,Tr\,1}] - E[Y_{i\,Tr\,0}] = \beta + (\lambda_1 - \lambda_2)$$

- Instead subtract change for controls from change for treated:

$$\begin{aligned} DID &= (E[Y_{i\,Tr\,1}] - E[Y_{i\,Tr\,0}]) - (E[Y_{i\,C\,1}] - E[Y_{i\,C\,0}]) \\ &= ((\beta + \alpha_{Tr} + \lambda_1) - (\alpha_{Tr} + \lambda_0)) - ((\alpha_C + \lambda_1) - (\alpha_C + \lambda_0)) \\ &= (\beta + \lambda_1 - \lambda_0) - (\lambda_1 - \lambda_0) \\ &= \beta \end{aligned}$$

DID: two groups & two time periods



Common trend assumption: In absence of intervention, the treatment group would have had the same trend in *Y* as the control group.

DID: two groups & two time periods

Example: Card & Krueger (AER, 1994)

- What is the effect of increase in minimum wage on employment?
- Prediction economic theory: a rise in the minimum wage leads perfectly competitive employers to cut employment.
- Card and Krueger investigate effect of increase in minimum wage from $4.25 to $ 5.05 in New Jersey on April 1, 1992.
- Data on 410 fast-food restaurants (Burger King, Wendy's,...):
    - in New Jersey (**treatment group**)
    - and Pennsylvania (**control group**)
    - in February/March 1992 (**before**)
    - and in November/December 1992 (**after**)

DID: two groups & two time periods

Example: Card & Krueger (AER, 1994)

Data on fast food restaurants:

```
  obs:           820
 vars:             3                            21 Feb 2013 16:06
 size:         7,380
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
state           byte   %8.0g                    1 if New Jersey; 0 if Pennsylvania
employment      float  %9.0g                    employment (fte) in fast food restaurant
time            float  %9.0g                    0 if before, 1 if after
-------------------------------------------------------------------------------
Sorted by:

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       state |       820    .8073171    .3946469          0          1
  employment |       794    21.02651    9.422746          0         85
        time |       820          .5    .5003052          0          1
```

DID: two groups & two time periods

Example: Card & Krueger (AER, 1994)

Mean employment by state, time period:

-> state = Pennsylvania, time = before

Variable	Obs	Mean	Std. Dev.	Min	Max
employment	77	23.33117	11.85628	7.5	70.5

-> state = Pennsylvania, time = after

Variable	Obs	Mean	Std. Dev.	Min	Max
employment	77	21.16558	8.276732	0	43.5

-> state = NewJersey, time = before

Variable	Obs	Mean	Std. Dev.	Min	Max
employment	321	20.43941	9.106239	5	85

-> state = NewJersey, time = after

Variable	Obs	Mean	Std. Dev.	Min	Max
employment	319	21.02743	9.293024	0	60.5

DID: two groups & two time periods
Example: Card & Krueger (AER, 1994)

- Y_{igt} is employment in restaurant i in state g at time t:

	Before ($t = 0$)	**After ($t = 1$)**
New Jersey ($g = Tr$)	$\widehat{E[Y_{i\,Tr\,0}]} = 20.44$	$\widehat{E[Y_{i\,Tr\,1}]} = 21.03$
Pennsylvania ($g = C$)	$\widehat{E[Y_{i\,C\,0}]} = 23.33$	$\widehat{E[Y_{i\,C\,1}]} = 21.17$

- $\hat{\beta}^{DID} = (21.03 - 20.44) - (21.17 - 23.33) = 2.75$
- Counter-intuitive result: Employment increased as consequence of increase in minimum wage
- Note: small change in NJ, but downward trend in PA
- Common trend assumption: In absence of intervention employment in NJ would have had same downward trend as PA

DiD: general set-up

- DID-estimator can be obtained by estimating this equation by OLS

$$Y_{igt} = \beta_0 + \beta_1 \cdot X_{gt} + \beta_2 G_i + \beta_3 D_t + u_{igt}$$

  - with $G_i = 1$ for the treatment group and 0 for the control group,
  - $D_t = 1$ if after & $D_t = 0$ if before,
  - and $X_{gt} = G_i \times D_t = 1$ if treated and 0 otherwise

- If we observe outcomes at $t = 0$ & $t = 1$ for each i, we can also take first differences:

$$\begin{aligned} Y_{ig1} - Y_{ig0} = & \quad (\beta_0 - \beta_0) - \beta_1 \cdot (X_{g1} - X_{g0}) + \beta_2(G_i - G_i) + \beta_3(D_1 - D_0) + (u_{ig1} - u_{ig0}) \\ \triangle Y_{ig} = & \quad \beta_1 X_g + \beta_3 + \triangle u_{ig} \end{aligned}$$

