Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment

Duflo (2011) - American Economic Review

The questions of whether investments in infrastructure can cause an increase in educational attainment, and whether an increase in educational attainment causes an increase in earnings are basic concerns for development economists.

This paper exploits a dramatic change in policy to evaluate the effect building schools has on education and earnings in Indonesia. In 1973, the Indonesian government launched a major school construction program, the Sekolah Dasar INPRES program. Between 1973 - 1974 and 1978 - 1979, more than 61,000 primary schools were constructed-an average of two schools per 1,000 children aged 5 to 14 in 1971.

Using a large cross section of men born between 1950 and 1972 from the 1995 intercensal survey of Indonesia (SUPAS), the author linked an adult’s education and wages with district-level data on the number of new schools built between 1973-1974 and 1978-1979 in his region of birth.

The exposure of an individual to the program was determined both by the number of schools built in his region of birth and by his age when the program was launched. After controlling for region of birth and cohort of birth effects, interactions between dummy variables indicating the age of the individual in 1974 and the intensity of the program in his region of birth are plausibly exogenous variables, and are used as instruments in the wage equation.

The estimates suggest that each new school constructed per 1,000 children was associated with an increase of 0.12 to 0.19 in years of education and 1.5 to 2.7 percent in earnings for the first cohort fully exposed to the program. This implies estimates of economic returns to education ranging from 6.8 to 10.6 percent.

The baseline specifications\(^1\) using data from 1995 and having Schooling and Log of wages as outcome variables (\(y_{ijk} \in \{S_{ijk}, w_{ijk}\}\)) could be summarized as:

\[
y_{ijk} = c + \alpha_j + \beta_k + \gamma \cdot (P_j \times T_i) + \delta \cdot (C_j \times T_i) + \epsilon_{ijk}
\]

Where \(i\) indexes individuals, \(j\) indexes districts and \(k\) indexes the cohort of birth of the individual in 1974. \(P_j\) represents the intensity (High or Low) of the program in district \(j\).

\(^1\)These are the most basic set of estimation equations. I’m using just two definitions of cohorts and two intensities of the program. In the document, however, the author presents a detail analysis using a full set of cohorts of birth and using the complete variation in the intensity of the program (not just using the dichotomous definition of High and Low).
and $T_i$ is an indicator variable equals one if the individual $i$ was 2 to 6 years old in 1974 (a.k.a Young Cohort); $C_j$ is a vector of district characteristics including enrollment rate in 1971 and the allocation of water and sanitation programs.

In equation (1) the the sample of individuals used are people from 2 to 6 or 12 to 17 years old in 1974 (a.k.a Old Cohort). Under the assumption that previous to the implementation of the program places with high and low levels of the program exhibited a similar trend in educational attainment and earnings (Parallel trend assumption) the coefficient of interest $\gamma$, that captures the effect of the program on $y_{ijk}$.

Furthermore, under the additional assumption that the program had no effect on wages other than increasing educational attainment (Exclusion Restriction), the author estimates the returns to education running a 2SLS model as follows:

First Stage: 
$$S_{ijk} = c + \alpha_j + \beta_k + \gamma_1 \cdot (P_j \times T_i) + \delta \cdot (C_j \times T_i) + \epsilon_{ijk}$$

Second Stage: 
$$w_{ijk} = d + \alpha_j + \beta_k + \gamma_2 \cdot \hat{S}_{ijk} + \delta \cdot (C_j \times T_i) + \eta_{ijk}$$

An therefore $\gamma_2$ captures the effect of one year of education on earnings. Moreover, the author also presents the 2SLS estimates correcting for Selection into the labor market using 1) The propensity score of being selected to the labor market (Similar but less powerful than the “Heckit” approach that we reviewed in class) and 2) additional results imputing the average earnings of sector of activity for self-employed individuals (Those who do not have information on wages in the intercensal data).

Finally, the author presents a cost-benefit analysis for the program concluding that the internal rate of return of this investment in infrastructure was quite high between 8.8 and 12 percent.

Two key graphs to check in the presentation: