ECON 326: Economics of Developing Countries
TA Session 4

Vaidehi Parameswaran (Northwestern Econ)

January 2025

Today's Agenda

- Benefits of Education - Duflo (2001)
- Problem Set 2 - Some Pointers

Duflo (2001)

Schooling and Labour Market Consequences of School Construction in Indonesia

Overview

- A very influential paper that uses non-experimental methods
- Evaluates the impact of a large school construction program in Indonesia
- This was not an RCT, but a quasi-experimental design
- The author uses a difference-in-difference strategy to estimate the impact of the program
- Combined with an IV strategy to estimate returns to school
- Findings are consistent with other papers (e.g., Angrist & Krueger 1991)

The Benefits of Education

- Why do people send their children to school?
  - Knowledge as an end in itself
  - Labour market: earnings, access to formal jobs, profitability of own business
  - Social status
  - Political attitudes and participation
- To economists, education is a form of human capital - you invest in it and get returns in the form of higher earnings (and other benefits)
- This is especially interesting in developing countries, where the returns to education are often higher than in developed countries

Why is it hard to estimate the returns to schooling?

- Mincer equation: $\ln(W) = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \varepsilon$
- But we know by now that correlations are not causations
- What are possible biases?
  - Selection bias: People who go to school are different from those who don't
  - Reverse causality: People who are have higher expected potential earnings may choose to go to school
  - Omitted variables: There may be other factors that affect both schooling and potential earnings
- Ideal experiment?
- Can we randomly assign "education" to people?

Assigning an "Instrument"

- But one can randomly assign a student to a program which may increase their education
  - Examples of such interventions?
- Then we can exploit the effect an intervention has on education
  - If the intervention has no direct effect on earnings (or other outcomes), but it does affect earnings in the data, then you can infer that it affected earnings through education alone, and hence that education affects earnings
- Today's paper uses this insight
- Then we can apply this technique in a non-experimental setting

Sekolah Dasar INPRES

- In 1973, Indonesian government prioritised "equity" across provinces
- Launched a large school construction program
- Between 1973-1974 and 1978-1979, 61,807 schools were built
- The largest and the fastest school construction program in the world
- Financed by oil boom revenues
- Explicitly targeted children who were previously not enrolled in school
    - Number of schools in each district was proportional to number of children of primary school age not enrolled in 1972

Data

- Intercensal survey data in 1995
  - Focus on men born between 1950 and 1972
- Match with district-level census data and number of schools built in each district

Empirical Strategy (I)

- Goal: Estimate the effect of INPRES on schooling and labour market outcomes.
- Key idea: Children born before 1962 were 12 years or older in 1974 and were not affected by the program.
- So children in Indonesia belong to one of these cells:

	INPRES Intensity in Region of Birth	
Cohorts	High	Low
(A) Aged 2-6 in 1974:	High Exposure to INPRES	Low Exposure to INPRES
(B) Aged 12-17 in 1974:	Low Exposure to INPRES	Low Exposure to INPRES

Empirical Strategy (II)

	INPRES Intensity in Region of Birth	
Cohorts	High	Low
(A) Aged 2-6 in 1974:	High Exposure to INPRES	Low Exposure to INPRES
(B) Aged 12-17 in 1974:	Low Exposure to INPRES	Low Exposure to INPRES

- Naive approach: Compare cohorts A and B to estimate the effect of INPRES
- But this can underestimate the actual effects of INPRES:
  - Only those in high-intensity regions get the full benefits of the policy
- And it's hard to argue that cohorts A and B *only* differ in exposure to INPRES
  - If Indonesia is experiencing growth over time, we expect cohort A to do better, regardless of whether INPRES is implemented

Empirical Strategy (III)

	INPRES Intensity in Region of Birth	
Cohorts	High	Low
(A) Aged 2-6 in 1974:	High Exposure to INPRES	Low Exposure to INPRES
(B) Aged 12-17 in 1974:	Low Exposure to INPRES	Low Exposure to INPRES

- Another naive approach: Compare high vs low intensity regions
- But this is misleading as well:
  - INPRES targeted regions that were worse off to begin with
  - So we expect people in high-intensity regions to do worse, regardless of INPRES