- Main assumption: In absence of intervention treatment and control groups would have **common trends**

$$E\left[u_{igt}|X_{gt}, G_i, D_t\right] = E\left[u_{igt}|G_i, D_t\right] \qquad or \qquad E\left[\triangle u_{ig}|X_g\right] = 0$$

DID: general set-up, two groups

Example: Card & Krueger (AER, 1994)

$$Y_{igt} = \beta_0 + \beta_1 \cdot X_{gt} + \beta_2 G_i + \beta_3 D_t + u_{igt}$$

```
1 . gen treatment= state* time

2 . regress employment treatment state time, robust

  Linear regression                                  Number of obs =        794
                                                        F(  3,   790) =       1.40
                                                        Prob > F      =     0.2404
                                                        R-squared     =     0.0074
                                                        Root MSE      =     9.4056
```

employment	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
treatment	2.753606	1.795451	1.53	0.126	-.7708128	6.278024
state	-2.891761	1.438696	-2.01	0.045	-5.71588	-.067642
time	-2.165584	1.641212	-1.32	0.187	-5.387236	1.056067
_cons	23.33117	1.345741	17.34	0.000	20.68952	25.97282

DID: general set-up, two groups

- We don't need panel data to apply difference-in-differences
- We can use repeated cross-sections to estimate

$$Y_{igt} = \beta_0 + \beta_1 \cdot X_{gt} + \beta_2 G_i + \beta_3 D_t + u_{igt}$$

- Repeated cross-section: a collection of cross-sectional data sets, where each cross-section corresponds to a different time period.
- Additional assumption: composition of treatment and control groups do not change over time.
- Minimum wage example: you don't need to observe exactly the same fast food restaurants in $t = 0$ & $t = 1$ ...
- ....as long as the sample of restaurants in New Jersey & Pennsylvania at $t = 0$ & $t = 1$ are random draws from the same population of fast food restaurants.

Natural experiment: Draft eligibility, veteran status & earnings

- The increase in the minimum wage in New Jersey (but not Pennsylvania) is an example whereby the treatment is "as if" randomly assigned
- We now turn to an example where the "as if" randomization partially affects the treatment.
- **Research question:** Does serving in the military affect future earnings?
- **Treatment of interest:** veteran status
- **Natural experiment:** During the Vietnam War draft *eligibility* was determined by a national lottery system based on birthdays
  - men with a low lottery number were *eligible* to be drafted into the military
  - men with a high lottery number were *not eligible* to be drafted.

Natural experiment: Draft eligibility, veteran status & earnings

- Serving in the military might have a positive effect on future earnings (training)
- Serving in the military could also have a negative effect (psychological problems/bad health)
- Estimating the effect of veteran status on earnings by OLS will likely give a biased estimate, because veteran status is correlated with (unobserved) individual characteristics.
- Draft lottery during Vietnam war randomly assigned draft eligibility.
- Draft eligibility partially determines actual military service.
- Angrist (AER, 1990) used the draft eligibility as an instrumental variable to estimate the causal effect of veteran status on earnings.

Natural experiment: Draft eligibility, veteran status & earnings

Sample of about 13500 men born in 1950

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad X_i = \pi_0 + \pi_1 Z_i + v_i$$

- Y_i is earnings observed in 1981
- $X_i = 1$ if served in the military and 0 if not
- $Z_i = 1$ if individual was draft-eligible in 1970
  - randomly assigned a low lottery number (below cut-off)
- $Z_i = 0$ if individual was not draft-eligible in 1970
  - randomly assigned high lottery number (above cut-off)

Natural experiment: Draft eligibility, veteran status & earnings

Draft eligibility is a valid instrument if

Instrument exogeneity: $Cov(Z_i, u_i) = 0$

1. **Independence:** Draft eligibility is uncorrelated with unobserved characteristics that affect earnings
   - Draft eligibility was randomly assigned by a national lottery and therefore uncorrelated with (unobserved) characteristics
2. **Exclusion restriction:** Draft eligibility does not have a direct effect on earnings, only effect is via veteran status.
   - Assumption might be violated if men with low draft lottery number stayed in school longer to avoid being drafted.

