

Economic Policy Analysis (EC406)

Difference in Differences

John Van Reenen
Centre for Economic Performance
London School of Economics

* These lecture notes are incomplete. Reading the lecture notes without attending lectures can be very misleading.

The Program Evaluation Problem

Homogenous Treatment Effects and Linear Model

Assume homogenous treatment effects

- The treatment has the same effect across individuals

Suppose that the program takes place at date k

- $d_i = 1$ if an individual participates in the program
- $d_i = 0$ otherwise

Consider the following linear regression model

$$y_{it} = \beta_0 + \alpha d_i + u_{it} \quad \text{if } t > k$$

$$y_{it} = \beta_0 + u_{it} \quad \text{if } t \leq k$$

Where, for simplicity, there are no control variables

The Program Evaluation Problem

Homogeneous Treatment Effects and Linear Model

Important econometric concern is that $E(u|d) \neq 0$

$$E(y | d = 1) - E(y | d = 0) = \alpha + E(u | d = 1) - E(u | d = 0)$$

⇒ OLS yields biased and inconsistent estimates of the treatment effect

Consider following error components model

$$u_{it} = \eta_i + \theta_t + \mu_{it}$$

- η_i is fixed effect (unobserved heterogeneity)
- θ_t are time dummies
- μ_{it} is idiosyncratic error

Make the identifying assumption

$$E(\mu | d) = 0$$

Difference in Differences Estimator

t_0 denotes pre-treatment periods for which data are available

t_1 denotes post-treatment periods for which data are available

Average change in outcome (pre and post-treatment) for treatment group
minus average change in outcome for control group

$$\hat{\alpha}^{DID} = \left(\bar{y}_{t_1}^{Treat} - \bar{y}_{t_0}^{Treat} \right) - \left(\bar{y}_{t_1}^{Control} - \bar{y}_{t_0}^{Control} \right)$$

Identification comes from the *differential* change between the two groups
pre and post-treatment

- difference out unobserved fixed effects
- difference out common time effects

Implicitly make assumptions about individuals' participation decision,
 $E(\mu / d) = 0$

Difference in Differences (OLS Representation and Control Variables)

Difference in Differences estimator can also be represented as the following OLS regression

$$y_{it} = \alpha(d_i \times post_t) + \eta_i + \theta_t + \mu_{it}$$

- $d_i = 1$ if individual i participates in the program and 0 otherwise
- $post_t = 1$ post-treatment and 0 otherwise

This representation allows us to control for time-varying considerations which affect the outcome and vary between the two groups

$$y_{it} = \alpha(d_i \times post_t) + X_{it}\beta + \eta_i + \theta_t + \mu_{it}$$

- Assumes the same coefficients on the controls between the groups and over time
- This assumption can be relaxed

Difference in Differences (Allowing for Heterogeneous Time Effects)

Suppose that there are heterogeneous time effects between the treatment and control groups

$$u_{it} = \eta_i + \psi_g \theta_t + \mu_{it}$$

The difference in differences estimator identifies

$$E(\hat{\alpha}^{DID}) = \alpha + (\psi_T - \psi_C)(\theta_{t_1} - \theta_{t_0})$$

To control for heterogeneous time effects, find another time interval during which a similar macro trend occurred

Trend-adjusted difference in differences

$$\hat{\alpha}^{TADID} = \left[\left(\bar{y}_{t_1}^{Treat} - \bar{y}_{t_0}^{Treat} \right) - \left(\bar{y}_{t_1}^{Control} - \bar{y}_{t_0}^{Control} \right) \right] \\ - \left[\left(\bar{y}_{t_{**}}^{Treat} - \bar{y}_{t_*}^{Treat} \right) - \left(\bar{y}_{t_{**}}^{Control} - \bar{y}_{t_*}^{Control} \right) \right]$$

Empirical Application : School Construction in Indonesia (*Esther Duflo, AER, 2001*)

In 1973, Indonesian government launched a major school construction program, Sekolah Dasar INPRES program

- Between 1973-4 and 1978-9, more than 61,000 primary schools were constructed
 - average of 2 schools per 1,000 children aged 5-14 in 1971
- Enrollment rates among children aged 7-12 increased from 69% in 1973 to 83% by 1978
- In contrast to an absence of capital expenditure and a decline in enrollment in the early 1970s

Exposure of an individual to the program was determined by the number of schools built in her region of birth and her age when the program was launched

Research Questions and Methodology

Question: What was the effect of the Indonesian INPRES school construction program on?

