ELSEVIER
Journal of Development Economics 74 (2004) 163–197
JOURNAL OF Development ECONOMICS
www.elsevier.com/locate/econbase

The medium run effects of educational expansion: evidence from a large school construction program in Indonesia

Esther Duflo*

Department of Economics, MIT, E52-252g, 50 Memorial Drive, Cambridge, MA 02142, USA

Abstract

This paper studies the medium run consequences of an increase in the rate of accumulation of human capital in a developing country. From 1974 to 1978, the Indonesian government built over 61,000 primary schools. The school construction program led to an increase in education among individuals who were young enough to attend primary school after 1974, but not among the older cohorts. 2SLS estimates suggest that an increase of 10 percentage points in the proportion of primary school graduates in the labor force reduced the wages of the older cohorts by 3.8–10% and increased their formal labor force participation by 4–7%. I propose a two-sector model as a framework to interpret these findings. The results suggest that physical capital did not adjust to the faster increase in human capital.

JEL classification: O15; O11; O12; I2
Keywords: General equilibrium; Externalities; Social returns; Education; Adjustment

1. Introduction

Evaluations of social programs in developing economies tend to focus on the short run and "partial equilibrium" effects of these programs, and do not try to assess their macroeconomic consequences. Empirical studies of the determinants of economic growth form a largely independent subfield that uses predominantly cross-country data sets. This division is unfortunate. While aggregate cross-country data is readily available and simple to use, it can lead to misleading conclusions, either because aggregate data is of poor

* Tel.: +1-617-2587013; fax: +1-617-2531330.
E-mail address: eduflo@mit.edu (E. Duflo).

0304-3878/$ - see front matter
doi:10.1016/j.jdeveco.2003.12.008

quality (Krueger and Lindahl, 2001; Atkinson and Brandolini, 2001) or because regressions are mis-specified (Banerjee and Duflo, 2000). Conversely, policy recommendations based on "partial equilibrium" analysis can be misleading, if the "general equilibrium" effects undo the direct effects of the policy (Heckman et al., 1998).

Moreover, the aggregate response to large programs is of independent interest and can be a fruitful source for identifying macroeconomic relationships. In particular, large programs are well-identified shocks. Studying the economy's aggregate response to these shocks is an occasion to understand the process of adjustment. The adjustment of an economy to shocks is the objective of macroeconomic studies of the "medium run" (Solow, 2000). In particular, macroeconomists and labor economists have long been interested in labor supply shocks, such as changes in cohort sizes (Welch, 1979), the level of education of the labor force (Katz and Murphy, 1992), changes in level of education by cohorts (Card and Lemieux, 2001), or adverse labor supply shocks in Europe (Blanchard, 1997). The speed and efficiency of adjustment are important dimensions of the effects of a range of economic policies. Trade policy analysis, for example, often assumes immediate adjustment of production decisions, which could be extremely misleading. The long-term effects of economic crises, as well as the appropriate policy response, are closely linked to whether they lead to efficient or inefficient restructuring, which is linked to the ability of the economy to allocate factors efficiently (Caballero and Hammour, 2000). Thus, studying the aggregate consequences of large programs can inform economic policy beyond the specific program considered.

Most studies of the medium run consequences of labor supply shocks focus on the US or on Europe. Yet, the response of the economy to a shock is closely related to its market institutions.[1] In developing countries, because market institutions (credit market, contractual enforcement and labor market regulations) are less effective, one might expect the adjustment process to be particularly sluggish. Caballero and Hammour (2000) argue that institutional failures, because they lead to mis-allocation of resources and inefficiently slow restructuring, are at the root of under-development. The evidence on medium term adjustment in developing countries is, however, extremely limited.

This paper studies the effects of a dramatic policy change that had differential effects on different cohorts and different regions of Indonesia on the allocation of the labor force across sectors and on wages. 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 built. In earlier work (Duflo, 2001), I showed that the program had an impact on the education and wages of the cohorts exposed to it. This paper studies the behavior of wage rates and formal labor force participation of those who were not directly exposed to the program, from 1986, 12 years after the school construction program was initiated (this is when the first generation exposed to the program first entered the labor force) to 1999. This is therefore a study of the "medium" run aggregate effects of the program. Until 1997, this was a period of rapid growth for the Indonesian economy: between 1986 and 1999, the economy grew by over 50% and the

[1] See Blanchard and Wolfers (2000) for a comparison of the reaction to the labor supply shocks in the 1970s across European countries with different labor market institutions.

share of the labor force in manufacturing doubled (from 6% to 13%). Industrialization occurred throughout Java and in concentrated pockets in the other islands (Miguel et al., 2001).

I first show that the program led to faster increases in the fraction of primary school graduates in the regions where it was more important, between 1986 and 1999. This increase is strikingly similar to that which would have been predicted in the absence of any migration. I then proceed to look at the effect of the program on the wages and the formal labor force participation of the cohorts that were not directly exposed to it, because they were already out of school when the program started. This allows me to look at the impact of the increase in education on factor returns, for a population whose skill level is not affected. It turns out that wages increased *less rapidly* from year to year in regions that received more schools. This holds even after controlling for the factors that determined the initial allocation and may have caused different growth trajectories across these regions.

Using interactions between the survey year and the number of INPRES schools per 1000 children as instruments for the fraction of educated workers in the region therefore suggests a *negative* effect of the proportion of primary school graduates on individual wages, keeping the individuals' own skill level constant. On the other hand, an increase in the fraction of educated workers seems to cause an increase in the participation of both educated and uneducated workers in the formal labor market. The negative impact of average education on individual wages does not seem to be explained by selection bias caused either by selective migration or by selective entry into the formal labor market.

I propose a simple two sector model as a framework to interpret these effects and their magnitude. Individuals can work either in the informal or in the formal sector. In the informal sector, they are self employed and labor (skilled and unskilled) is combined with land, a fixed factor. In the formal sector, labor (skilled an unskilled) is combined with capital, and individuals earn a wage. The production function in the formal sector exhibits constant returns to physical and human capital combined. The fact that the increase in the share of educated workers led to a movement of workers from the informal to the formal sector indicates that the elasticity of substitution between labor and land in the informal sector is smaller than the elasticity of substitution between labor and capital in the formal sector. The elasticity of the supply of capital with respect to the share of educated labor determines the predicted effect of the program on wages in the model. I compare two polar versions of the model. The benchmark version assumes costless adjustment of the capital stock. In this case, in the period under study (1986–1999, 12–25 years after the program was initiated), physical and human capital should grow at the same rate and there should be no relative fall in wages in regions where human capital grows faster. This holds in a closed economy model as well as in an open economy model where capital is accumulated nationally and efficiently allocated across regions. The second version, in contrast, compares the empirical estimates I obtain to the parameters predicted by the model in the absence of *any* adjustment of capital to the increase in education. These empirical estimates are close to what this version of the model would predict. This suggests that physical capital did not adjust to the regional differences in the rate of accumulation of human capital induced by the program.

The remainder of this paper is organized as follows. In Section 2, I describe the INPRES program and its effects on average education. In Section 3, I discuss the identification of the effects of average education on individual wages and derive the empirical specifications. Section 4 presents the results. Section 5 presents the model that organizes and explains the findings, and compares the estimates to what the two polar versions (costless adjustment of capital or no adjustment of capital) of the model would predict.

2. The program and its effects on average education

2.1. The Sekolah Dasar INPRES program

In 1974, the Indonesian government initiated a large primary school construction program, the Sekolah Dasar INPRES program. Between 1974 and 1978, 61,807 new buildings were constructed, doubling the number of available schools per capita. More schools were put in regions where initial enrollment rates were low, which caused important regional variations in the intensity of the program. Using a large household survey conducted in 1995 (the SUPAS 1995) linked to data on the number of schools constructed in each individual's region of birth, Duflo (2001) showed that the growth in education between cohorts unexposed to the program and cohorts exposed to the program was faster in regions that received more INPRES schools. This difference can be attributed to the program with a reasonable level of confidence, because no similar pattern is present when comparing cohorts that were not exposed to the program. In addition, the program affected mostly primary school completion, whereas omitted factors would have affected other levels of schooling as well. This pattern is summarized in Fig. 1, reproduced from

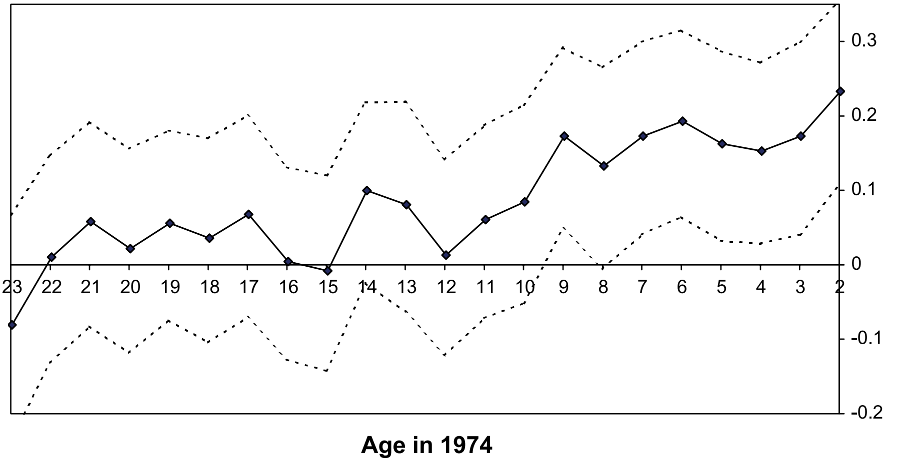

Fig. 1. Coefficients of the interactions of age in 1974* program intensity in the region of birth in the education equation.

Duflo (2001). Each point on the solid line summarizes the effect one more school built per 1000 children had on the average education of children born in each cohort.[2] Children in Indonesia normally go to primary school until age 12 (although delay at school entry and repetition are not uncommon); therefore, one would expect the effect of the program to be 0 for children who reached 12 before 1974, when the first schools were built, and increase progressively as the program affects younger cohorts. This is exactly what the picture shows.

If migration flows were either small or not affected by the program, one would expect to see a similar pattern when comparing the evolution of average education among adults over the years in the different regions. As the generations exposed to the program enter the labor market, one should see the average education (and in particular the fraction of primary school graduates) increase faster in the regions that received more schools.

2.2. Data and empirical specification

The data for this paper comes primarily from the annual Indonesian Labor Force Survey (SAKERNAS), from 1986 to 1999. These surveys are repeated cross sections, of approximately 60,000 households. The surveys contain information on province and district (kabupaten) of residence (but not of birth), education level achieved, labor force participation, type of employment, number of hours worked in the last week and wages for individuals who work for a wage in their primary occupation. I restrict the sample to men. Using this data, I construct the average hourly wage as weekly wage divided by hours worked on this occupation. An individual is considered as part of the formal sector if he works for a wage in his primary occupation. The survey questions and definitions are homogenous between 1986 and 1999. I restrict the sample to males aged 20–60, and I exclude Jakarta, where migration makes it difficult to compare samples across years. Descriptive statistics are presented in Table 1. The fraction of individuals born after 1962, and therefore theoretically exposed to the program, in the age groups 20–40 and 20–60, increases progressively over the years. I consider the proportion of primary school graduates in each region in each year. There are a total of 3826 district-year cells, with an average of 287 individual observations in each cell in the full sample.[3] All regressions are performed on this aggregate data set, and each cell is weighted by the number of observations used to construct it.

Consider comparing two regions in 1986 and 1999, one which received a large number of INPRES schools per capita, while the other received a small number of schools. The fraction of people who were young enough to be exposed to the INPRES program is bigger in 1999 than in 1986. We know that the gains in years of education of these younger cohorts, relative to the older ones, were bigger in the regions that received more schools. If the effect of the program was not undone by migration, one would expect the

[2] These are the coefficients obtained by regressing years of education on the interactions between the number of schools built per capita in the individual's region of birth and year of birth dummies, after controlling for year and region of birth fixed effects.