Instrument Relevance: $Cov(Z_i, X_i) \neq 0$

- Draft eligibility should affect probability of serving in the military.
- Can be checked by running first stage regression & testing $H_0 : \pi_1 = 0$

Natural experiment: Draft eligibility, veteran status & earnings

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad X_i = \pi_0 + \pi_1 Z_i + v_i$$

	First stage	2SLS
Dependent variable:	Served in Military	Earnings ($1000)
Served in military		-2.741**
		(1.324)
Eligible for draft (lottery nr. below cut-off)	0.159***	
	(0.040)	
First stage F-statistic	15.80	

Note: ** significant at 5% level, *** significant at 1% level

The Wald estimator

There is an alternative way of computing the instrumental variable estimator:

- Recall the formula of the instrumental variable estimator

$$
\begin{aligned}
\hat{\beta}_{IV} &= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})(Z_i-\bar{Z})}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Z_i-\bar{Z})} \\
&= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})(Z_i-\bar{Z})/\frac{1}{n-1}\sum_{i=1}^{n}(Z_i-\bar{Z})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Z_i-\bar{Z})/\frac{1}{n-1}\sum_{i=1}^{n}(Z_i-\bar{Z})^2} \\
&= \frac{S_{ZY}/S_Z^2}{S_{ZX}/S_Z^2}
\end{aligned}
$$

- $\frac{S_{ZY}}{S_Z^2}$ is the OLS estimator when regressing Y_i on Z_i
- $\frac{S_{ZX}}{S_Z^2}$ is the OLS estimator when regressing X_i on Z_i

The Wald estimator

When the instrument Z_i is binary:

- Estimating $Y_i = \gamma_0 + \gamma_1 Z_i + \varepsilon_i$ by OLS gives the following differences estimator
$$\widehat{\gamma}_1 = \frac{S_{ZY}}{S_Z^2} = E\,[\widehat{Y_i|Z_i = 1]} - E\,[\widehat{Y_i|Z_i = 0]}$$
- Estimating $X_i = \pi_0 + \pi_1 Z_i + u_i$ by OLS gives the following differences estimator
$$\widehat{\pi}_1 = \frac{S_{ZX}}{S_Z^2} = E\,[\widehat{X_i|Z_i = 1]} - E\,[\widehat{X_i|Z_i = 0]}$$
- We therefore have that the IV estimator equals the so called Wald estimator

$$\hat{\beta}_{IV} = \frac{S_{ZY}/S_Z^2}{S_{ZX}/S_Z^2} = \frac{E\,[\widehat{Y_i|Z_i = 1]} - E\,[\widehat{Y_i|Z_i = 0]}}{E\,[\widehat{X_i|Z_i = 1]} - E\,[\widehat{X_i|Z_i = 0]}}$$

The Wald estimator

$\widehat{\pi}_1$	$\widehat{\gamma}_1$	Wald estimate
$E[\widehat{X_i \mid Z_i = 1]} - E[\widehat{X_i \mid Z_i = 0]}$	$E[\widehat{Y_i \mid Z_i = 1]} - E[\widehat{Y_i \mid Z_i = 0]}$	$\widehat{\beta}_{IV} = \frac{E[\widehat{Y_i \mid Z_i=1]} - E[\widehat{Y_i \mid Z_i=0]}}{E[\widehat{X_i \mid Z_i=1]} - E[\widehat{X_i \mid Z_i=0]}}$
0.159 (0.040)	-0.436 (0.211)	-2.741 (1.324)

- $\frac{-0.436}{0.159} = -2.741$
- Using a natural experiment, the draft lottery, as instrumental variable we find that serving in the military reduces future earnings by 2741 dollar.
- Note: this is based on the assumption of a homogenous treatment effect: $\beta_{i1} = \beta_1$
- We assume that the effect of serving in the military on earnings is the same for all men.

Heterogeneous treatment effect & OLS

- When we write
$$Y_i = \beta_0 + \beta_1 X_i + u_i$$
we assume that the effect of a unit change X_i equals β_1 for all i.

- What if $Y_i = \beta_0 + \beta_{1i} X_i + u_i \quad \textit{with} \quad \beta_{1i} \neq \beta_1$

- If we have a (natural) experiment where the treatment X_i is ("as if") randomly assigned and we estimate the effect of X_i on Y_i by OLS we get

$$\begin{aligned}
\widehat{\beta}_{OLS} = \frac{S_{YX}}{S_X^2} \longrightarrow \frac{Cov(Y_i, X_i)}{Var(X_i)} &= \frac{Cov(\beta_0 + \beta_{1i} X_i + u_i, X_i)}{Var(X_i)} \\
&= \frac{Cov(\beta_0, X_i) + Cov(\beta_{1i} X_i, X_i) + Cov(u_i, X_i)}{Var(X_i)} \\
&= \frac{0 + E[\beta_{1i}] Cov(X_i, X_i) + 0}{Var(X_i)} \\
&= E[\beta_{1i}]
\end{aligned}$$

- With heterogeneous effects OLS will give a consistent estimate of the average treatment effect.