- Education
- Earnings

Methodology: Difference in Differences estimation

- Region of birth
- Cohort of birth

(exposure of an individual to the program determined by the number of schools built in her region of birth and her age when the program was launched)

Research Questions and Methodology

Question: What is the rate of return to education?

$$\ln W_i = X_i \beta + \rho E_i + \mu_i$$

Econometric problem: the error term, which includes unobserved ability, may be correlated with individuals' education decision

Exploit exogenous variation in education due the INPRES school construction program

Instrumental variables estimation

- Interactions between region of birth & cohort of birth as instruments

Identifying assumption

- These interactions only affect wages through years of education

Data

Intercensal survey of Indonesia 1995

- Focus on men born between 1950 and 1972
- Between ages 24 and 2 in 1974 when first schools constructed

Sample

- Information on 152,989 individuals
- Average of 7.98 years of completed education (6 years of education corresponds to graduation from primary school)
- Subsample of 60,633 individuals work for a wage (potential sample selection issues)

Match individual-level survey data with district-level census data and with the number of schools scheduled to be constructed in each district under the INPRES program

- Actual school construction followed scheduled construction closely

Program Background

1973 launch of the Sekolah Dasar INPRES school construction program

- Between 1973-4 and 1978-9, more than 61,000 primary schools were constructed
- Cost of over 500 million 1990 US dollars (1.5% of GDP in 1973)
- Once an INPRES school established, government recruited the teachers and paid their salaries
 - Stock of schools doubled
 - Stock of teachers grew by 43%

Program designed to explicitly target children who had not previously been enrolled in school

- Presidential instructions listed exact number of schools to be constructed in each district
- Number to be constructed in a district proportional to the number of primary school aged children *not enrolled in school* in 1972

Table 2 : Allocation of Schools

TABLE 2—THE ALLOCATION OF SCHOOLS

	Log(INPRES schools) ^a
Log of number of children aged 5–14 in the region	0.78 (0.027)
Log(1 – enrollment rate in primary school in 1973) ^b	0.12 (0.038)
Number of observations	255
R^2	0.78

Notes: Standard errors are in parentheses.

^a The dependent variable is the log of the number of INPRES schools built between 1973 and 1978.

^b The enrollment rate in primary school is the number of children enrolled in primary school in 1973 (obtained from the Ministry of Education and Culture) divided by the number of children aged 5–14 in the region in 1973.

Identification Strategy

Indonesian children normally attend primary school during age 7-12

- Abstracting from grade repetition, all children born in 1962 or earlier were 12 or older in 1974 when the first schools were constructed, and hence were not affected by the program
- For younger children, exposure to the program is an increasing function of their date of birth

Because school construction was related to enrollment rates in 1972 and these varied widely across regions, program intensity also varies with region of birth

- Region of birth highly correlated with region of education (91.5% of children still living in district of birth at age 12)
- Since all individuals in the sample born before the program started, region of birth exogenous with respect to the program

Difference in Differences

Distinguish between “high program” and “low program” regions

- Based on positive or negative residuals in a regression of number of schools constructed on number of children aged 5-14 in 1973

Distinguish between individuals with

- Little or no exposure to the program (aged 12 to 17 in 1974)
- Exposed the entire time in primary school (aged 2 to 6 in 1974)

Consider the difference in differences between these two groups

- Causal effect of the program

Identifying assumption

- In the absence of the program, the increase in educational attainment would not have been systematically different in high and low program regions (motivation for falsification exercise)

Table 3 : Means of Education and Log(Wage) by Cell

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)
<i>Panel A: Experiment of Interest</i>						
Aged 2 to 6 in 1974	8.49 (0.043)	9.76 (0.037)	-1.27 (0.057)	6.61 (0.0078)	6.73 (0.0064)	-0.12 (0.010)
Aged 12 to 17 in 1974	8.02 (0.053)	9.40 (0.042)	-1.39 (0.067)	6.87 (0.0085)	7.02 (0.0069)	-0.15 (0.011)
Difference	0.47 (0.070)	0.36 (0.038)	0.12 (0.089)	-0.26 (0.011)	-0.29 (0.0096)	0.026 (0.015)
<i>Panel B: Control Experiment</i>						
Aged 12 to 17 in 1974	8.02 (0.053)	9.40 (0.042)	-1.39 (0.067)	6.87 (0.0085)	7.02 (0.0069)	-0.15 (0.011)
Aged 18 to 24 in 1974	7.70 (0.059)	9.12 (0.044)	-1.42 (0.072)	6.92 (0.0097)	7.08 (0.0076)	-0.16 (0.012)
Difference	0.32 (0.080)	0.28 (0.061)	0.034 (0.098)	0.056 (0.013)	0.063 (0.010)	0.0070 (0.016)

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

Basic Education Specification

A more efficient estimation strategy exploits the full variation in treatment intensity across regions and cohorts

