Slide 1

Education–What works?

Esther Duflo

14.771, Fall 04

Slide 2

Some facts and questions about education

There are enormous disparities in educational outcomes:

- ➜ Around the world
- ➜ Across regions in the same country
- ➜ By gender within countries
- ➜ By income levels
- ➜ By urban/rural residence

Slide 3

Descriptive Evidence from India

Probe report: survey of schools and households in 234 villages in Bihar, Madhya Pradesh, rajasthan, UP, Himachal Pradesh: Northern Indian states.

Some striking facts from this report and other accumulated evidence

- ➜ Very poor education performance
- ➜ Fairly high motivation by parents
- ➜ Child labor does not seem to be a big constraint
- ➜ Direct cost are small but not negligible
- ➜ School availability is not a constraint at the primary school level, may matter more at the post primary school
- ➜ School quality is dismal (teacher absence;facilities

Slide 4

Possible interventions to improve school participation and performance

- ➜ Affecting the direct costs scholarship programs, vouchers, school construction.
- ➜ Affecting the opportunity costs: child labor ban, mandatory schooling, conditional income transfers, school meals, incentives to learn
- ➜ Affecting the returns by changing school quality textbooks, teacher training, teacher incentives, class size, remedial education, computer assisted learnings
- ➜ Improving income levels: unconditional transfers.
- ➜ No intervention specifically in education: foster economic growth to improve the returns to education; improve health.

School Construction: Indonesia: set up

Slide 5

The INPRES school construction program

Second five year plan (1974-79)-Oil shock.

- A large program:
  - 61,807 primary schools constructed from to 1973/74 to 1978/79.

    Number of schools multiplied by 2. 1 schools for every 500 children.
  - A *change* in policy: Before 1973, no construction, ban on recruiting for public service positions.
- A program meant to favor low-enrollment regions.

  Allocation rule: number of schools constructed in a district proportional to the number of children (ages 7 to 12) not enrolled in primary school.

Data and sources of variation

Slide 6

SUPAS 95: A survey done in 1995: after the children educated in these schools have completed their schooling, and have started working.

- 150,000 men born 1950-1972

- Variables: education, year and region of birth, wages.

Sources of variation

Slide 7

Two factors affect the intensity of the program.

- Year of birth : Examples
  - Born in 1962 or earlier: 12 or older in 1974. Not exposed to the program.
  - Born in 1967: 7 in 1974, 12 in 1979. Some exposure to the program.
  - What would we find if we compare the education of those born before and after 1962? Would this be a good measure of the impact of the program? Why?
- Region of birth

  The government was targeting low enrollment regions ⇒ substantial variation in program intensity across districts.

  What would we find if we compare regions with high and low construction?

The "Difference in differences" methodology

Slide 8

- **Basic idea**

Suppose that there are two regions in the data: a "high program" region, and a "low program" region.

Suppose that we have the age group of the individuals: "young people", born after 1967 and who could fully benefit from the schools, and "old people" born before 1962, and who could not benefit at all from the schools.

So in total, we have four groups: YOUNG and High program, OLD and high program,.... See the DD in the table.

Under what assumption is the DD is good measure of the program? Is is likely to be satisfied?

CONTROL EXPERIMENT

We have a possibility to check that the assumption is not rejected in the available data.

Slide 9

Suppose we fill the same boxes, but we now compare the "OLD" to the "VERY OLD". Neither of them benefited from the program: what do we expect to see if the assumption is satisfied? What do we expect to see if the assumption is not satisfied?

Table:what do we see?

EXTENDING DIFFERENCE IN DIFFERENCES

(1) Using all the regional variation

There are 280 districts in Indonesia, and we know how many schools each district has received: grouping the region into two groups is throwing away some information!

Slide 10

Before, we had 2 regional group, and 2 age group, we formed 4 age-region group. Now we have 280 regional group, 2 age group, how many groups can we form? What are these groups?

First, we form the average for each group. We will note S_{Yj} the average education of the young in any region j, and S_{Oj} the average education of the young in any region j.

What can we do next?
-Take the difference between young and old in all the regions
-Plot the differences against the number of school constructed per 1000 child during the INPRES program (see graph)
- What do we wee? What does this suggest?
- Suppose we run the regression:

Slide 11

$$S_{Yj} - S_{Oj} = \alpha P_j + v_j \quad (1)$$

Where can you see the slope of this regression?
- See table 4: what is the result of running this regression?
What can we conclude?
Under what assumption is this conclusion valid?
Any suggestion to test this assumption?

(2) Using regional and age variation

The last generalization is that we don't have only 3 age groups (young, old, and very old): we have 23 age group (everybody born between 1950 and 1972).

Slide 12

Note S_{j2}, S_{j3},....,S_{jk},... S_{j24} the average education of people born in region j, and who were of age 2, 3, ... k,...24, when the program started.

Suppose we run the regression:

$$S_{j2} - S_{j24} = \alpha_2 P_j + v_{j2}$$

What is α_2?

What is α_{23}? What should α_{23} be equal to?

In general, suppose that for all ages k we run the regression:

$$S_{jk} - S_{j24} = \alpha_k P_j + \upsilon_{jk}$$

For what values of k should we see a positive α_k? (remember that children attend primary school until age 12). Should we see the coefficient be larger for younger children or older children?