[3] There are on average 185 observations per cell of individuals born before 1962, including 61 with wage data.

Table 1
Descriptive statistics

	Fraction born after 1962		Percentage of primary school graduates		Ratio of uneducated/educated informal sector	Skill premium	Implied β
	20–40	20–60	20–40	20–60	(20–60)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1986	0.22	0.15	0.69	0.61	0.36	0.40	0.53
1987	0.25	0.16	0.71	0.63	0.33	0.41	0.55
1988	0.32	0.21	0.73	0.64	0.29	0.33	0.53
1989	0.36	0.23	0.74	0.65	0.29	0.39	0.57
1990	0.42	0.28	0.76	0.68	0.29	0.35	0.55
1991	0.46	0.31	0.79	0.71	0.22	0.31	0.58
1992	0.49	0.32	0.79	0.71	0.23	0.35	0.60
1993	0.55	0.36	0.8	0.72	0.19	0.41	0.68
1994	0.58	0.38	0.81	0.74	0.21	0.36	0.63
1995	0.62	0.42	0.79	0.72	0.21	0.43	0.67
1996	0.67	0.44	0.85	0.79	0.23	0.37	0.62
1997	0.71	0.46	0.84	0.78	0.17	0.41	0.71
1998	0.77	0.51	0.87	0.81	0.17	0.38	0.69
1999	0.82	0.53	0.88	0.82	0.13	0.35	0.73

1. Average wage is average of the log of monthly wage. 2. Means are weighted by the number of observations in each district-year cell. 3. See text for the definition of β.

average education (in particular, the proportion of primary school graduates) to have grown faster between 1986 and 1999 in the region that received more schools. This suggests comparing the difference in educational attainments between 1999 and 1986 in these two regions. More generally, this suggests that, if one runs a regression of the difference between average educational attainment in 1999 and 1986 on the number of schools per capita built in each region, one should see a positive coefficient. Clearly, this reasoning also applies to any year-to-year difference.

In summary, this suggests the following specification:

$$\overline{S_{jt}} = \mu_t + v_t + \sum_{l=1987}^{1999} (\lambda_l * P_j)\gamma_{1l} + \sum_{l=1987}^{1999} (\lambda_l * C_j)\delta_{1l} + \epsilon_{jt} \tag{1}$$

where $\overline{S_{jt}}$ is the proportion of primary school graduates among adults in year t in region j, μ_t is a survey year fixed effect and v_j is a region fixed effect; P_j is the number of INPRES schools built between 1974 and 1978 in district j, and λ_l is a survey year dummy ($\lambda_l = 1$ if $t = l$ and 0 otherwise). C_j is a vector of initial conditions that are introduced as control variables. In particular, it may be important to control for the enrollment rate in 1971, since it was a determinant of the placement of the program. Note that the first-order effect of a higher enrollment rate in 1971 is a difference in *level* of education, which should affect all cohorts, and all survey years identically, and therefore be absorbed by the region fixed effect. Only a *change* in the rates at which children attend school in a region will lead to a change in the rate at which average education increases from year to year. Therefore, controlling for the enrollment rate in 1971, interacted with year dummies, is important only to the extent that changes in enrollment rates are correlated with levels. I also control for the number of children in 1971.

2.3. Results

Since the generations exposed to the program have already started entering the sample in 1986, one would expect all the coefficients of the interactions between program intensity and survey year dummies to be positive and increasing. Columns 1, 2, 4 and 5 in Table 2 show these coefficients for the specification that includes enrollment rates as a control, for adults aged 20–40, and for adults aged 20–60, in the whole sample and in a sample that excludes urban districts.[4] The fraction of "young" (or exposed) people among individuals aged 20–40 increases faster than among those aged 20–60, and one would expect the coefficients to be larger and more significant in the former group. The group of individuals aged 20–60, however, corresponds better to "the labor market", and will therefore be important in the second stage of this analysis. The coefficients are increasing for both groups, they are jointly significant, and, as expected, they are larger in the group aged 20–40. They become individually significant from 1991 in the 20–40 group and from 1996 in the 20–60 group.

This pattern could have been caused by factors other than the increase in education due to INPRES, for example by migration of educated workers into districts that received more INPRES schools. If I had data from earlier years, it would be possible to use "pre-program" data to test the identification assumption that the increase in education levels over time would not have been systematically different in regions where a different number of schools was built, even in the absence of the program. No comparable survey was realized before 1986. If the pattern was due to something other than the effect of the program on education, however, one would see a faster (or slower) increase in the education over the years, even in the subsample of those who were not exposed to the program (individuals born in 1962 or before). To check this, I estimated a specification similar to Eq. (1), with the fraction of primary school graduates among individuals born in 1962 or before as the dependent variable. The coefficients are presented in columns 3 and 6 in Table 2. Fig. 2 gives a graphical summary of these estimates: it shows the coefficients (and their confidence intervals) for the entire group aged 20–40, and for the group of individuals born before 1962. There is no systematic increase among the group born before 1962. The coefficients in the two equations are significantly different from each other. This indicates that the increase in average education is likely due to the program, rather than to other factors.

2.4. Does migration undo the effect of local infrastructure development?

Although migration flows were not very important in Indonesia over the period, they are far from negligible. In 1995, if one excludes Jakarta, 12% of individuals in the SUPAS

[4] The specification without enrollment rates as a control is very similar and is therefore omitted.

Table 2
First stage regressions

Sample	Sample: urban and rural areas			Sample: rural areas only		
	20–40	20–60	Born before 1962	20–40	20–60	Born before 1962
	(1)	(2)	(3)	(4)	(5)	(6)
1986	omitted	omitted	omitted	omitted	omitted	omitted
1987	0.0028	−0.0004	−0.0026	0.0009	−0.0015	−0.0023
	(0.008)	(0.0075)	(0.0077)	(0.0099)	(0.009)	(0.0091)
1988	0.0015	−0.0028	−0.0050	−0.0032	−0.0079	−0.0088
	(0.0064)	(0.0058)	(0.0059)	(0.0078)	(0.0069)	(0.0069)
1989	0.0014	−0.0012	−0.0048	−0.0010	−0.0043	−0.0068
	(0.0053)	(0.0047)	(0.005)	(0.0066)	(0.0056)	(0.0061)
1990	0.0027	−0.0042	−0.0106	−0.0002	−0.0071	−0.0126
	(0.0061)	(0.0056)	(0.0059)	(0.0072)	(0.0065)	(0.007)
1991	0.0151	0.0074	−0.0002	0.0110	0.0032	−0.0034
	(0.0058)	(0.0055)	(0.0055)	(0.0067)	(0.0062)	(0.0065)
1992	0.0124	0.0006	−0.0074	0.0082	−0.0058	−0.0126
	(0.0065)	(0.0061)	(0.0061)	(0.0076)	(0.0068)	(0.007)
1993	0.0205	0.0088	−0.0017	0.0143	0.0012	−0.0082
	(0.0065)	(0.0062)	(0.0059)	(0.0072)	(0.0067)	(0.0067)
1994	0.0195	0.0076	−0.0044	0.0145	0.0005	−0.0103
	(0.0068)	(0.0068)	(0.0069)	(0.0074)	(0.0072)	(0.0078)
1995	0.0202	0.0074	−0.0064	0.0161	0.0003	−0.0127
	(0.0064)	(0.0063)	(0.0061)	(0.0072)	(0.0068)	(0.007)
1996	0.0260	0.0112	−0.0027	0.0207	0.0016	−0.0122
	(0.0064)	(0.0066)	(0.0069)	(0.007)	(0.0066)	(0.0075)
1997	0.0252	0.0149	0.0039	0.0190	0.0070	−0.0010
	(0.0071)	(0.0072)	(0.0076)	(0.0077)	(0.0076)	(0.0086)
1998	0.0340	0.0190	0.0044	0.0300	0.0110	−0.0040
	(0.0077)	(0.0074)	(0.0072)	(0.0073)	(0.0066)	(0.0076)
1999	0.0290	0.0154	0.0045	0.0253	0.0076	−0.0029
	(0.0082)	(0.0082)	(0.0083)	(0.0078)	(0.0073)	(0.0086)
Number of cells	3826	3826	3826	3140	3140	3140
F-statistic	7.23	3.40	1.29	5.02	1.68	0.95

Effect of the program on the proportion of individuals completing primary school or more.
Coefficients of interactions of the intensity of the program and survey year dummies.
1. The program intensity is the number of INPRES schools built between 1974 and 1978, divided by the number of children in 1971. 2. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions. 3. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell. 4. The *F*-statistic is for the hypothesis that the set of interactions is jointly insignificant. 5. The standard errors are corrected for auto-correlation within kabupaten.

sample did not live in their province of birth and 24% did not live in their district of birth (17% if one excludes the urban districts). Among individual born in 1962, 13% did not live in their province of birth and 25% did not live in their district of birth. It is therefore interesting to study whether out migration of educated workers dampened the effect of the program on average education in the labor market.

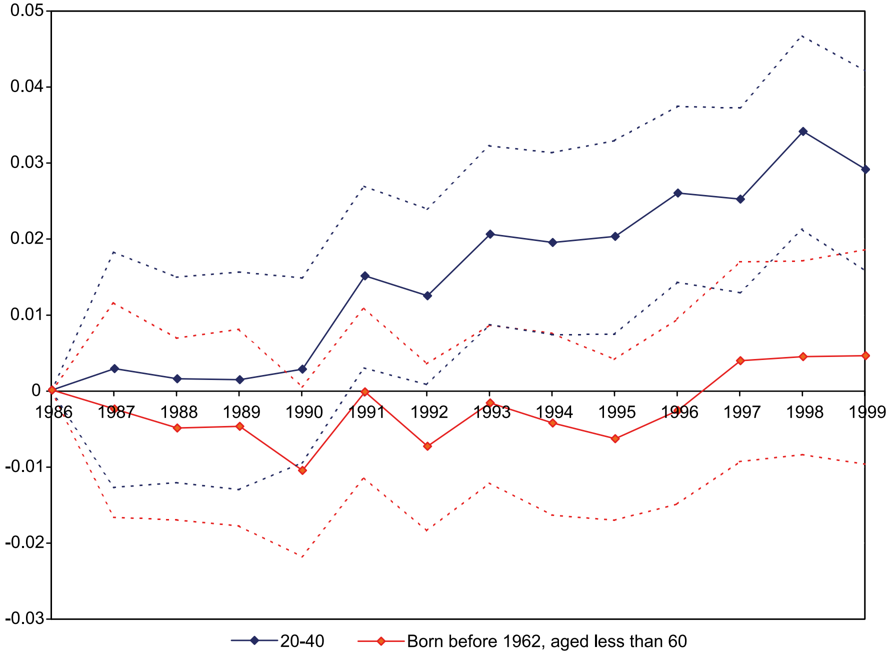

Fig. 2. Coefficients of the interactions of program intensity and survey year dummies. Dependent variable: % of primary school graduates.

The results in Section 2.3 already suggest a partial answer to this question. The program affected the average education of adults, which indicates that its effects were not totally undone by migration. We can, however, make this result more precise by comparing the effect the program should have had, in the absence of any off-setting effect of migration, with the effect it actually had. The results from this exercise are important in the context of an increasing focus on decentralization, notably in Indonesia. If local governments believe their communities are not getting any benefits from investment in education because educated people migrate (with their human capital), decentralization of school finance may lead the public financing of education to decline.[5]

To get at this question, I first estimated the effect of the number of INPRES schools constructed per capita in an individual's district of *birth* on the probability that an individual completed primary school, for each cohort. To this end, I used the SUPAS 1995 data. The SUPAS (Intercensal Survey of Indonesia) is a sample of over 200,000 households. It is representative at the district level. It is conducted every 10 years by the Central Bureau of Statistics of Indonesia. The survey collects the same information as the SAKERNAS (which it replaced in 1995), as well as more information about household members, including their province and district of birth. The sample for this analysis is men born between 1950 and 1972 (there are 152,989 individuals in the sample).

[5] Bound et al. (2000) ask the same question for college education in the US.

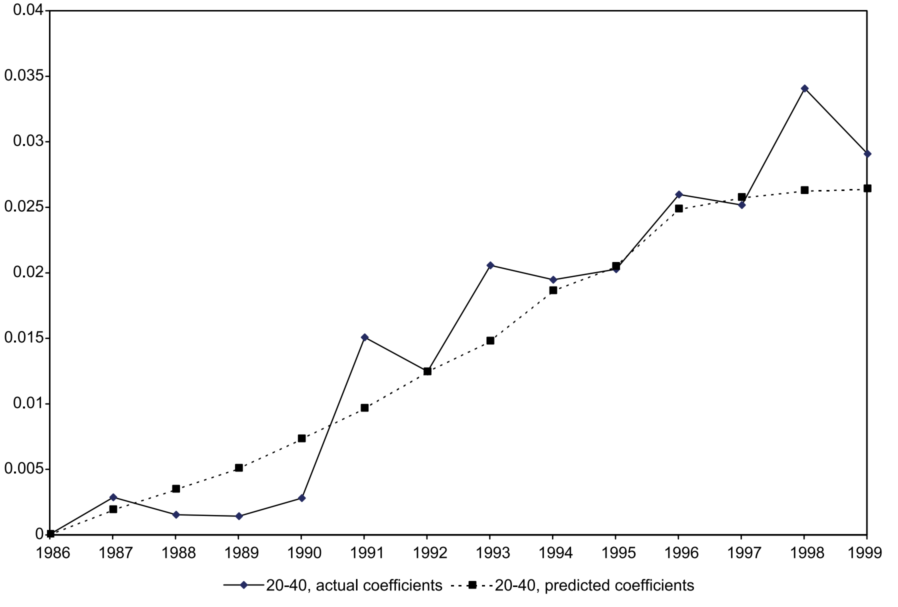

Fig. 3. Actual and predicted coefficients of the interactions of program intensity and survey year dummies. Dependent variable: % of primary school graduates.

Using this data, I regressed a dummy indicating whether an individual completed primary school on a set of district of birth fixed effects, cohort of birth fixed effects, and interactions between the number of schools constructed in one's district of birth and year of birth dummies. The equation estimated is identical to Eq. (11) in Duflo (2001), except that it used the primary school completion (instead of years of education) as the dependent variable. Denote the estimated effect of the program on the cohort born in year k as $\hat{\pi}_k$. I used the 1995 data to compute the proportion of primary school graduates among those aged 20–40 in each year (from 1986 to 1999) *born* in each district before 1962. Denote this average by $\widetilde{S_{oj}}$. For each survey year t and year of birth k, denote by ϕ_{kt} the share of those aged 20–40 who were born in year k.[6] We can then compute the proportion of primary school graduates predicted by the program in each district and each year as:

$$\widetilde{S_{jt}} = \widetilde{S_{oj}} + P_j\left(\sum_{k=1962}^{t-20} \hat{\pi}_k \phi_{kt}\right) \tag{2}$$

Note that this predicted value does not contain any information specific both to the district and the year considered. There is therefore no source of mechanical relationship between $\widetilde{S_{jt}}$ and $\overline{S_{jt}}$. The first observation is that $\widetilde{S_{jt}}$ and $\overline{S_{jt}}$ are strongly correlated. The regression of the actual share of primary school graduates on the predicted share leads to a coefficient of 0.84 (with a t-statistic of 77). This, however, is not very informative, because a large part of this correlation is driven by those born before 1962, and would therefore

[6] All the individuals who are born in 1962 or before are in the same cohort k.

still exist even if all the educated young had migrated out of the high program districts. The following experiment is more informative. I run the same specification as in Eq. (1), but I use as the dependent variable the *predicted* education, $\widetilde{S_{jt}}$. These coefficients indicate how the average education of adults would have been affected in each region in the absence of any offsetting effect of migration. The coefficients γ_{1l} obtained in this specification are plotted in Fig. 3, along with the coefficients obtained when estimating Eq. (1) with the actual average as the dependent variable. The two sets of coefficients are surprisingly close to each other. In particular, there is no evidence that the predicted effect is bigger than the actual effect.

3. Identifying the effect of a change in average education

3.1. Conceptual framework

Consider an economy with two sectors. Assume that there are only two types of workers, educated (with a primary education or more) and uneducated (no primary education). Assume that the formal sector employs educated labor, uneducated labor and capital, and the informal sector employs educated and uneducated labor and land.[7] Individuals are self-employed in the informal sector, and receive a wage in the formal sector. As in Harris and Todaro (1970) and other dual economy models of development, economic growth happens as the formal sector expands. This is reflected in the assumption that land is a fixed factor. The share of the labor force employed in the formal sector and their wages are the two variables of interest.

The production functions in the formal and informal sectors are given by $fc(A_F, E_F, U_F, K)$ and $g(A_I, E_I, U_I, T)$, respectively, where K and T are the stock of capital and land, respectively; E_F and U_F denote educated and uneducated labor employed in the formal sector, respectively; E_I and U_I denote educated and uneducated labor employed in the informal sector; and A_F and A_I are "productivity" parameters. We will treat the total population as a constant (nothing is affected by allowing steady population growth). The wages, as well as the fraction of educated and uneducated workers working in each sector, are determined jointly in equilibrium as a function of the number of educated and uneducated workers (in the economy as a whole), the stock of capital, the stock of land, and the parameters A_F and A_E.

Normalizing the entire labor force to 1, and denoting S the share of educated workers, we can therefore write the wage and formal employment functions as[8]: $\ln(w_E) = \phi_E(A_F, A_I, K(S), S, T)$, $\ln(w_u) = \phi_U(A_F, A_I, K(S), S, T)$, $E_F = \psi_E(A_F, A_I, K(S), S, T)$ and $U_F = \psi_U(A_F, A_I, K(S), S, T)$.

K is explicitly written as a function of S, to reflect the fact that a change in the proportion of educated workers has a direct effect (the effect of the share of educated

[7] I could allow land and capital to be present in both sectors, but we would have to model migration of capital and land between sectors, for which I have little data.

[8] Writing the wage function directly in logarithm, for convenience.

workers on the wage), and an indirect effect due to the accumulation of physical capital in response to this increase. The elasticity of physical capital with respect to the share of educated labor is an empirical question: in the long run, one might expect the physical capital to adjust to a change in the fraction of educated workers, while in the very short run, adjustment will be much more limited. In the "medium run", the speed of adjustment of the capital stock will depend on the flexibility of the production function and the availability of finance for the installation of new capital.

Consider a Taylor expansion of the wage function around $S=0$:

$$\phi_{\mathrm{s}}(A_{\mathrm{F}t}, A_{\mathrm{I}t}, K_t(S_t), S_t, T_t) \simeq \phi_{\mathrm{s}}(A_{\mathrm{F}t}, A_{\mathrm{I}t}, K_t(S_t = 0), S_t = 0, T_t) + S_t \frac{\partial \phi_{\mathrm{s}}}{\partial S} + S_t \frac{\partial K}{\partial S} \frac{\partial \phi_{\mathrm{s}}}{\partial K}.$$

Denoting $K_t(S_t=0)$ by $\tilde{K}_t$, this can be rewritten:

$$\ln(w_{\mathrm{s}t}) = \left(\frac{\partial \phi_{\mathrm{s}}}{\partial S} + \frac{\partial K}{\partial S} \frac{\partial \phi_{\mathrm{s}}}{\partial K}\right) S_t + \phi_{1\mathrm{s}}(A_{\mathrm{F}t}, A_{\mathrm{I}t}, \tilde{K}_t, T_t)$$

I, therefore, seek to estimate the coefficient α_{s} in the expression:

$$\ln(w_{\mathrm{s}t}) \simeq \alpha_{\mathrm{s}} S_t + \phi_{1\mathrm{s}}(A_{\mathrm{F}t}, A_{\mathrm{I}t}, \tilde{K}_t, T_t) \tag{3}$$

In this expression, α_{s} reflects the direct effect of S on the wage, as well as any indirect effect due to the α_{s} response of the stock of physical capital to the stock of human capital. The sign of is not determined *a priori*. If capital adjusts slowly, or if there are diminishing returns to capital and labor combined, α_{s} will tend to be negative, reflecting the fact that the increase in the share of educated workers increases the quantity of labor (measured in efficiency units). If capital adjusts rapidly and there is no fixed factor, α_{s} would be zero or even positive in the presence of an external effect of education (as in Lucas, 1988), or if an increase in the share of educated workers led to a more than offsetting increase in the stock of physical capital (Acemoglu, 1996).

Consider running a regression of the wage of the uneducated workers on the share of educated workers in the economy, without controlling for the stock of capital. Clearly, if there is any correlation between the level of capital (and the productivity of the formal and informal sector) and the share of educated people, the coefficient of average education will be upwardly biased.

Since there are many districts and years, one could instead compare wage growth across districts. Using Eq. (3), the growth of the log wage between two periods is given by:

$$\ln(w_{\mathrm{s}t}) - \ln(w_{\mathrm{s}t-1}) \simeq \alpha_{\mathrm{s}}(S_t - S_{t-1}) + \phi_{1\mathrm{s}}(A_{\mathrm{F}t}, A_{\mathrm{I}t}, \tilde{K}_t, T_t) - \phi_{1\mathrm{s}}(A_{\mathrm{F}t-1}, A_{\mathrm{I}t-1}, \tilde{K}_{t-1}, T_{t-1}) \tag{4}$$

If we estimate this relationship using an OLS regression, and we omit the term $\phi_{1\mathrm{s}}(A_{\mathrm{F}t}, A_{\mathrm{I}t}, \tilde{K}_t, T_t) - \phi_{1\mathrm{s}}(A_{\mathrm{F}t-1}, A_{\mathrm{I}t-1}, \widetilde{K_{t-1}}, T_{t-1})$, the coefficient α_{s} will be biased if there is a correlation between physical capital accumulation and human capital accumulation. In almost any model of human and physical capital accumulation based upon optimizing individuals, the increase in the share of educated workers and the rate of physical capital accumulation will be related: in particular, both are determined by the discount rate in the economy. In order to estimate the parameter α_s consistently, we therefore need an instrument, correlated with the increase in the share of educated workers, but not with the evolution in the other factors in the economy.[9]

A potential instrument for $(S_t - S_{t-1})$ in our setting is the number of primary schools constructed by the INPRES program. To understand how the instrument works and its limitations, suppose first that the government allocated the schools randomly. Each school reduces the effective cost of schooling, and therefore increases the enrollment rate among all the future young generations (but not that of the older generations). The increase in the number of schools combined with the fact that the young generations enter the labor market progressively starting in the late 1980s changes the rate of growth of S over time. The modification in the rate of growth is a function of the number of schools built, which suggests that if the schools had been allocated randomly, the number of INPRES schools would form an ideal instrument.

In practice, however, the government allocated more schools in regions where enrollment rates at the primary school level were lower. The evidence presented in this paper and in Duflo (2001) suggests that the rate of growth of human capital was not systematically correlated with the program before it was initiated. Nevertheless, the level of the program will not be a valid instrument for $(S_t - S_{t-1})$ if it is correlated with the rate of capital accumulation. This would happen if educational attainments in 1971 were correlated with capital accumulation between 1986 and 1999. Regions with a lower level of educational attainment tend to be poorer and could therefore have been growing faster, if there had been a tendency for Indonesian regions to converge. In practice, Indonesian provinces exhibited very little convergence in gross provincial product per capita until 1996 (Hill, 1996). Nevertheless, to control for possible convergence, I will control for enrollment rate in 1971. We will also present the results in the rural sample separately, and omit the years 1998 and 1999 to allow for the fact that the Indonesian crisis hit wages in richer regions, and in particular cities, much more than in poorer regions and in rural areas (Frankenberg et al., 1999), causing some convergence of wage rates between regions.

3.2. Empirical specifications

The wage of individual i observed in district j in year t is given by:

$$\ln(w_{ijt}) = S_i(\ln(w_{\mathrm{E}jt}) - \ln(w_{\mathrm{U}jt})) + \ln(w_{\mathrm{U}jt}) + v_{ijt}, \tag{5}$$

where S_i is a dummy indicating whether the individual has graduated from primary school. The error term v_{ijt} reflects all the other factors that determine wage, besides individual and average education.

[9] For example, Moretti (1999) proposed to instrument for $(S_t - S_{t-1})$ with the share of young people in the base year, on the grounds that education will grow faster in regions which have more young people. The problem remains, however, that the share of young people in the base year is very likely to influence physical capital accumulation as well.

Substituting the expression for $\ln(w_{Ejt})$ from Eq. (3) and including all the variables which we do not measure in the error term, we obtain a relationship between individual wage, individual education level and regional human capital in the district at date t:

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

where b_{jt}=$(\ln(w_{Ejt}) - \ln(w_{Ujt}))$ (the skill premium) and $\mu_t + v_j + \epsilon_{jt} = \phi_{1U}(A_{Fjt}, A_{Ijt}, \tilde{K}_{jt}, T_j)$.

As we have seen, estimating this equation by OLS (treating b_{jt} as a random coefficient) could be very misleading, because of the correlation between ϵ_{jt}, μ_t or v_j and S_{jt}. In addition, Acemoglu and Angrist (2000) show that, even if there is no omitted district level variable, the OLS estimate of α_U will be a biased estimate of the effect of S_{jt} on $\ln(w_{ijt})$ if the estimate of b_{jt} is biased for any reason (such as measurement error in the education variable or the endogeneity of education). They propose to instrument for both S_i and S_{jt}. Alternatively, one could instrument S_{jt} with a variable that does not affect S_i, an individual's education, which is the approach I take here.

The nature of the INPRES program suggests the following instrumental variable strategy. All individuals who were born in 1962 or before were not affected by the program (we have verified in Section 3.1 that the average education in this group did not grow faster from year to year in the districts that received more INPRES schools). On the other hand, the program affected the average education by affecting the education of those born *after* 1962. Therefore, the intensity of the INPRES program is a potential instrument for the average education, which does not affect individual education, when restricting the sample to those born before 1962.

To derive the empirical specification, take the average of Eq. (6) for all individuals born before 1962 in each district-year cell:

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

where $\overline{S_{jto}}$ is the proportion of primary school graduates among the old (born before 1962) in year t in district j. Taking the first difference of this equation and rearrange the terms, we obtain:

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

The evolution of the skill premium $(b_{jt} - b_{jt-1})$ is itself a function of the evolution in the number of educated workers in the region, so that the effect of the evolution of primary school graduates on the average wages of the individual born before 1962 is finally given by an expression of the following form:

$$\overline{\ln(w_{ijt})} - \overline{\ln(w_{ijt-1})} = (S_{jt} - S_{jt-1})\alpha + \mu_t' + \epsilon_{jt}' \tag{8}$$

Subject to the caveats discussed in Section 3.1, we can use the number of INPRES schools (P_j) as an instrument for $S_{jt} - S_{jt-1}$ in Eq. (8), possibly after controlling for variables such as the enrollment rate and the wage in 1986 (a vector C_j). We have verified that P_j is uncorrelated with $(\overline{S_{jto}} - \overline{S_{jt-1o}})$, which is now included in the error term.

A joint test of the validity of the strategy and the seriousness of the problem suggested by Acemoglu and Angrist (2000) is to use as dependent variable the average of the residual of a regression of individual wages on individual education. If this equation is correctly specified, it should lead to the similar, but more precise estimate (since $(\overline{S_{jto}} - \overline{S_{jt-1o}})$ will not be part of the error term anymore).

Thus the reduced form with 2 years of data would be written:

$$\overline{\ln(w_{jt})} - \overline{\ln(w_{jt-1})} = \mu_t + \gamma_2 P_j + \delta_2 C_j + \xi_{jt}$$

This equation can be generalized to incorporate all available years, leading to a reduced form equation similar to Eq. (1):

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

where μ_t and γ_j are year and district fixed effects, respectively.

Eqs. (1) and (9) form, respectively, the first stage and the reduced form of an instrumental variables strategy to estimate Eq. (7).

The same reasoning applies to formal labor force participation, and the same specification can be estimated with formal labor force participation instead of wages. Finally, we can also estimate equations similar to Eq. (7), using the average skill premium as dependent variable. The variables I consider here (wages, education, skill premium, formal labor force participation) are likely to be auto-correlated over time. Bertrand et al. (2001) show that this can cause severe downward bias in the estimated standard errors. I thus correct standard errors in all equations using a generalization of the White variance formula, which allows for a flexible auto-correlation process within any state.

Since the sample of individuals not affected by the program is different every year, this specification may suffer from sample selection. First, the program may have induced selective migration by old people, potentially correlated with their productivity, and therefore with their wages. Second, I will show that the program affected the proportion of old people who work for a wage: it also opens some room for selection bias, since it is possible that workers with the lowest productivity switched to the wage sector. Section 4.5 will present additional evidence (using two other data sets) on whether these two possibilities for sample selection affected the results.

4. Results

Summary statistics for the sample of people born before 1962, and aged 60 or less in the survey year, are presented in Table 3. The proportion of primary school graduates among them increased from 59% to 74% between 1986 and 1989 (this reflects the fact that individuals present in the sample belong to later cohorts in later years). We determine participation in the formal sector by noting whether an individual receives

Table 3
Descriptive statistics

Survey year	Individuals born before 1962, aged less than 60				
	% primary school graduates	% working for wage	Average wage	S.D. of wage	Skill premium
	(1)	(2)	(3)	(4)	(5)
1986	0.59	0.31	6.56	0.71	0.44
1987	0.60	0.32	6.58	0.66	0.48
1988	0.60	0.31	6.55	0.65	0.40
1989	0.61	0.33	6.61	0.66	0.46
1990	0.62	0.33	6.66	0.67	0.46
1991	0.65	0.34	6.70	0.66	0.44
1992	0.64	0.33	6.77	0.67	0.47
1993	0.65	0.34	6.81	0.73	0.51
1994	0.66	0.36	6.88	0.71	0.47
1995	0.64	0.36	6.93	0.72	0.57
1996	0.71	0.35	6.98	0.68	0.50
1997	0.69	0.36	7.08	0.72	0.54
1998	0.72	0.33	6.82	0.69	0.54
1999	0.74	0.33	6.86	0.70	0.57

Sample: individuals aged less than 60 and born before 1962.

a wage. This fraction is a little over 30%. The average wage, in real terms, increased by about 50% between 1986 and 1997, and declined by 22% between 1997 and 1999.

4.1. Reduced form results

The reduced form results (the estimates of the coefficients γ_{2l} in Eq. (9)) are presented in Table 4 and in Fig. 4a and b. These two figures summarize the reduced form effects on wages and on formal employment. Although none of these coefficients is individually significantly different from zero, the reduced form coefficients in the wage equation are declining (in contrast to the coefficients of average education, which are increasing). The reduced form coefficients on the probability of working for a wage are increasing. In the sample that includes both urban and rural areas, the coefficients increase from 1997 to 1999, which probably reflects the differential impact of the crisis. In the rural sample, they are monotonically declining.

4.2. The effects of average education on wage rates

The main sample for the analysis is all the individuals aged 20–60 who were born before 1962.[10] I will consider two independent variables. First, the fraction of primary

[10] It means that, for each survey year, there is both a cohort effect and an age effect. I have run all the specifications in a sample, which maintains a constant cohort composition, and the results were very similar.

Table 4
Reduced form regressions

	Sample: urban and rural areas 20–60 years old, born before 1962			Sample: rural areas only 20–60 years old, born before 1962		
	Wages	Residual wage	Formal employment	Wages	Residual wage	Formal employment
	(1)	(2)	(3)	(4)	(5)	(6)
1986	omitted	omitted	omitted	omitted	omitted	omitted
1987	0.0025	− 0.0064	− 0.0087	0.0209	0.0094	− 0.0059
	(0.0187)	(0.014)	(0.0056)	(0.018)	(0.0139)	(0.0056)
1988	− 0.0041	− 0.0051	− 0.0040	0.0108	0.0103	− 0.0047
	(0.0152)	(0.0141)	(0.004)	(0.0139)	(0.0136)	(0.0042)
1989	0.0022	− 0.0011	− 0.0074	0.0155	0.0095	− 0.0081
	(0.0161)	(0.0139)	(0.0047)	(0.0159)	(0.0141)	(0.0044)
1990	− 0.0140	− 0.0111	− 0.0027	0.0047	0.0056	− 0.0030
	(0.0155)	(0.0133)	(0.0054)	(0.0165)	(0.0143)	(0.0054)
1991	− 0.0061	− 0.0079	0.0035	0.0082	0.0047	0.0039
	(0.0187)	(0.0146)	(0.0051)	(0.0185)	(0.0149)	(0.0049)
1992	− 0.0173	− 0.0116	0.0019	− 0.0055	0.0021	− 0.0002
	(0.0163)	(0.0141)	(0.0055)	(0.0186)	(0.0145)	(0.0055)
1993	− 0.0166	− 0.0132	0.0069	− 0.0097	− 0.0038	0.0074
	(0.0156)	(0.0137)	(0.0055)	(0.0203)	(0.017)	(0.0054)
1994	− 0.0222	− 0.0189	0.0103	− 0.0149	− 0.0170	0.0068
	(0.0165)	0.0136)	(0.007)	(0.0224)	(0.0175)	(0.0069)
1995	0.0031	− 0.0010	0.0071	0.0044	0.0079	0.0058
	(0.0145)	(0.0134)	(0.0056)	(0.0188)	(0.015)	(0.0051)
1996	− 0.0190	− 0.0173	− 0.0035	− 0.0158	− 0.0116	− 0.0051
	(0.0176)	(0.0143)	(0.0071)	(0.0251)	(0.0178)	(0.0062)
1997	− 0.0192	− 0.0155	0.0061	− 0.0119	− 0.0166	0.0042
	(0.0174)	(0.0169)	(0.0068)	(0.0232)	(0.0183)	(0.0066)
1998	− 0.0079	− 0.0141	0.0083	− 0.0190	− 0.0072	0.0105
	(0.0189)	(0.0157)	(0.0074)	(0.0219)	(0.0164)	(0.0072)
1999	0.0034	− 0.0104	0.0078	− 0.0188	− 0.0189	0.0074
	(0.019)	(0.0159)	(0.0071)	(0.0238)	(0.0185)	(0.0076)
Number of cells	3804	3804	3804	3119	3119	3119

Effect of the program on wages, residual wages and formal employment among individuals born before 1962. Coefficients of interactions of the intensity of the program and survey year dummies.
1. The program intensity is the number of INPRES schools built between 1974 and 1978, divided by the number of children in 1971. 2. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions. 3. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell. 4. The standard errors are corrected for auto-correlation within kabupaten.