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

Distinguish between “high program” and “low program” regions

- Individual i , born in region j in year k
- T_i is dummy indicating whether or not the individual belongs to the young cohort (aged 2-6 in 1974 versus 12-17 in 1974)
- β_{1k} is cohort of birth fixed effect
- α_{1j} is district of birth fixed effect
- P_j denotes the intensity of the program in the region of birth (schools constructed per 1,000 children)
- C_j is a vector of region-specific variables

γ_1 is a differences in differences estimate of the effect of the program

Table 4 : Effect of the Program on Education and Wages

TABLE 4—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS ON DUMMIES AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN

One school built per 1,000 children increased the education of children aged 2-6 in 1974 by 0.12 years for the whole sample and 0.20 years for the sample of wage earners

	Observations	Years of education					
		(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Experiment of Interest: Individuals Aged 2 to 6 or 12 to 17 in 1974</i>							
<i>(Youngest cohort: Individuals ages 2 to 6 in 1974)</i>							
Whole sample	78,470	0.124 (0.0250)	0.15 (0.0260)	0.188 (0.0289)			
Sample of wage earners	31,061	0.196 (0.0424)	0.199 (0.0429)	0.259 (0.0499)	0.0147 (0.00729)	0.0172 (0.00737)	0.0270 (0.00850)
<i>Panel B: Control Experiment: Individuals Aged 12 to 24 in 1974</i>							
<i>(Youngest cohort: Individuals ages 12 to 17 in 1974)</i>							
Whole sample	78,488	0.0093 (0.0260)	0.0176 (0.0271)	0.0075 (0.0297)			
Sample of wage earners	30,225	0.012 (0.0474)	0.024 (0.0481)	0.079 (0.0555)	0.0031 (0.00798)	0.00399 (0.00809)	0.0144 (0.00915)
<i>Control variables:</i>							
Year of birth*enrollment rate in 1971		No	Yes	Yes	No	Yes	Yes
Year of birth*water and sanitation program		No	No	Yes	No	No	Yes

Notes: All specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummies and the number of children in the region of birth (in 1971). The number of observations listed applies to the specification in columns (1) and (4). Standard errors are in parentheses.



Identification

Identifying assumption

- There are not omitted cohort-varying and region-specific effects correlated with the program
 - Mean reversion that would have taken place in the absence of the program?
 - Allocation of other governmental programs across regions correlated with the allocation of INPRES schools?

Inclusion of control variables

Falsification exercise

- Compare cohort aged 12-17 with cohort aged 18-24 in 1974



Generalized Reduced Form Specifications

Consider a specification with generalized year of birth interactions

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \sum_{l=2}^{23} (P_j \times d_{il}) \gamma_{1l} + \sum_{l=2}^{23} (C_j \times d_{il}) \delta_{1l} + \varepsilon_{ijk}$$

- d_{il} is a dummy that indicates whether individual i is of age l in 1974
- The dummy for individuals of age 24 in 1974 is omitted, so that they form the base group

Figure 1 graphs the estimated γ_{1l}

- As expected, the estimated coefficients fluctuate around 0 until age 12 and start increasing after age 12
- Estimated coefficients significantly different from 0 after age 8

Also consider restricted specification, where γ_{1l} are equal to 0 for $l \geq 13$