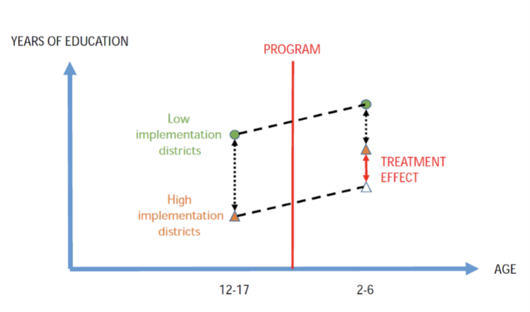
Empirical Strategy (IV)

	INPRES Intensity in Region of Birth	
Cohorts	High	Low
(A) Aged 2-6 in 1974:	High Exposure to INPRES	Low Exposure to INPRES
(B) Aged 12-17 in 1974:	Low Exposure to INPRES	Low Exposure to INPRES

- Solution: Use both the variation from cohorts and the variation from intensity and combine the two naive approaches
- For each region, we compare the outcomes of cohort A relative to cohort B
- Then we compare the changes in outcomes between high and low intensity regions
- What is the key assumption for this to work?
  - Factors that make high and low intensity regions different affect each cohort identically

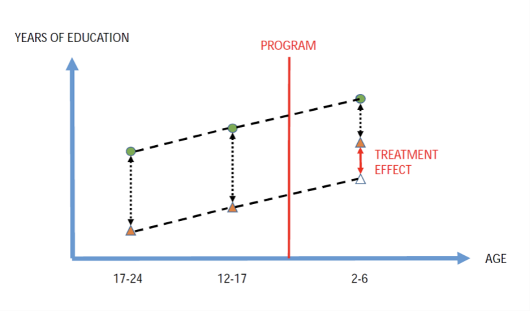
Empirical Strategy (V)

- This was a classic example of a difference-in-difference strategy
- The key identifying assumption: **Parallel Trends Assumption**
  - In the absence of INPRES, the outcomes of cohorts A and B in high and low intensity regions would have evolved in the same way



Empirical Strategy (VI)

- Fundamentally untestable assumption
- But we can verify its plausibility by looking at pre-trends
- Compare a new placebo cohort (C) to cohort B
  - If the outcomes of cohorts B and C evolve the same way, then we have suggestive evidence that there were no different intergenerational effects by region



Empirical Strategy (Regression)

- Let P_j be program intensity in region (number of schools per 1,000 children)
- Let T_i be an indicator for cohort A (the young cohort)
- Let C_j be a vector of region characteristics
- Duflo (2001) estimates the effect of the program on schooling (S_{ijk}) and earnings y_{ijk} of individual i born in region j in year k using the following regression

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \gamma_1 T_i * P_j + \delta_1 T_i * C_j + \epsilon_{ijk}$$
$$y_{ijk} = c_2 + \alpha_{2j} + \beta_{2k} + \gamma_2 T_i * P_j + \delta_2 T_i * C_j + \nu_{ijk}$$

Empirical Strategy (IV)

- First stage:

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \gamma_1 T_i * P_j + \delta_1 T_i * C_j + \epsilon_{ijk} \quad (1)$$

- Reduced form:

$$y_{ijk} = c_2 + \alpha_{2j} + \beta_{2k} + \gamma_2 T_i * P_j + \delta_2 T_i * C_j + \nu_{ijk} \quad (2)$$

- Second stage:

$$y_{ijk} = c + \alpha_j + \beta_k + b\hat{S}_{ijk} + \delta T_i * C_j + \nu_{ijk} \quad (3)$$

- Coefficient of interest:

$$\hat{b} = \frac{\hat{\gamma_2}}{\hat{\gamma_1}} \quad (4)$$

Results (I)

- These are the key results from the diff-in-diff strategy
- INPRES raises schooling by 0.12 years and wages by 0.026 log points ($\sim 2.6\%$)

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Results (II)

- Do we believe that INPRES *only* affects wages through education?
  - If so, these results imply huge returns to education (over 20%!)
  - To see this, take the ratio of both effects above: $\frac{2.6\%}{0.12}$

Table 3—Means of Education and Log(Wage) by Cohort and Level of Program Cells

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Results (III)

- More formally: within regions, exposure to INPRES can be used as an IV
  - The first stage of the IV is the left panel of the table below
  - And the right panel is its reduced form