school graduates in the sample aged 20–60 (a reasonable approximation of the average education in the labor market); second, the fraction of primary school graduates among the 20–40 sample. The INPRES program directly affected the latter (since the older affected people were 37 in 1999). The former was affected as a consequence: Focusing on the 20–

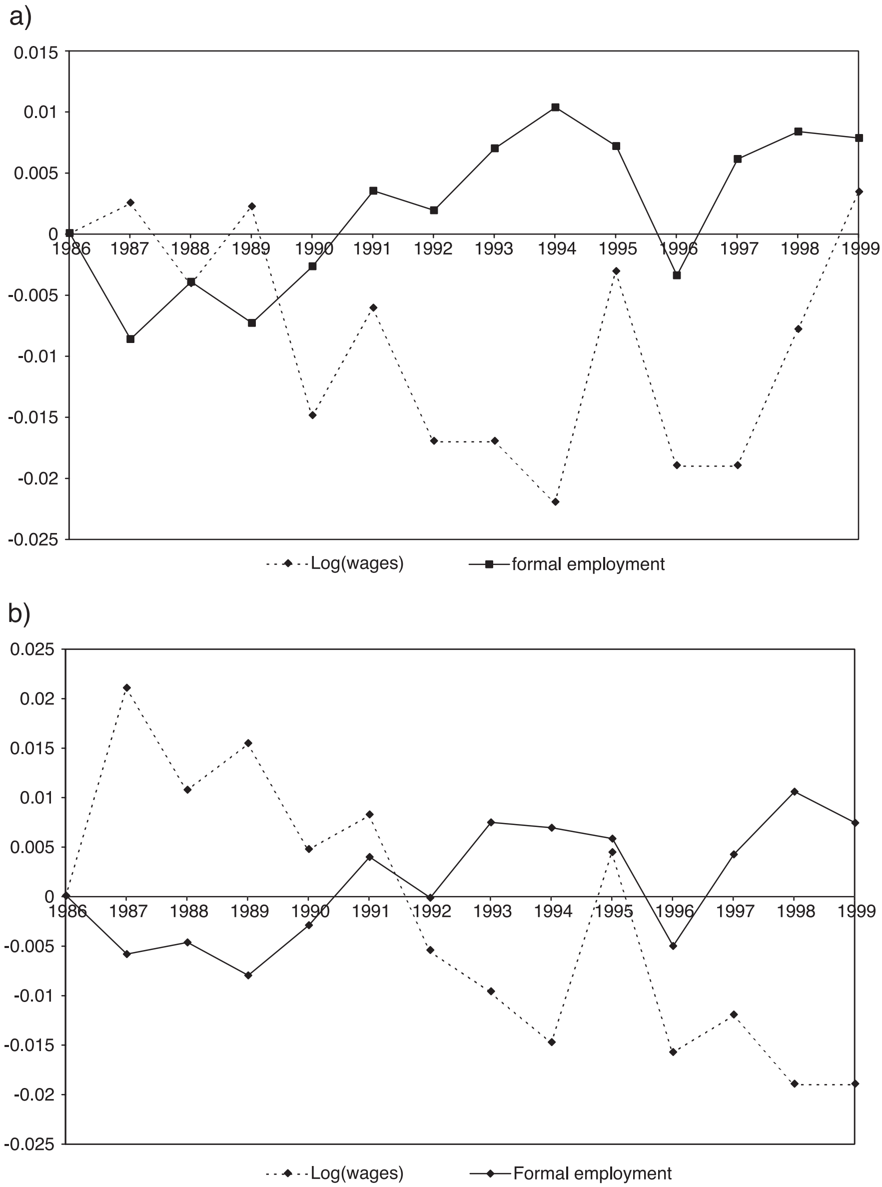

Fig. 4. (a) Coefficients of the interactions of program intensity and survey year dummies. Dependent variables: log(wage) and formal sector employment (individuals born before 1962 and aged less than 60). Sample: urban and rural regions. (b) Coefficients of the interactions of program intensity and survey year dummies. Dependent variables: average log(wage) and average formal sector employment among individuals born before 1962 and aged less than 60. Sample: rural regions.

Table 5
OLS estimates of the impact of average education on individual wages

	Independent variable: % of primary school graduates in the 20–40 sample		Independent variable: % of primary school graduates in the 20–60 sample	
	Sample: rural and urban areas	Sample: rural areas only	Sample: rural and urban areas	Sample: rural areas only
	(1)	(2)	(3)	(4)
Panel A: OLS, without region fixed effects (1986–1999)				
Log (wage)	0.852 (0.035)	0.754 (0.041)	0.881 (0.030)	0.875 (0.037)
Log (wage) residual	0.195 (0.024)	0.226 (0.028)	0.228 (0.021)	0.312 (0.027)
Skill premium	− 0.458 (0.041)	− 0.531 (0.047)	− 0.423 (0.038)	− 0.533 (0.045)
Formal employment	0.431 (0.016)	0.162 (0.015)	0.459 (0.015)	0.170 (0.015)
Formal employment among educated workers	0.095 (0.015)	− 0.036 (0.015)	0.108 (0.013)	− 0.035 (0.015)
Formal employment among uneducated workers	0.035 (0.011)	0.0046 (0.011)	0.045 (0.011)	0.0088 (0.011)
Panel B: OLS, with region fixed effects (1986–1999)				
Log (wage)	0.409 (0.041)	0.436 (0.046)	0.537 (0.041)	0.594 (0.047)
Log (wage) residual	0.040 (0.032)	0.060 (0.035)	0.086 (0.032)	0.121 (0.037)
Skill premium	− 0.082 (0.073)	− 0.092 (0.080)	− 0.094 (0.077)	− 0.106 (0.085)
Formal employment	0.160 (0.016)	0.134 (0.017)	0.228 (0.016)	0.201 (0.018)
Formal employment among educated workers	− 0.0019 (0.020)	0.0079 (0.022)	0.017 (0.021)	0.031 (0.022)
Formal employment among uneducated workers	0.012 (0.016)	0.015 (0.017)	0.031 (0.018)	0.033 (0.018)

Men aged 20–60 and born before 1962.
1. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions.
2. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell.

40 variable puts more accent on the source of identification.[11] The results presented here focus on the share of primary school graduates among males. Using instead the share of primary school graduates among males and females combined leads to almost identical results.

Table 5 presents OLS estimates for Eq. (7), where the dependent variable is the average wage and the average of the residual wage (after controlling for individual education and age). The first panel does not include district fixed effects, while the second one does. As expected, results obtained from specifications that do not control for individual education are bigger. They lump together the "social" and the "private" returns. We should, therefore, focus on the coefficients of average education in the wage residual equation. The difference

[11] In addition, the first stage is stronger for this variable, which minimizes the problems that can arise from using weakly correlated instruments. One would expect the results with the 20–60 average to be a scaled up version of the results obtained with the 20–40 average.

between the first and the second panel illustrates the remark made in Section 4.1: the OLS estimates are much bigger than the corresponding fixed effects estimates, which suggests that they are very strongly upward biased: educational attainments are higher in regions where wages are higher, but this is as likely to come from a relationship running from income to education as from the opposite relationship.[12] The OLS results are all positive and significant, while the OLS results with fixed effect are positive, but significant only in the specification that has the proportion of primary school graduates among those aged 20–60 as the dependent variable. These point estimates suggest small positive effects: an increase of 10 percentage points in the share of primary school graduates among those aged 20–60 is associated with an increase of 0.8% in the wages, after controlling for individual education. These coefficients are less than one tenth of those estimated by Moretti (1999) for the impact of the share of college graduates in the US.

Table 6 presents the instrumental variables results. The results are presented for the entire sample, and for a sample that excludes the years 1998 and 1999, since the crisis hit different regions differently. The first line of each panel presents the results on wages. In the full sample, the estimates become more negative when the crisis years are removed, while the estimates are not affected in rural areas. The second line presents the results using the residual wages as the dependent variable. None of the estimates is significant. The estimates obtained using the residual wage or the actual wage as the dependent variable are very similar, which is reassuring: since the INPRES instruments affect only average education, and not individual education, controlling for individual education should not affect the estimate, which is what we find here. The estimates using the residual wage are somewhat more precise. They suggest that an increase of 10 percentage points in the share of primary school graduates among the 20–60-year-old led to a decrease of 3.8% in wages in the full sample and to a decrease of 9.9% in the sample of rural areas. Without using the last 2 years of data, the coefficients are, respectively, -4.4% and -9%. The negative coefficients are significant (at the 10% level) in the rural sample only.

Focusing on the share of primary school graduates among the 20–40 year-olds (for which the first stage has more explanatory power), the story is the same: an increase of 10 percentage points in the share of primary school graduates leads to a decrease of 2.9% in the wage of the old in the full sample and to a decrease of 6.3% in the rural sample.

4.3. Skill premium

The third line in panels A and B of Table 5 presents the results of estimating by OLS (with and without district dummies) an equation similar to Eq. (7), but where the dependent variable is the difference between the average wages of educated and uneducated workers. Without district dummies, the estimate is negative, large (about -0.45) and significant. With district dummies, the estimates are negative, but much

[12] There are many reasons, besides those emphasized here, which would lead OLS coefficient to be biased upwards. First, there may be a wealth effect in education: Glewwe and Jacoby (2000) find an important wealth effect in Vietnam. Second, with economic growth, expected returns to education improve and this may lead to a higher demand for education (see Foster and Rosenzweig, 1996 for microeconomic evidence of the Indian green revolution and Bils and Klenow, 2000 for a re-interpretation of the cross-country evidence along these lines).

Table 6
2SLS estimates of the impact of average education on individual wages

	Independent variable: % of primary school graduates in the 20–40 sample		Independent variable: % of primary school graduates in the 20–60 sample	
	Sample: rural and urban areas	Sample: rural areas only	Sample: rural and urban areas	Sample: rural areas only
	(1)	(2)	(3)	(4)
Panel A: years 1986–1999				
Log (wage)	– 0.204 (0.443)	– 0.834 (0.701)	– 0.208 (0.615)	– 0.871 (0.837)
Log (wage) residual	– 0.292 (0.355)	– 0.633 (0.431)	– 0.379 (0.512)	– 0.994 (0.556)
Skill premium	– 0.434 (0.916)	– 0.982 (1.408)	– 0.596 (1.197)	– 0.636 (1.645)
Formal employment	0.441 (0.159)	0.454 (0.203)	0.661 (0.238)	0.745 (0.352)
Formal employment among educated workers	0.432 (0.197)	0.501 (0.259)	0.543 (0.264)	0.713 (0.406)
Formal employment among uneducated workers	0.379 (0.203)	0.409 (0.232)	0.510 (0.354)	0.318 (0.318)
Panel B: years 1986–1997				
Log (wage)	– 0.358 (0.493)	– 0.710 (0.821)	– 0.451 (0.716)	– 0.480 (0.801)
Log (wage) residual	– 0.330 (0.412)	– 0.588 (0.529)	– 0.437 (0.618)	– 0.902 (0.602)
Skill premium	– 0.225 (1.033)	– 0.635 (1.461)	– 0.291 (1.488)	0.536 (1.576)
Formal employment	0.463 (0.183)	0.442 (0.233)	0.716 (0.282)	0.694 (0.379)
Formal employment among educated workers	0.428 (0.229)	0.473 (0.301)	0.530 (0.317)	0.622 (0.479)
Formal employment among uneducated workers	0.478 (0.249)	0.449 (0.277)	0.624 (0.415)	0.263 (0.319)

Men aged 20–60 and born before 1962.
1. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions. 2. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell. 3. The instruments are interactions between survey year dummies and the program intensity. 4. The standard errors are corrected for auto-correlation within kabupaten.

smaller (about – 0.09) and insignificant. OLS seems again to be biased upwards (in absolute value).