Figure 1 : Coefficients on Interactions

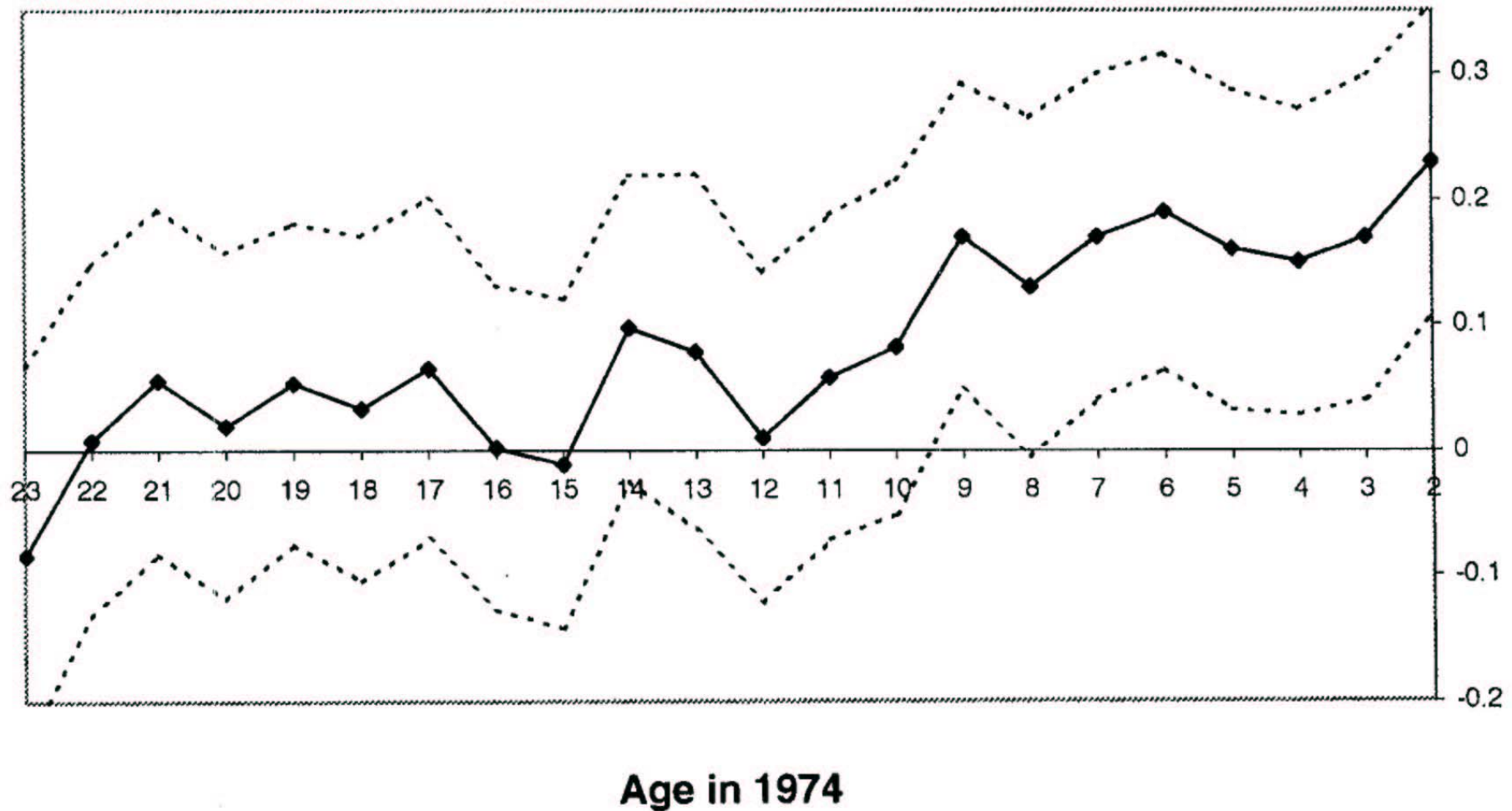


FIGURE 1. COEFFICIENTS OF THE INTERACTIONS AGE IN 1974* PROGRAM INTENSITY IN THE REGION OF BIRTH IN THE EDUCATION EQUATION

Table 5 : Effect on Education and Wages (Restricted Estimation)

TABLE 5—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN DUMMIES INDICATING AGE IN 1974 AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN REGION OF BIRTH

Age in 1974	Dependent variable: years of education					Dependent variable: log(hourly wage)			
	Whole sample			Sample of wage earners					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
12	-0.035 (0.047)	-0.025 (0.048)	0.002 (0.054)	-0.040 (0.077)	-0.010 (0.078)	0.009 (0.091)	0.016 (0.013)	0.019 (0.013)	0.027 (0.015)
11	0.011 (0.046)	0.025 (0.047)	0.018 (0.051)	0.008 (0.073)	0.014 (0.074)	-0.003 (0.083)	-0.014 (0.012)	-0.013 (0.013)	-0.009 (0.014)
10	0.059 (0.047)	0.049 (0.049)	0.078 (0.054)	0.10 (0.075)	0.092 (0.076)	0.13 (0.090)	0.0036 (0.013)	0.0042 (0.013)	0.0059 (0.015)
9	0.14 (0.039)	0.14 (0.041)	0.15 (0.044)	0.067 (0.065)	0.063 (0.066)	0.17 (0.077)	0.0095 (0.011)	0.010 (0.011)	0.018 (0.013)
8	0.088 (0.049)	0.11 (0.050)	0.11 (0.054)	0.19 (0.078)	0.20 (0.079)	0.28 (0.089)	0.019 (0.013)	0.021 (0.013)	0.027 (0.015)
7	0.12 (0.044)	0.14 (0.046)	0.16 (0.051)	0.11 (0.072)	0.13 (0.073)	0.16 (0.084)	-0.0095 (0.012)	-0.0049 (0.012)	0.0066 (0.014)
6	0.14 (0.042)	0.17 (0.044)	0.26 (0.049)	0.23 (0.070)	0.23 (0.070)	0.32 (0.084)	0.011 (0.012)	0.013 (0.012)	0.018 (0.014)
5	0.10 (0.043)	0.13 (0.045)	0.13 (0.050)	0.14 (0.075)	0.16 (0.075)	0.27 (0.088)	0.021 (0.013)	0.023 (0.013)	0.052 (0.015)
4	0.11 (0.039)	0.12 (0.041)	0.18 (0.046)	0.19 (0.069)	0.19 (0.069)	0.29 (0.082)	0.019 (0.012)	0.020 (0.012)	0.038 (0.014)
3	0.11 (0.044)	0.14 (0.046)	0.20 (0.053)	0.15 (0.079)	0.17 (0.080)	0.30 (0.097)	0.0079 (0.013)	0.013 (0.014)	0.027 (0.016)
2	0.14 (0.041)	0.19 (0.043)	0.19 (0.049)	0.20 (0.073)	0.22 (0.074)	0.25 (0.088)	0.016 (0.012)	0.023 (0.013)	0.040 (0.015)
<i>Control variables:^a</i>									
Year of birth*enrollment rate in 1971	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Year of birth*water and sanitation program	No	No	Yes	No	No	Yes	No	No	Yes
<i>F</i> -statistic ^b	4.03	5.18	6.15	2.70	2.74	4.38	1.13	1.29	2.05
<i>R</i> ²	0.19	0.19	0.17	0.14	0.14	0.13	0.14	0.15	0.13
Number of observations	152,989	152,495	143,107	60,633	60,466	55,144	60,633	60,466	55,144