Heterogeneous treatment effects & 2SLS

- If we estimate a 2SLS model with heterogeneous effects

$$Y_i = \beta_0 + \beta_{1i} X_i + u_i \qquad \textit{with} \qquad \beta_{1i} \neq \beta_1$$

$$X_i = \pi_0 + \pi_{1i} Z_i + v_i \qquad \textit{with} \qquad \pi_{1i} \neq \pi_1$$

  the IV-estimator $\widehat{\beta}_{IV}$ will not be a consistent estimator of the average treatment effect in the whole population.

- Instead the IV-estimator will be a consistent estimator of the

local average treatment effect (LATE): the average treatment effect in the sub-population of those who are affected by the instrument.

Draft-lottery example: the average causal effect of military service on earnings *for men who complied with draft eligibility status* is equal to $\widehat{\beta}_{IV} = -2,741$ (dollars)

Compliers: men that would serve in the military if draft eligible but would not serve if draft ineligible.

Regression discontinuity design

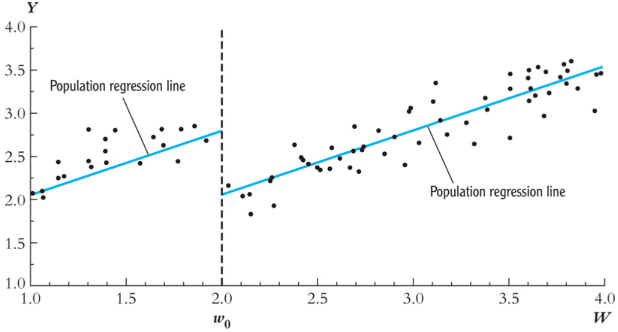
Another example of a quasi-experiment is a (fuzzy) regression discontinuity design:

- If treatment occurs when some continuous variable W crosses a threshold w_0, then you can estimate the treatment effect by comparing
  - individuals with W just below the threshold (treated)
  - individuals to these with W just above the threshold (untreated).
- If the direct effect on Y of W is continuous, the effect of treatment should show up as a jump in the outcome.
- The magnitude of this jump estimates the treatment effect.

Regression discontinuity design

Sharp regression discontinuity design in a picture:

Treatment occurs for everyone with $W < w_0$, and the treatment effect is the jump or "discontinuity."



Regression discontinuity design

Sharp regression discontinuity design: everyone on one side of the threshold w_0 gets treatment, those on the other side do not get the treatment.

- This is an example of a quasi experiment whereby the treatment is "as if" randomly assigned conditional on W
- Treatment effect can be estimated by estimating equation below by OLS

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$$

with $X_i = 1$ if $W < w_0$ & $X_i = 0$ if $W \geq w_0$

- Assuming that the direct effect of W_i on Y_i is linear and

$$E\left[u_i | X_i, W_i\right] = E\left[u_i | W_i\right]$$

Regression discontinuity design

Fuzzy regression discontinuity design: crossing the threshold w_0 influences the probability of treatment, but that probability is between 0 and 1.

- This is an example of a quasi experiment whereby the treatment is partially affected by "as if" randomization conditional on W
- Treatment effect can be estimated by using 2SLS

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i \qquad X_i = \pi_0 + \pi_1 Z_i + \pi_2 W_i + v_i$$

with $Z_i = 1$ if $W < w_0$ & $Z_i = 0$ if $W \geq w_0$

- Assuming that the direct effect of W_i on Y_i is linear and

$$E\left[u_i | Z_i, W_i\right] = E\left[u_i | W_i\right]$$

Quasi-experimental estimates of class size

- Angrist and Lavy (1999) use a fuzzy RD design based on the interpretation of the Talmud by 12th century rabbinic scholar Maimonides.

- According to Maimonides' rule:

  "Twenty five children may be put in charge of one teacher. If the number in the class exceeds twenty-five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed"

- Since 1969 Maimonides' rule is used to determine the division of enrollment cohorts into classes in Israeli public schools

- Angrist and Lavy use this maximum class size rule as a source of exogenous variation to estimate the effect class size on test scores in elementary school.