The third line in panels A and B of Table 6 presents the instrumental variables estimates of the same equation. The IV estimates of the effect of the share of primary school graduates on the primary education premium are either negative or positive and always insignificant. The education premium does not seem to have been affected by the increase in the number of primary school graduates. This suggests that, in at least one sector of the economy, educated and uneducated workers are close substitutes.

4.4. Formal labor force participation

The fourth line in panels A and B of Table 6 presents the instrumental variables results for formal labor force participation (corresponding OLS results are presented in Table 5). The dependent variable is the fraction of people who work for a wage. The 2SLS estimates suggest that there is a strong positive effect of the fraction of primary school graduates on the probability that someone works for a wage. In the rural and urban sample combined, a 10% increase in the proportion of primary school graduates among the 20–40 year-olds leads to a 4.5% increase in the probability of working for a wage. A 10% increase in the proportion of primary school graduates among the 20–60 year-olds leads to a 6.6–7.5% increase. The coefficients are significant in all of the specifications and are very similar across specifications.

4.5. Sample selection

There are two possible sources of sample selection. First, there might be selective migration. Second, since the program affected the proportion of people for whom we observe wages, the probability of selection in the sample is affected by the instruments. In particular, one can imagine a situation where the program pushed the "marginal" self-employed into the formal labor force and these marginal employees receive a lower wage. Moreover, new entrants into the labor force have less experience, which should lower their wages.

Fig. 2 (and columns 3 and 6 in Table 4) suggests that the average education of individuals born before 1962 in the sample was not affected by the program: along this observable dimension, the sample remains comparable over time. Likewise, when I regress the education level of individuals who were born before 1962 *and who earn a wage* on the interactions between the program intensity and the survey year dummies, there is no distinct pattern in this regression (the F statistic of the interactions is 1.03), which indicates that, along observable characteristics at least, the composition of the formal labor force did not change as a result of the program.

However, there may have been selective migration along unobserved dimensions (if low productivity old people are attracted to the program regions for example, or if high productivity old people leave the region), which will cause a downward bias in the effect of the program on wages. The SAKERNAS data does not indicate whether an individual is a migrant, and I do not have any income measure for individuals who are not working for a wage. We thus need other sources of information to shed light on this issue.

First, the SUPAS data set (the 1995 Intercensal Survey of Indonesia described earlier) has the individual's region of birth as well as his region of residence. To investigate whether there are differences in productivity between migrants and non-migrants that are correlated with the INPRES program, I form for each district the difference between the logarithm of the hourly wage of the migrants and that of the non-migrants (among those born before 1962 currently residing in the district). Column 1 of Table 7 presents a regression of this variable on the number of INPRES schools built per capita in the region. The coefficient on the number of schools is actually positive (but insignificant), which suggests that there is no downward sample selection bias. In column 2, I construct the difference between the wage of those who migrated out of their region of birth and those who stayed. This difference is unrelated to the level of the program. There is thus no evidence that selective migration is likely to bias the results downward.

Second, I use the SUSENAS data (a nationally representative survey of about 50,000 households, which has an income and a consumption supplement once every 5 years). I

Table 7
Additional evidence on sample selection

	Log (wage migrant in) − log(wage stayer)	Log (wage migrant out) − log(wage stayer)	Formal sector premium 1993 − formal sector premium 1987
	(1)	(2)	(3)
Program intensity	0.0229 (0.0159)	0.011 (0.0164)	− 0.00207 (0.0160)
Number of cells	285	288	272

1. The program intensity is the number of INPRES schools built between 1974 and 1978, divided by the number of children in 1971. 2. The formal sector premium is defined as: log(income of wage earner/income of non-wage earners). 3. The enrollment rate and the number of children in 1971 are introduced as controls in the regressions. 4. The data for columns (1) and (2) is the SUPAS 1995. 5. The data for column (3) is the SAKERNAS, 1987 and 1993.

use the income modules from the SUSENAS from 1987 and 1993 to compute the ratio of the household income of self-employed to the household income of wage earners. The ratio is very stable between 1987 and 1993: self-employed earn 17.85% less than employed in 1987 and 18.5% less in 1993. I then regress the difference in the log of this ratio (Table 7, column 3) on the level of the program and found no relationship (the coefficient of the INPRES program is 0.0021, with a *t*-statistic of 0.130).

On balance, it appears that the relative wage loss of the old generation in regions where the program increased the supply of primary school graduates cannot be attributed to a composition effect.

We can summarize the results from this section as follows. An increase in the share of the educated workers leads to:

- A decline in the wage of older workers, whose level of education did not change. The point estimate is large (as large as the skill premium itself, or even larger in rural areas), although it is significant only in some specifications.
- No change in the skill premium among older workers.
- An increase in the share of the labor force employed in the formal sector, among the old. The point estimates are large (a 10% increase in the share of educated workers leads to an increase of at least 4% in the share of old workers employed in the formal sector) and significant.
- No change in the difference between formal and informal sector earnings.

In Section 5, I build a model which can explain these effects and serve as a framework to interpret their magnitude.

5. Model and interpretation

What do these results tell us about the response of the economy to an increase in the education of the labor force? In this section, I use a simple two-sector model as a framework to interpret these effects and their magnitude. In Section 5.1, I set up the model and study the effect of education on wages and the allocation of labor across sectors, taking capital as given. In Section 5.2, I compare the predictions of the model to the data. I specify two polar cases for the accumulation of capital. In the first case, there is no adjustment cost. In the second case, physical capital accumulation does not adjust at all to changes in the rate of human capital accumulation. In this simple model, both assumptions have the same implication for formal labor force participation, but very different implications for wage rates.

Below, I provide some justification for the specific assumptions of the model.

The fact that the skill premium was not affected suggests that, at least in one sector, educated and uneducated workers are very strong substitutes. Since in the formal sector, workers are combined, while they are self employed in the informal sector (we can think about this sector as small scale agriculture), we will take as a starting point that educated and uneducated workers are perfect substitutes in the informal sector. Suppose that the informal sector combines land and human capital, and the formal sector combines physical and human capital. The informal sector is characterized by a downward sloping demand for effective units of labor, even in the long run (because land is a fixed factor). In the formal sector, the slope of the labor demand depends on how capital adjusts to changes in the composition of the labor force. If capital does not adjust to an increase in effective labor supply, labor demand will be downward sloping. It would be flat if this increase led to an offsetting increase in the supply of capital, or even slope upwards if there were technological externalities, or if faster capital accumulation more than offset human capital accumulation (as in Acemoglu, 1996).

Consider a simple competitive model, where wages are equalized in the formal and informal sector (for a given level of skill). Fig. 5 illustrates the effect of the increase in

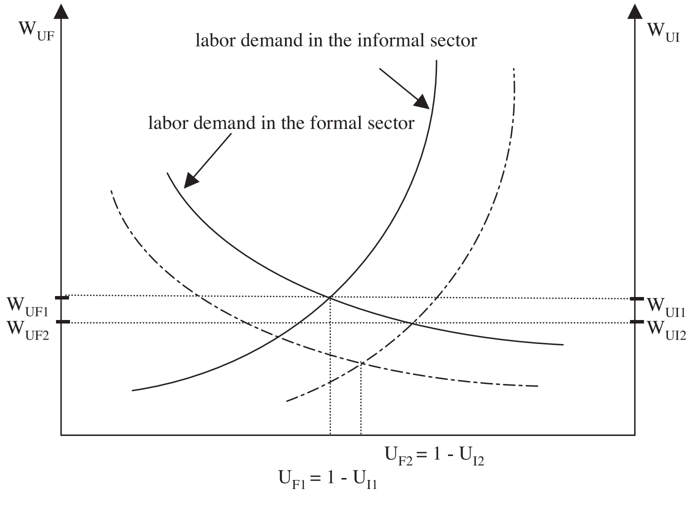

Fig. 5. Effect of an increase in the share of educated workers on wages and formal employment.

educated workers induced by the INPRES reform on uneducated old workers. The total number of old uneducated workers and their allocation between the formal and the informal sector are depicted on the *X*-axis. The left *Y*-axis presents the wage in the formal sector and the right *Y*-axis presents the wage in the informal sector. The increase in the number of educated workers leads to a shift in the labor supply expressed in efficiency units, which is akin to a shift to the right of the labor demand in the informal sector, and a shift to the left in the formal sector. If the labor demand in the formal sector is downward slopping, the effect on wages is unambiguously negative. The effect of an increase in the number of educated workers on the allocation of labor between formal and informal sector depends on the ratio of the elasticity of substitution between capital and labor in the formal sector and that between land and labor in the informal sector. The observation that the increase in the proportion of educated workers led to a shift from the informal to the formal sector indicates that the elasticity of substitution between labor and capital is bigger than the elasticity of substitution between labor and land. We will capture this with the simplifying assumption that the informal sector uses a Leontieff technology.

5.1. Model

This section builds a specific model along the lines described above. We start by describing the allocation of labor and the determination of wages, taking capital as given. In Section 5.2, I will compare the predictions of this model to the data, using two polar cases for the adjustment of capital to the increase of the share of educated workers: no adjustment cost vs. no adjustment whatsoever (infinite adjustment costs).

There is a mass 1 of workers in the economy, a fraction S of which are educated. There are two sectors, an informal (small-scale agriculture) and a formal sector (industry).

Suppose that the informal sector is characterized by a Leontieff production function in terms of efficiency units:

$$Y_{\mathrm{I}} = \min(T, H_{\mathrm{I}}),$$

where H_{I} is the number of efficiency units employed in the informal sector and T is the (normalized) amount of available land.

In addition, assume that educated and uneducated workers are perfect substitutes in the informal sector:

$$H_{\mathrm{I}} = U_{\mathrm{I}} + hE_{\mathrm{I}},$$

where h is the relative efficiency of educated workers. Land is distributed optimally between educated and uneducated workers and both earn their marginal productivity, so that the ratio of the educated to the uneducated wage is h.

The production function in the formal sector is Cobb-Douglas in human and physical capital, with constant returns to scale:

$$Y_{\mathrm{F}} = A_{\mathrm{F}} K^{\alpha} H_{\mathrm{F}}^{1-\alpha}$$

Human capital is a Cobb-Douglas aggregate of educated and uneducated workers:

$$H_{\mathrm{F}} = E_{\mathrm{F}}^{\beta} U_{\mathrm{F}}^{1-\beta}$$

Wages in the formal sector are given by the marginal productivity of each factor:

$$w_{\mathrm{EF}} = (1-\alpha) A_{\mathrm{F}} K^{\alpha} H_{\mathrm{F}}^{-\alpha} \beta E_{\mathrm{F}}^{\beta-1} U_{\mathrm{F}}^{1-\beta}$$

and

$$w_{\mathrm{UF}} = (1-\alpha) A_{\mathrm{F}} K^{\alpha} H_{\mathrm{F}}^{-\alpha} (1-\beta) E_{\mathrm{F}}^{\beta} U_{\mathrm{F}}^{-\beta}.$$

In equilibrium, the ratio $w_{\mathrm{EF}}/w_{\mathrm{UF}}$ must be equal to h. Therefore, the ratio of uneducated to educated labor in the formal sector must satisfy the relationship:

$$\frac{U_{\mathrm{F}}}{E_{\mathrm{F}}} = h \frac{1-\beta}{\beta} \tag{10}$$

Together with the relationships $E_{\mathrm{F}} + E_{\mathrm{I}} = S$, $U_{\mathrm{F}} + U_{\mathrm{I}} = 1 - S$ and $h_{\mathrm{EI}} + U_{\mathrm{I}} = T$, this determines the employment of each category of workers in each sector:

$$E_{\mathrm{F}} = \frac{\beta}{h}[1 + S(h-1) - T] \tag{11}$$

$$E_{\mathrm{I}} = \frac{\beta}{h}[T - 1 - S(h-1) + \frac{h}{\beta} S] \tag{12}$$

$$U_{\mathrm{F}} = (1-\beta)[1 + S(h-1) - T] \tag{13}$$

$$U_{\mathrm{I}} = T - \beta[T - 1 - S(h-1) + \frac{h}{\beta} S] \tag{14}$$

Production in the formal sector can be expressed as a function of E_{F}:

$$Y_{\mathrm{F}} = A_{\mathrm{F}} \left(\left(h \frac{1-\beta}{\beta} \right)^{1-\beta} \right)^{1-\alpha} K^{\alpha} E_{\mathrm{F}}^{1-\alpha}$$

The wage of an educated worker is therefore given by:

$$\ln(w_{\mathrm{E}}) = C + \alpha \ln(K) - \alpha \ln(E_{\mathrm{F}}), \tag{15}$$

$$\ln(w_{\mathrm{E}}) = C' + \alpha \ln(K) - \alpha \ln(1 + S(h-1) - T), \tag{16}$$

where C and C' are functions of α, β, A_{F} and h.