Notes: All specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummies and the number of children in the region of birth (in 1971). Standard errors are in parentheses.

^a The control group is comprised of individuals aged 13–24 in 1974.

^b The *F*-statistics test the hypothesis that the coefficients of the interaction between the year of birth dummies and the program intensity in the region of birth are jointly zero.

What was the Mechanism Behind the Effect of the Program?

Examine the impact of the program in different types of regions

Find the program had no effect in densely populated regions and a large effect in sparsely populated regions

- In sparsely populated regions, each new school significantly reduces the distance to school
- In densely populated regions, the main effect will be to increase the availability of school places or reduce school overcrowding
- Consistent with reductions in the distance traveled to school being an important effect of the program

Also find that the program had a larger effect in poor provinces

Table 6 : Program Effect and Returns to Education by Region

TABLE 6—PROGRAM EFFECT AND RETURNS TO EDUCATION BY CATEGORIES OF REGION OF BIRTH

	Characteristics of region of birth						
	Whole sample (1)	Density ^a		1976 Poverty ^b		Preprogram education ^c	
		<Median (2)	>Median (3)	High (4)	Low (5)	<Median (6)	>Median (7)
<i>Panel A: Effect of the Program on Education</i>							
Dependent variable: Years of education.							
Sample: individuals ages 2 to 6 or 12 to 17 in 1974							
Interaction (2–6 in 1974)*program intensity in region of	0.15 (0.026)	0.19 (0.035)	–0.014 (0.048)	0.13 (0.058)	0.083 (0.035)	0.14 (0.040)	0.13 (0.036)
<i>Panel B: Effect of the Program on Wages</i>							
Dependent variable: log(hourly wage). Sample: individuals ages 2 to 6 or 12 to 17 in 1974 (wage earners)							
Interaction (2–6 in 1974)*program intensity in region of	0.017 (0.0074)	0.032 (0.011)	–0.00084 (0.012)	0.051 (0.017)	–0.00083 (0.0094)	0.028 (0.013)	0.0046 (0.0095)
<i>Panel C: Returns to Education</i>							
Dependent variable: log(hourly wage). Sample: wage earners							
Years of education	0.078 (0.00062) [0.9]	0.11 (0.026) [0.86]	No First stage [0.88]	0.10 (0.028) [0.88]	No First stage [0.72]	0.12 (0.032) [0.72]	0.029 (0.052) [0.83]

Notes: Region of birth dummies, year of birth dummies, and interactions of year of birth dummies and the number of children and the enrollment in the region in 1971 are included in the regressions. Standard errors are in parentheses, *F*-statistics of the overidentification test are in square brackets.

^a The median density (the density for the region of birth for the median person in the weighted sample) is 308 habitants per square kilometer.

^b The high poverty provinces are the provinces where the proportion of people consuming less than 1,500 Rp per capita is larger than the national for rural regions (in the 1976 SUSENAS). I define “high poverty” as rural districts in these provinces, which are: Lampung, Central East Java, East Nusa Tenggara, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Maluku, Irian Jaya (World Bank, 1979).

^c The preprogram education is the average education in the region of birth for people born in 1962 or before. The median is 3.18 years.