Slide 13

Run the regressions in one operation:

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \sum_{l=2}^{23}(P_j * d_{il})\gamma_{1l} + \sum_{l=2}^{23}(\mathbf{C_j} * d_{il})\delta_{\mathbf{11}} + \epsilon_{ijk}, \quad (2)$$

Figure 2: Do the dots have the expected pattern?

ESTIMATING RETURNS TO EDUCATION

- ➜ Do the same for wage : Same patterns

$$y_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + \sum_{l=2}^{23}(P_j * d_{il})\gamma_{1l} + \sum_{l=2}^{23}(\mathbf{C_j} * d_{il})\delta_{\mathbf{11}} + \epsilon_{ijk}, \quad (3)$$

Slide 14

- ➜ What are the assumptions necessary to interpret this as the effect of education on wage?
- ➜ Consider using the policy to construct instruments for an instrumental variables of the effect of education on wages.
- ➜ What would be an IV estimate in the DD case?
- ➜ What would be an IV estimate in the difference in age case?
- ➜ What are candidates for the excluded instruments in equation **??**.

- ➜ We estimate the equation:

$$y_{ijk} = d + \alpha_j + \beta_k + S_{ijk}b + \eta_{ijk}, \quad (4)$$

Slide 15

using the interaction age*region of birth as instruments.

- ➜ Think about exclusion restrictions. What are the potential threat to them? What can be done to try to address them.
- ➜ Think about the interpretation of the IV. Who is affected by the experiment? Who are the returns to education calculated for

MARKET EQUILIBRIUM EFFECTS OF EDUCATION

- ➜ This experiment compares two people working in the same labor market, one of them having received more education than the other.
- ➜ What are we missing when we do this comparison?
- ➜ Potential externalities
  - ➜ Positive
  - ➜ Negative
- ➜ Heckman's criticism of experiments/natural experiments: by doing local comparison, they miss the market equilibrium effects.

Slide 16

ESTIMATING MARKET EQUILIBRIUM EFFECTS OF EDUCATION

Slide 17

- ➜ Ideal experiment: randomly assign different levels of education to different entire *markets*.
- ➜ The INPRES experiment does something that approximates this ideal experiment.
- ➜ Consider the older people who leave in a regions where many schools were built.
  - ➜ They did not directly benefit from the schools
  - ➜ However, as the newly educated cohorts enter the labor market, the average level of education in the labor market increases.
  - ➜ It increases more in schools where more schools were built.

Slide 18

- ➜ First stage:

$$S_{jt} = \mu_t + \nu_j + \sum_{l=1987}^{1999} (\lambda_l * P_j)\gamma_{1l} + \epsilon_{jt} \tag{5}$$

where $\overline{S_{jt}}$ is the average education in district j in year t, other notation as before.

- ➜ We seek to estimate the structural equation

$$\ln(w_{ijt}) = S_i b_{jt} + \alpha_U S_{jt} + \epsilon_{jt} + \mu_t + \nu_j + v_{ijt}, \tag{6}$$

where i is the individual, S_i is individual education, and b_{jt} are the returns to education in district j in year t.

Slide 19

- ➜ Why are the returns to education indexed by j and t.
- ➜ Which problems will we run into if we estimate this equation with OLS?
- ➜ First part of the strategy: focus on the old and average their salary by district and year.

$$\overline{\ln(w_{ijt})} = \overline{S_{jto}} b_{jt} + S_{jt}\alpha_U \epsilon_{jt} + \mu_t + \nu_j + \overline{v_{ijt}}, \tag{7}$$

Slide 20

- ➜ Second part of the strategy: differencing over time. Assume you have only two years:

$$\overline{\ln(w_{ijt})} - \overline{\ln(w_{ijt-1})} = (\overline{S_{jto}} - \overline{S_{jt-1o}})b_{jt-1}$$
$$+\overline{S_{jt-1o}}(b_{jt}-b_{jt-1})+(S_{jt}-S_{jt-1})\alpha_U+\mu_t-\mu_{t-1}+\epsilon_{jt}-\epsilon_{jt-1}+\overline{v_{ijt}}-\overline{v_{ijt-1}}$$

- ➜ because the old where not affected this reduces to:

$$\overline{\ln(w_{ijt})} - \overline{\ln(w_{ijt-1})} = (S_{jt} - S_{jt-1})\alpha + \mu'_t + \epsilon'_{jt}, \tag{8}$$

- ➜ Use P_j as instrument for $(S_{jt} - S_{jt-1})$
- ➜ What went into α?
- ➜ What went into the error term? Why is it legitimate to do that?

RESULTS

With many years of data, reduced form is written:

$$\overline{\ln w_{jt}} = \mu_t + \nu_j + \sum_{l=1987}^{1999} (\lambda_l * P_j)\gamma_{2l} + \delta_{2l} + \epsilon_{jt}, \quad (9)$$

Slide 21

We can now estimate

$$\overline{\ln(w_{ijt})} = +S_{jt}\alpha\epsilon_{jt} + \mu_t + \nu_j + \overline{v_{ijt}}, \quad (10)$$

OLS: positive effect. IV: negative effect, significant at 10% in rural areas.

CONCLUSION

Slide 22

- ➜ Education has positive returns
- ➜ No evidence of externalities.
- ➜ There are no convincing studies finding positive externalities.
- ➜ Why should government finance education?