The same equations (with different C and C') describe the wage of an uneducated worker.

5.2. Calibration

5.2.1. Effect of education on formal employment

The ratio of educated to uneducated workers in the formal sector, divided by the skill premium, should be constant, according to our model, and allows us to calculate β. In fact, the ratio of the number of uneducated workers to the number of educated workers is decreasing over time (Table 1, column 5), while the skill premium is stable (Table 1, column 6). The value of β obtained from Eq. (10) is therefore increasing. This may reflect factors not taken into account, such as the adoption of more skill-complementary technologies. The average value of β is 0.62.

Eqs. (11) and (13) indicate how employment of educated and uneducated workers in the formal sector responds to a change in S. Suppose that all the newly educated workers are "born" into the informal sector, and that young and old workers shift from the informal to the formal sector to restore the correct proportion between E_{F} and U_{F}. When S increases by 1%, the probability that an educated worker born before 1962 will be employed in the formal sector increases by $\beta/h((h-1)/S_{\mathrm{o}})$, and the probability that an uneducated worker born before 1962 will be employed in the informal sector increases by $(1-\beta)(h-1)/1-S_{\mathrm{o}}$, where S_{o} is the share of educated workers among the old.

Values of the parameters β, h and S_{o} are shown in Tables 1 and 3. The average value of $h-1$ (the skill premium) is close to 0.40. The average value of S_{o} is around 0.65. The model would therefore predict that an increase in S of one percentage point would make the educated and uneducated born before 1962, respectively, 0.29% and 0.41% more likely to work in the formal sector.

The coefficients we have estimated suggest that an increase in S of one percentage point leads to an increase in the probability of working for a wage of 0.43 (in the combined sample) and 0.71 (in the rural sample) for the educated workers. For the uneducated workers, the estimated effects are, respectively, 0.51 (in the combined sample) and 0.31 (in the rural sample). Therefore, the model does a reasonable job in predicting the shift from the informal to the formal sector for the uneducated, and under-predicts the shift of the educated from the formal to the informal sector, especially in the rural sample.

5.2.2. Effect of an increase in average education on wage rates

The effect of the increase in average education on the wages (at a given level of human capital) depends on the extent to which physical capital adjusts to the increase in human capital. We contrast the implications of two polar cases. In the first case, capital adjusts costlessly to the modification induced by INPRES: I first consider the situation where each district is a closed economy with local accumulation, and then a situation where capital is freely mobile across regions. Both models deliver the results that wage growth across regions should not be differentially affected. In the second case, there is no adjustment at all of the physical capital stock in response to the INPRES program, over the period under consideration.

- Costless capital stock adjustment, local capital accumulation.

To study the consequences of the INPRES model in the case of costless adjustment, we need to specify how human capital and physical capital are accumulated. Following Barro and i Martin (1995) (chapter 5), consider a one-sector endogenous growth model, with human and physical capital. Output is divided between consumption and investment in the two forms of 27 capital.

$$Y = A_F K^{\alpha} H_F^{(1-\alpha)} + T = C + I_K + I_H,$$

where I_K and I_H are the gross investment in physical and human capital, respectively. The changes in the capital stocks are given by:

$$\dot{K} = I_K - \delta K,$$

and

$$\dot{H}_F = \eta I_H(t - L) - \delta H_F.$$

Note two differences from the set up in Barro and i Martin (1995). First, there is a lag of duration L between investment in human capital and the date at which it will become active: this reflects the fact that children who go to school today will enter the job market only after a lag. Second, η represents the ratio at which each unit of investment is transformed into a unit of human capital.[13] Note that H could not accumulate indefinitely if the only way to increase H were to increase the proportion of primary school graduates. We should think of H as a Cobb-Douglas aggregate of all levels of education: I model the beginning of the process of development, where only two levels of education have been achieved.

The solution to this problem is obtained by setting up the Hamiltonian expression for the household choice between consumption and investment in human or physical capital, taking first order conditions with respect to C, I_H, I_K, K and H_F. The steady-state growth rate is equal to:

$$\gamma = \frac{1}{\theta}\left[A_{F\alpha}\left(\frac{K}{H_F}\right)^{\alpha-1} - \delta - \rho\right]$$

The steady-state value of K/H_F is pinned down by the condition of equality between net returns to investment in physical and human capital. It is the solution to the following implicit equation:

$$\eta e^{L\left[\delta - \alpha A_F\left(\frac{K}{H_F}\right)^{\alpha-1}\right]}\left(\frac{K}{H_F}\right)^{\alpha} = \frac{\alpha}{1-\alpha}. \tag{17}$$

Denote k^* the value of K/H_F given by this equation. The steady-state growth rate is given by:

$$\gamma^* = \frac{1}{\theta}\left[A_F \alpha k^{*(\alpha-1)} - \delta - \rho\right] \tag{18}$$

We can model the effect of the INPRES program as an increase in η, the rate at which investment is transformed into human capital accumulation: each new school

[13] In the dual economy model, η reflects both the cost of producing educated workers and the rate at which the educated workers are participating in the formal labor market.

makes it cheaper to accumulate human capital. Suppose that the INPRES program will last long enough that we can think about it as a permanent program (a permanent increase in η).[14] In the new steady state, the ratio k^* will be lower than in the old steady state, to maintain the equality between net returns to investment in human capital and physical capital (see Eq. (17)). The program should therefore cause a drop in the *levels* of wages (holding human capital constant). After the transition, consumption, human capital and physical capital will grow at the same new steady-state rate (given by Eq. (18)). In the new steady state, growth in human capital will be exactly matched by growth in physical capital, and wage growth will therefore be unaffected by growth in human capital. To see this, differentiate Eq. (15) with respect to time, and note that $\dot{K}/K = \dot{H}_F/H_F$, so that the two terms cancel.

The transition to the new steady state involves indeed a decline of K/H_F, and therefore a decline in wages. In this model, however, the adjustment in the ratio K/H_F should have taken place right after the program was initiated, and well *before* the rate of growth of H actually starts to accelerate. To see this, imagine that the economy were initially at a steady state, and consider the household decision the first day the schools are built: human capital investment has suddenly become cheaper, while physical capital investment has the same price and the same returns as before (since the human capital available today keeps growing at the old steady-state growth rate). Net returns to investing in human capital are therefore higher, all the investment takes the form of human capital, and the ratio K/H_F declines rapidly until it is low enough that investing in physical capital become profitable again. Simulations of the transitional dynamics of this model show that the ratio K/H_F gets very close to its steady-state value very fast after the program is initiated (in 1 year if there if households are allowed to transform physical capital into human capital, and in less than 3 years otherwise).[15] Afterwards, the ratio K/H_F stay almost constant.[16]

This very fast adjustment rate may seem unrealistic, but this is because the assumption of costless capital adjustment is not realistic either, even though it underlies, implicitly or explicitly, the effort to identify a human capital externality (in which case the share of educated workers could have a positive effect on wages). The important point is to note that, in this model, the share of educated workers should not affect wages from 1986 to 1999. The data do not seem to support this full adjustment model for Indonesia; however, during the entire period, wages grow more slowly in regions were human capital grows faster (see Fig. 4B).

[14] For our purpose, it would not be easy to distinguish a permanent temporary program, since the rate of capital accumulation was more rapid in the regions that received more schools for the entire period under consideration.

[15] For example, if γ_1 was initially 3.5% a year, assuming a value of 0.05 for δ, K/H_F would have declined by 8% every year. With reasonable parameter values for α, δ, ρ and θ ($\alpha = 0.3$, $\delta = 0.05$, $\rho = 0.02$, $\theta = 3$), and assuming that the effect of each school on the rate of growth of S we observe in the data is the steady-state effect (i.e. each school causes an increase of 0.5% in the growth rate), K/H_F should fall by about 10% for each school built (per 1000 children), which would have been achieved in 1.25 years. On average, two schools were built, so in 2.5 years, K/H_F should have fallen enough in the average region to reach the new steady-state value.

[16] In the simulations, there are small oscillations around the steady state values and they are progressively dampening.

• Costless capital stock adjustment, national capital accumulation and freely mobile capital across regions.

If capital is freely mobile across regions and there are no adjustment costs, the capital stock simply adjusts to the increase in human capital to equalize the returns to physical capital across regions. This determines the physical capital/human capital ratio k, which in turns determines the wage: thus (absent positive externalities), the evolution of wages should not be different in regions that received more schools (the level of wages may of course be related to the program). Once again, the data do not seem to support this model.

There were more than 10 years between the start of the program and the entry into the labor force of the generations that were exposed to it. Thus, even if capital accumulates with a gap (“time to build”), the conclusions of such a model would be unchanged.

Thus, it seems that, to account for the pattern present of the data, it is necessary to introduce frictions in the adjustment of the capital stock. We consider an extreme case of frictions.

• No adjustment of the capital stock in response to the increase in education.

The other extreme is a case where there is absolutely no response of physical capital accumulation to the increase in the stock of educated workers. The stock of capital thus evolves exogenously. In this case, if the INPRES program is a valid instrument (that is, if it is not correlated with the exogenous rate of capital accumulation), the stock of capital

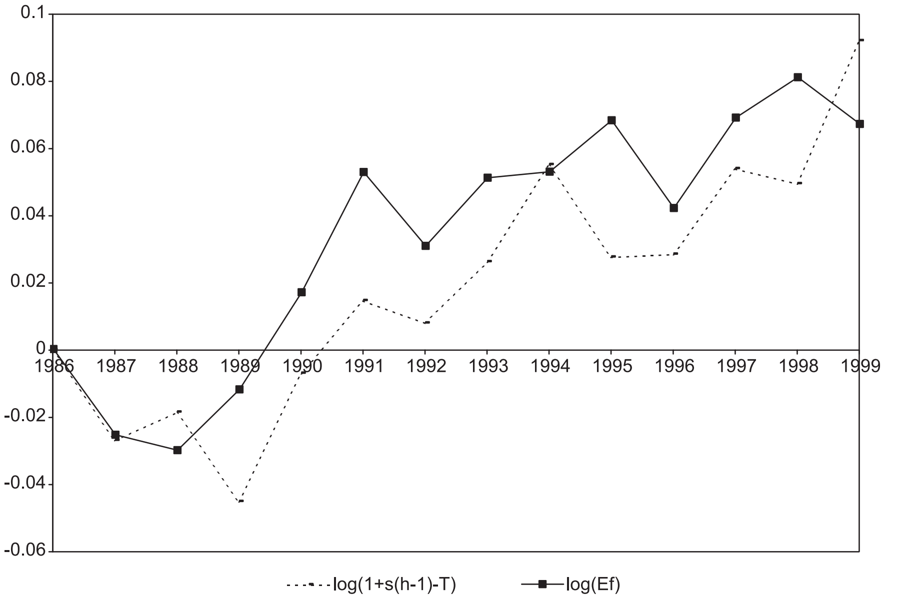

Fig. 6. Coefficient of the interaction between program intensity and survey year dummies. Dependent variable: log(Ef) and $\log(1+S(h-1)-T)$.