Effect on Wages

Consider an analogous basic wage specification

$$\ln W_{ijk} = c_2 + \alpha_{2j} + \beta_{2k} + (P_j T_i) \gamma_2 + (C_j T_i) \delta_2 + \varepsilon_{ijk}$$

Find one school constructed per 1,000 children increased wages by between 1.5 and 2.7 per cent (Table 4)

- Estimates increase in magnitude when add controls
- Again no evidence of effects in the falsification exercise

Consider a specification with generalized year of birth interactions

$$\ln W_{ijk} = c_2 + \alpha_{2j} + \beta_{2k} + \sum_{l=2}^{23} (P_j \times d_{il}) \gamma_{2l} + \sum_{l=2}^{23} (C_j \times d_{il}) \delta_{2l} + \varepsilon_{ijk}$$

- γ_{2l} should be zero for $l > 12$ and start increasing thereafter
- If the program affects wages only through education, the estimated γ_{2l} should track the estimated γ_{1l}

Figure 3 : Coefficients on the Interactions

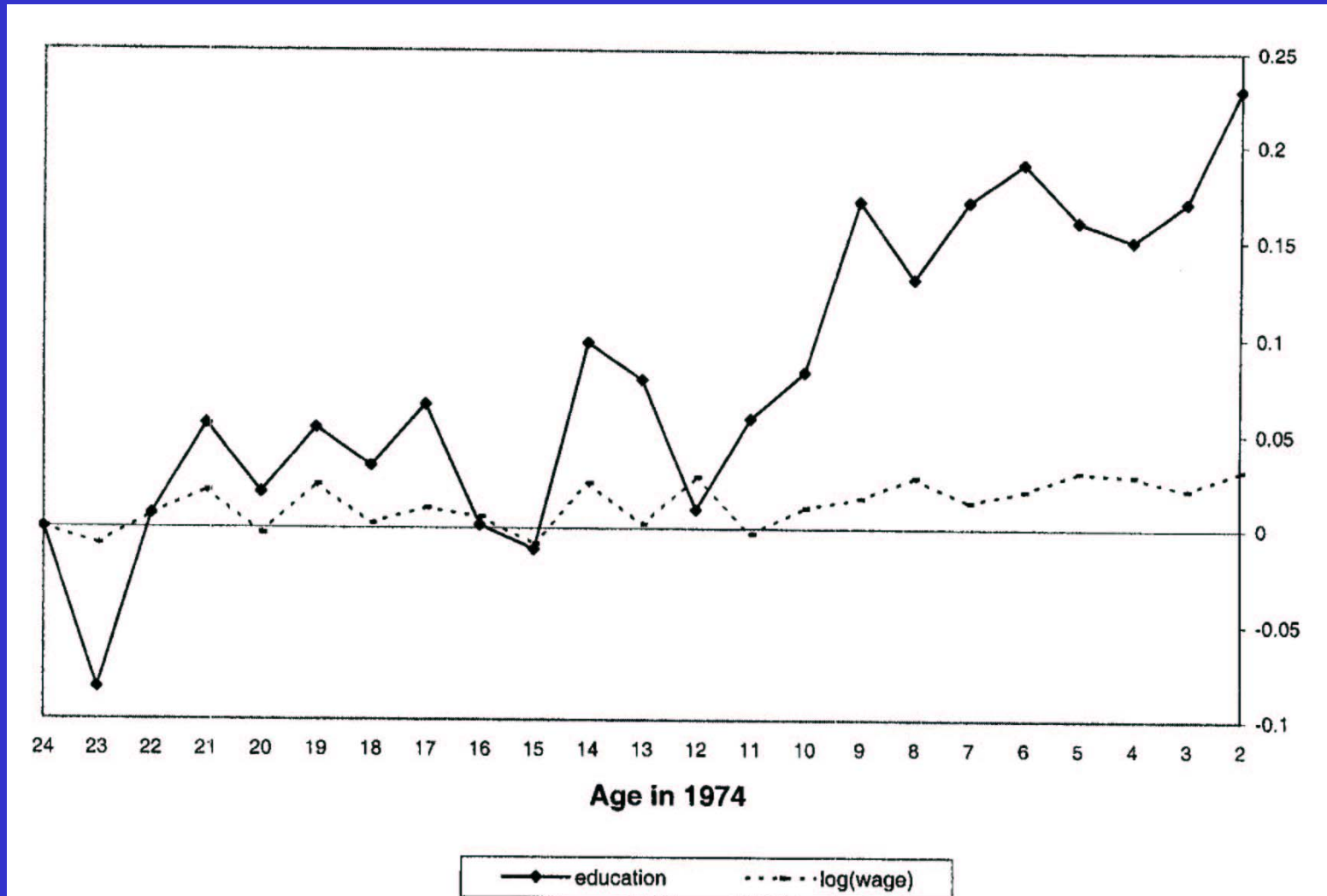


FIGURE 3. COEFFICIENTS OF THE INTERACTIONS AGE IN 1974* PROGRAM INTENSITY IN THE REGION OF BIRTH IN THE WAGE AND EDUCATION EQUATIONS