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Results (IV)

- But we have an **IV relevance** problem
- The effect of INPRES on schooling lacks significance (coef. $\sim$ std. error)
  - To address this, need to move towards a more sophisticated estimation strategy that is beyond the scope of today's session but is conceptually similar

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Results (V)

- Reassuringly, non-exposed cohorts ("old" and "very old") evolve similarly across high/low intensity regions
- So we can claim that there's no evidence for differential pre-trends

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

- We can extend the logic of the previous table in two ways:
  - Compare more than two cohorts
  - Measure INPRES intensity continuously (as schools built per student)
- This leads to more precise estimates of the returns to education, with smaller values that are more in line with the literature
- The figure below illustrates how this works: we compute the correlation between INPRES intensity and schooling for each cohort

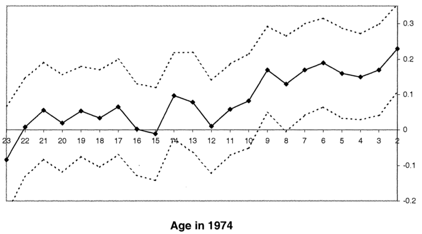


Figure 1. Coefficients of the Interactions Age in 1974* Program Intensity in the Region of Birth in the Education Equation

Results (VII)

- Plot of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ for each cohort

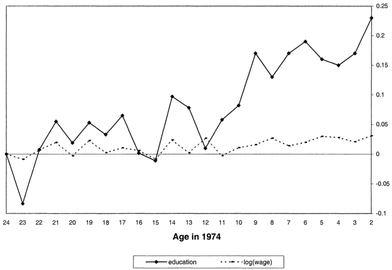


FIGURE 3. COEFFICIENTS OF THE INTERACTIONS AGE IN 1974* PROGRAM INTENSITY IN THE REGION OF BIRTH IN THE WAGE AND EDUCATION EQUATIONS

Results (VIII)

- Here are the main results of the paper
- Returns to education around 7%

Table 7—Effect of Education on Labor Market Outcomes: OLS and 2SLS Estimates

Method	Instrument	(1)
Panel A: Sample of Wage Earners		
Panel A1: Dependent variable: log(hourly wage)		
OLS		0.0776
		(0.000620)
2SLS	Year of birth dummies*program intensity in region of birth	0.0675
		(0.0280)
		[0.96]
2SLS	(Aged 2–6 in 1974)*program intensity in region of birth	0.0752
		(0.0338)
		(0.0338)

Cost-Benefit Analysis

- The paper ends by comparing INPRES's impacts to its costs
- Three big assumptions author needs to make:
  - How big is the dead-weight burden of taxes? **Assumes 0.2-0.6 for each unit of tax**
  - Does schooling increase earnings by increasing productivity or by signalling/rent capture? **Assumes the former**
  - Externalities of education? Assumes none
- Benefit of INPRES, in each year, is $\alpha \times GDP \times S \times \ E$
- α is the labour share of GDP, S is the share of wages earned by cohort that benefits from INPRES and E is the estimated average effect of the program
- Estimates costs of building and running INPRES schools
- Finds that INPRES had an internal rate of return between 9% and 12%
  - Higher than the interest rate of Indonesia's sovereign debt at the time
  - So it's profitable to fund more school building with government debt

Conclusion

- INPRES leads to more schooling and more earnings
- If we use exposure to INPRES (within region) as an IV, we obtain returns to education around 5-10%
  - In line with a large literature, including the seminal Angrist & Krueger (1991) paper
- But there are a few problems with this
  - INPRES did more than just school building
  - Poor regions may catch up with rich regions independently of INPRES
  - Negative spillovers and general equilibrium effects? Impacts on broader labour markets due to stock of educated workers?

Problem Set 2

Comments

- When discussing spillover effects from an RCT intervention into the control group:
  - It's not that the intervention becomes less effective - it could be more effective depending on the type of spillover
  - Estimates are biased towards zero if we are (incorrectly) assuming that the control group is not exposed to treatment
- Creating categorical variables in Stata, generating numeric variables
- Important note: for the control group, even though the beneficiary variable is missing, make sure you assign it a value of zero for the indicator mother or father variables.

See you next time!