Quasi-experimental estimates of class size

- Angrist and Lavy link test score data with information on class size, enrollment and other school characteristics
- They estimate the following specification by 2SLS

$$Y_{sc} = \beta_0 + \beta_1 X_{sc} + \beta_2 W_s + u_{sc}$$

$$X_{sc} = \pi_0 + \pi_1 Z_{sc} + \pi_2 W_s + v_{sc}$$

Y_{sc} is average test score in class c in school s

X_{sc} is the size of class c in school s

W_s is school enrollment

Z_{sc} is predicted class size

Quasi-experimental estimates of class size

- Predicted class size is based on Maimonides' rule:

$$Z_{sc} = W_s / \left[int((W_s - 1)/40) + 1 \right]$$

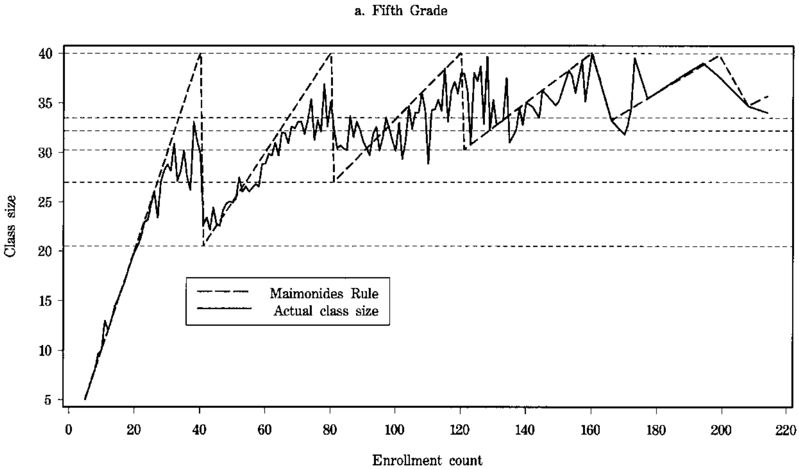
Z_{sc} is predicted class size for school s

W_s is beginning-of-the-year enrollment for school s (for particular grade)

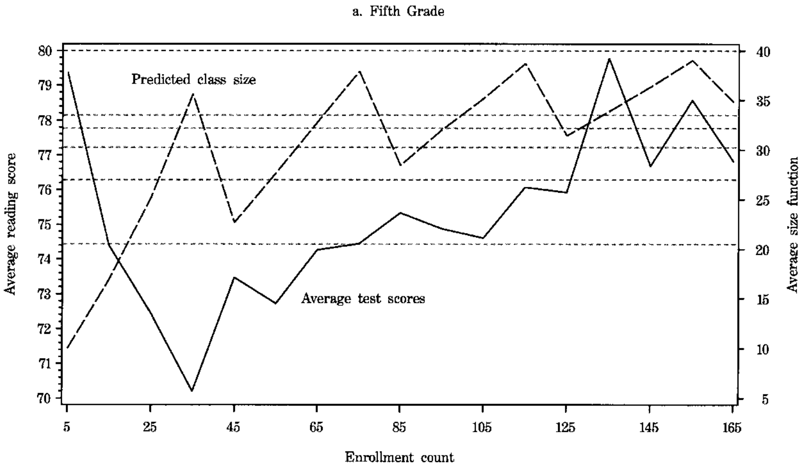
$int(n)$ denotes largest integer less than or equal to n

- Equation captures the fact that according to Maimonides' rule
  - enrollment cohorts of 1–40 should be grouped in a single class,
  - enrollment cohorts of 41–80 should split into 2 classes of average size 20.5–40,
  - enrollment cohorts of 81–120 should be split into 3 classes of average size 27–40, and so on.

Quasi-experimental estimates of class size



Quasi-experimental estimates of class size



Quasi-experimental estimates of class size

Predicted class size Z_{sc} valid instrument if it satisfies:

Instrument relevance: $Cov(X_{sc}, Z_{sc}|W_s) \neq 0$

- Can be checked by estimating first stage regression.

Instrument exogeneity: $Cov(u_{sc}, Z_{sc}|W_s) = 0$

- predicted class size (Z_{sc}) depends on enrollment (W_s)
- enrollment also has direct impact on student achievement (Y_{sc}) for other reasons than class size (X_{sc})
- hence, predicted class size as such is not an exogenous instrument
- however, assuming that effect of enrollment has been adequately controlled for in test scores equation the remaining variation of predicted class size serves as exogenous instrument

Quasi-experimental estimates of class size

	First stage	2SLS
Dependent variable:	Class size	Math test score
Class size (X_{sc})		-0.230**
		(0.092)
Predicted class size (Z_{sc})	0.542***	
	(0.027)	
Enrollment (W_s)	0.043***	0.041***
	(0.005)	(0.012)
First stage F-statistic	402.97	

Note: ** significant at 5% level, *** significant at 1% level

- Increasing class size by 1 pupil decreases average class test scores by 0.23 points.
- Results rely on assumption that direct effect of enrollment is linear.
- Angrist & Lavy (1999) therefore also estimate models with more flexible functions of enrollment.