Estimating the Rate of Return to Education

$$\ln W_{ijk} = d + \alpha_j + \beta_k + S_{ijk} \rho + \xi_{ijk}$$

OLS estimates will yield biased estimates of the rate of return to education (ρ) if $\text{Cov}(\xi, S) \neq 0$

- For example, due to unobserved ability

Under the following identifying assumptions

- Differences in wages across cohorts would not have varied systematically from one region to another in the absence of the program
- The program had no direct effect on wages

Interactions between age in 1974 and program intensity in the region of birth are valid instruments for education in the above wage equation

OLS and IV Estimates

OLS estimates

- Estimated return to education is 7.8 % and not affected by the introduction of control variables

IV estimates

- Without control variables, estimated rate of return is 6.75 % and not statistically different from the OLS estimate
- Estimated rate of return increases when condition on control variables

Sargan test of the model's overidentifying restrictions

- Unable to reject the null hypothesis that the excluded exogenous variables are uncorrelated with the wage equation residuals

Instruments jointly statistically significant in the first-stage regression

Table 7 : Effect of Education on Labor Market Outcomes

TABLE 7—EFFECT OF EDUCATION ON LABOR MARKET OUTCOMES: OLS AND 2SLS ESTIMATES

Method	Instrument	(1)	(2)	(3)	(4)
<i>Panel A: Sample of Wage Earners</i>					
<i>Panel A1: Dependent variable: log(hourly wage)</i>					
OLS		0.0776 (0.000620)	0.0777 (0.000621)	0.0767 (0.000646)	
2SLS	Year of birth dummies*program intensity in region of birth	0.0675 (0.0280) [0.96]	0.0809 (0.0272) [0.9]	0.106 (0.0222) [0.93]	0.0908 (0.0541) [0.9]
2SLS	(Aged 2–6 in 1974)*program intensity in region of birth	0.0752 (0.0338) (0.0338)	0.0862 (0.0336) (0.0336)	0.104 (0.0304) (0.0304)	
<i>Panel A2: Dependent variable: log(monthly earnings)</i>					
OLS		0.0698 (0.000601)	0.0698 (0.000602)	0.0689 (0.000628)	
2SLS	Year of birth dummies*program intensity in region of birth	0.0756 (0.0280) [0.73]	0.0925 (0.0278) [0.63]	0.0913 (0.0219) [0.58]	0.134 (0.0631) [0.7]
<i>Panel B: Whole Sample</i>					
<i>Panel B1: Dependent variable: participation in the wage sector</i>					
OLS		0.0328 (0.00311)	0.0327 (0.000311)	0.0337 (0.000319)	
2SLS	Year of birth dummies*program intensity in region of birth	0.101 (0.0210) [0.66]	0.118 (0.0197) [0.93]	0.0892 (0.0162) [1.12]	
<i>Panel B2: Dependent variable: log(monthly earnings), imputed for self-employed individuals</i>					
OLS		0.0539 (0.000354)	0.0539 (0.000354)	0.0539 (0.000355)	
2SLS	Year of birth dummies*program intensity in region of birth	0.0509 (0.0157) [0.68]	0.0745 (0.0136) [0.58]	0.0346 (0.0138) [1.16]	
Control variables:					
	Year of birth*enrollment rate in 1971	No	Yes	Yes	Yes
	Year of birth*water and sanitation program	No	No	Yes	No
	Propensity score, propensity score squared	No	No	No	Yes
<i>Notes:</i> Year of birth dummies, region of birth dummies, and the interactions between year of birth dummies and the number of children in the region of birth in 1971 are included in the regressions. Standard errors are in parentheses. <i>F</i> -statistics of the test of overidentification restrictions are in square brackets.					

Robustness

Rate of return to education (Quality versus Quantity)

- Estimates of rate of return biased if the program affects both the quantity and quality of education. However:
 - Changes in average pupil/teacher ratio between 1973 and 1977 were not systematically related to the number of INPRES schools constructed
 - The program had no effect on educational attainment (a) of individuals completing 9 years of education or more, (b) in densely populated regions (expect it to if quality changed)

Sample selection bias

- Sample of individuals working for a wage is a non-random sample
 - Heckman selection model
 - Impute an income for the self-employed and re-estimate the model for the extended sample

Cost – Benefit Analysis

Undertake a wider cost-benefit analysis which compares

- Costs of building and operating the new schools
- Additional wealth that the schools generated in the form of increased wages
 - Assuming the increase in wages represents an increase in human capital

Relates back to earlier lectures

The cost-benefit analysis requires additional assumptions

It underestimate the total benefit of the program

- Increase in education may have beneficial effects on other outcomes besides wages (e.g. fertility, child mortality etc)

Costs of School Construction

Yearly costs of school construction

$$C(t) = rK + rTC + (W(t) \times 1.25)$$

Yearly costs of school construction

- K is total capital cost (paid at construction)
- TC is total training cost of new teachers (paid at construction)
- $W(t)$ is total salaries of teachers in year t
- 1.25 incorporates overheads on teachers salaries
- r is real interest rate

Consider two alternative assumptions about deadweight burden of taxation (0.2 and 0.6)

$$\tilde{C}(t) = 1.20 \times C(t)$$

Benefits of School Construction

Yearly benefits of school construction. Assume

- Increase in wages attributed to the program represents an increase in the productivity of labor
- Effect on women working for a wage and self-employed people is the same as the effect on men working for a wage
- Share of total labor income received by people of a given age is constant over time and equal to the share received in 1995

$$B(c,t) = \alpha GDP(t)\phi(c,t)E(c)$$

- B is benefit for cohort c , year t
- GDP is Gross Domestic Product
- $\phi(c,t)$ is share of labor income received by cohort c in year t
- $E(c)$ estimated average effect of the program on cohort c

Total yearly benefits are the sum across cohorts of $B(c,t)$

Discounted Sum of Net Benefits

Take basic education estimates from Columns (1) and (2) of Table 4

- Evaluate discounted sum of net benefits over interval 1973-2050
- Consider alternative assumptions about GDP growth

Main findings

- School construction program takes a long time to generate positive returns
- Returns generated are large
 - Internal rates of return vary from 8.8% - 12%, well above average interest rate on government debt
- To a large extent, the benefits are driven by the rapid growth of Indonesian GDP from 1973-97
 - Did the school construction program play a role here?
 - Generalizability to other contexts?

Table 8 : Evaluation of the Program's Net Return

TABLE 8—EVALUATION OF THE PROGRAM'S NET RETURN

	Deadweight loss coefficient			
	0.2		0.6	
	(1)	(2)	(3)	(4)
<i>Panel A: Results</i>				
Control for year of birth*enrollment rate	No	Yes	No	Yes
First year where benefit > costs (discount rate = 5 percent)				
In annual value	1996	1996	1997	1997
In discounted sum	2005	2002	2009	2005
Discounted sum of net benefits in 2050 (growth rate after 1997 = 5 percent, discount rate 5 percent)				
In million 1990 U.S.\$	13,025	13,096	11,340	18,807
As a fraction of Indonesia's GDP in 1973	0.30	0.36	0.31	0.52
Divided by initial costs	24.1	24.2	21.0	35.0
Discounted sum of net benefits in 2050 (growth rate after 1997 = 2 percent, discount rate 5 percent)				
In million 1990 U.S.\$	6,691	11,589	5,008	9,905
As a fraction of Indonesia's GDP in 1973	0.18	0.32	0.14	0.27
Divided by initial costs	12.4	21.4	9.26	18.3
Discounted sum of net benefits in 2050 (growth rate from 1973 = 2 percent, discount rate 5 percent)				
In million 1990 U.S.\$	-631.6	1,200	-2,315	-483
As a fraction of Indonesia's GDP in 1973	-0.017	0.033	-0.063	-0.013
Divided by initial costs	-1.16	2.22	-4.28	-0.89
Internal rate of return ^a				
Growth rate after 1997 = 5 percent	0.102	0.118	0.0895	0.105
Growth rate after 1997 = 2 percent	0.088	0.106	0.0750	0.0915
Growth rate from 1973 = 2 percent	0.0443	0.059	0.0326	0.0467
<i>Panel B: Assumptions and Parameters</i>				
Population growth rate after 1997	0.015			
Yearly teacher's salary in 1973 (1990 U.S. dollars)	363			
Yearly teacher's salary in 1995 (1990 U.S. dollars)	2,467			
Total recurrent costs/teacher salary	1.25			
Total cost of construction (million 1990 U.S. dollars)	522			
Number of schools constructed	61,800			
Lifetime of the schools (years)	20			
Share of labor income in GDP	0.7			

Notes: The estimates underlying these calculations are taken from Table 5 [columns (7) and (8)]. Program effect has been set to 0 for children aged 7 or older in 1974.

^a The internal rate of return is the interest rate such that the net present value of the project at infinity is 0.

Conclusions

Considered difference in differences as a method of program evaluation

Identifying assumptions

- Idiosyncratic error term in the outcome equation is uncorrelated with the treatment (and hence the participation decision)
- Difference out time-invariant unobserved heterogeneity
- Difference out common time effects
- Can generalize the approach to allow heterogeneous time effects

Examined the Indonesian Sekolah Dasar INPRES school construction program as a natural experiment

- Difference in differences estimates of the effect on education and wages, exploiting interactions between cohort & region of birth
- Instrumental variables estimates of the rate of return to education using these interactions as instruments for education