Table A1
Alternative first stage: effect of the program on log(Ef) and log(1 + $S(h-1)-T$)

	Independent variable: log(Ef)		Independent variable: log(1 + $S(h-1)-T$)	
	Sample: urban and rural areas	Sample: rural areas only	Sample: urban and rural areas	Sample: rural areas only
	(1)	(2)	(1)	(2)
1986	omitted	omitted	omitted	omitted
1987	−0.0242 (0.0429)	−0.0344 (0.0521)	−0.0244 (0.0546)	−0.0320 (0.0663)
1988	−0.0279 (0.022)	−0.0502 (0.0259)	−0.0131 (0.0304)	−0.0248 (0.0361)
1989	−0.0109 (0.0285)	−0.0331 (0.0349)	−0.0409 (0.0319)	−0.0697 (0.0373)
1990	0.0185 (0.0313)	0.0033 (0.0345)	−0.0057 (0.0324)	−0.0135 (0.0387)
1991	0.0540 (0.0292)	0.0436 (0.0313)	0.0153 (0.03)	0.0058 (0.0321)
1992	0.0320 (0.0307)	0.0109 (0.0325)	0.0102 (0.0306)	−0.0176 (0.0338)
1993	0.0518 (0.0308)	0.0339 (0.0332)	0.0250 (0.0335)	0.0008 (0.0373)
1994	0.0534 (0.0345)	0.0238 (0.0392)	0.0548 (0.0354)	0.0220 (0.0404)
1995	0.0681 (0.0307)	0.0515 (0.0328)	0.0266 (0.0352)	0.0198 (0.0363)
1996	0.0459 (0.0344)	0.0231 (0.0343)	0.0313 (0.0363)	0.0060 (0.0401)
1997	0.0703 (0.0375)	0.0510 (0.0414)	0.0522 (0.0404)	0.0327 (0.0477)
1998	0.0807 (0.0351)	0.0708 (0.0371)	0.0417 (0.0517)	0.0137 (0.0615)
1999	0.0691 (0.0357)	0.0598 (0.0383)	0.0918 (0.0476)	0.1131 (0.063)
Number of cells	3808	3122	3336	2762
F-statistic	3.28	2.49	1.69	1.30

Sample: men aged 20–60.
1. The program intensity is the number of INPRES schools built between 1974 and 1978, divided by the number of children in 1971. 2. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions. 3. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell. 4. The *F*-statistic is for the hypothesis that the set of interactions is jointly insignificant. 5. See text for variables definition. 6. The standard errors are corrected for auto-correlation within kabupaten.

enters the error term in Eq. (15) and is uncorrelated with the level of the program. The best way to check whether our model fits the facts is to use $\ln(E_F)$ or $\ln(1+S(h-1)-T)$ as the dependent variable instead of S.[17] Fig. 6 shows that these variables are also influenced by the program. The first stage coefficients for these two variables are similar, which is reassuring (Table A1).[18] The figure also looks similar to Fig. 2. In Table 8, I present the results of estimating Eq. (15), using the same strategy outlined for estimating the effect of S on wages. In practice, I estimate Eq. (7), with $\ln(E_F)$ or $\ln(1+S(h-1)-T)$ as the endogenous regressor of interest.

The coefficients of a subset of the specifications estimated in Table 6 are presented in Table 8. For $\ln(E_F)$ (columns 1 and 2), the pattern is similar to what we found for S: the coefficients are negative, but insignificant. The point estimate is −0.07 in the urban

[17] $\ln(E_F)$ is the proportion of the labor force that is educated and works in the formal sector. $\text{Ln}(1+S(h-1)-T)$ is calculated using the district and year-specific skill premium. T is calculated using its definition: $T=H_I=U_I+h*E_I$.

[18] I estimate the same specification as in Eq. (1), but I use $\ln E_F$ and $\ln(1+S(h-1)-T)$, respectively, as the dependent variables. Full results are presented in Table A1.

Table 8
SLS estimates of the impact of human capital on wages

	Dependent variable: log(wage) for men 20–60, born before 1962			
	Independent variable: log(Ef) among sample aged 20–60		Independent variable: $\log(1+S(h-1)-T)$ among sample aged 20–60	
	Sample: rural and urban areas	Sample: rural areas only	Sample: rural and urban areas	Sample: rural areas only
	(1)	(2)	(3)	(4)
Panel A: years 1986–1999	– 0.073 (0.164)	– 0.219 (0.205)	– 0.365 (0.217)	– 0.341 (0.22)
Panel B: years 1986–1997	– 0.092 (0.165)	– 0.165 (0.208)	– 0.396 (0.213)	– 0.360 (0.275)

Log(Ef) and $\log(1+S(\mathrm{h}-1)-\mathrm{T})$ as measures of human capital.
1. Survey year dummies, region dummies, interactions between survey year dummies and the enrollment rate in 1971, and interactions between survey year dummies and the number of children are included in the regressions. 2. Regression run using kabupaten-year averages, weighted by the number of observations in each kabupaten-year cell. 3. The instruments are interactions between survey year dummies and program intensity. 4. The standard errors are corrected for auto-correlation within kabupaten.

sample and -0.22 in the rural sample. The latter estimate of -0.22 is close to -0.3, the conventional guess for α.

The results using $\ln(1+S(h-1)-T)$ as the independent variable are more encouraging. The point estimates are similar in the combined sample and in the rural sample. The estimates range between -0.34 and -0.40, and are all significant at the 10% level of confidence (after correcting the standard errors for auto-correlation over time within regions). These estimates (especially those obtained in the combined sample) are higher than what I obtained using $\ln(E_{\mathrm{F}})$ (they should in principle be the same). They are very close to (and not significantly different from) -0.3. These estimates are fairly precise and their order of magnitude is reasonable.

Note that the model predicts that the effect of the program on wages should be the same in the formal and in the informal sector. The SAKERNAS does not have data on income for those who do not work for a wage. However, the regression in column 3 of Table 7, discussed in Section 4.5, suggests that the ratio of the income of the wage earners to that of non-wage earners was not affected by the program: the difference between the value of this ratio in 1993 and in 1987 is not correlated at all with the intensity of the program.

This set of estimates therefore suggests that the accumulation of physical capital happens essentially independently from the accumulation of human capital. Even 25 years after the program was initiated, physical capital does not seem to have been accumulated to employ the new efficiency units of labor created by the program. This conclusion seems consistent with other evidence on the pattern of industrialization in Indonesia during the period. After 1984, the Indonesian Government cut all previously existing tax incentives to geographical dispersion and small businesses (Hill, 1996). Miguel et al. (2001) show that there was divergence in industrialization across regions between 1985 and 1995: industrialization was faster during this period in regions where its level was higher in 1985. Moreover, they show that change in the rate of industrialization seem uncorrelated with school construction (not only INPRES schools, but all primary and junior high schools) in the previous decade, as well as with other infrastructure variables (roads or electricity).

6. Conclusion

This paper argued that the INPRES program, a large school construction program undertaken by the Indonesian government in the 1970s, constitutes a good case study to empirically examine the impact of average primary schooling on the wages of older cohorts. This program modified the enrollment rates of the young generations, thus inducing a long-lasting change in the rate of human capital accumulation in the regions it affected most. We can study the impact of this shock on the supply of educated workers on an "old" generation that did not directly benefit from the program. It provides a natural solution to the identification problems inherent to any attempt to identify the effect of the average of a regressor while trying at the same time to control for it.

The instrumental variables estimates presented in this paper suggest that the effect of average education on individual wages (holding skill level constant) is negative: in places where average educational attainments grew faster because of the program, wages grew more slowly. This does not seem to be due to sample selection bias due to migration or increases in labor force participation. This is in sharp contrast to the OLS estimates, which are strongly positive, and to the fixed effect estimates, which are closer to zero but still positive. Such a strong bias in the OLS estimates suggests that the cross-country relationship between output per capita and education is likely to be affected by the same upward bias. The effects of average education on the participation in the formal labor market are, however, positive.

Both sets of estimates are shown to be consistent with a simple dual-economy model, where the number of efficiency units that can be productively employed in the informal sector is limited by the availability of a fixed factor (land, for example). The increase in the productivity of the labor force is entirely absorbed by the formal sector, which explains the increased participation. Wages decrease to the extent that physical capital does not fully adjust to the increase in human capital. I contrast three versions of this model. In a closed economy endogenous growth version with costless adjustment of the capital stock, the faster increase in human capital after 1986 should be matched with a corresponding increase in the stock of physical capital and wages should be unaffected. In an open economy model where capital moves freely, the capital–labor ratio adjusts to maintain equal returns to physical capital across regions, and the program should have no effect on wages. If capital accumulation is unresponsive to the increase in human capital, this predicts that the coefficient on the human capital variable should be the negative of the elasticity of production with respect to capital.

The extent to which wages fall in response to the increase in human capital available in the formal sector suggests that there was little or no reaction of the physical capital to the increase in educational opportunities in the regions where INPRES built more schools. The estimates I obtain are reasonable if this is the right model of the world.

What is absent in this paper (and left for future work) is an explanation of why the capital stock did not adjust. The program was publicly announced, and the increase in the stock of primary school graduates occurred gradually over 10 years after the program started. It is a puzzle that 25 years after the program was initiated, the labor demand looks like a "short run" labor demand curve. Future work (and, probably, more data) is needed to determine to what extent this is due to "myopic" behavior of investors, who fail to recognize the increase in the education level of the labor force, very large adjustment costs of the capital stock, to a financing constraint, or a combination of the three.

The results in this paper are important because, contrary to what is often assumed (on the basis of the experience of South-East Asian countries), acceleration in the rate of accumulation of human capital is not necessarily accompanied by economic growth. Several countries (in Africa, in particular) had very rapid expansion in education, but dismal economic growth (Kremer and Thomson, 1998). It is important to understand why this can be the case.

Models of credit constraints (Banerjee and Newman, 1993; Galor and Zeira, 1993; Aghion and Bolton, 1997) could be combined with models of costly adjustment of technology to study the effects of education on economic growth given the actual constraints faced by developing economies. The work of Caballero and Hammour (1998, 2000) comes closest to doing this. Their model combines costly adjustment with credit constraints but does not model growth. It cannot therefore be directly applied to the question of what happens when the growth rate of human capital increases. Once built, such a model could then be compared to actual evolutions, in exercises similar to Blanchard's (1997) analysis of the "medium run".

Acknowledgements

I thank participants at the conference "New Research on Education in Developing Countries" at the Center for Research on Economic Development and Policy Reform at Stanford University for comments, and particularly Hanan Jacoby for his discussion of the paper. I also thank two referees and the editor for very useful comments. I thank Lucia Breierova and Shawn Cole for excellent research assistance, Daron Acemoglu, Joshua Angrist, Abhijit Banerjee, Robert Barro, Ricardo Caballero, David Card, Michael Kremer, Emmanuel Saez and Jaume Ventura for very helpful discussions, and Guido Lorenzoni for his insights about transitional dynamics.

References

Acemoglu, D., 1996. A microfoundation for social increasing returns in human capital accumulation. Quarterly Journal of Economics 111 (3), 779–804.

Acemoglu, D., Angrist, J., 2000. How large are human capital externalities? Evidence from compulsory schooling laws. NBER Macroeconomics Annual, 9–59.

Aghion, P., Bolton, P., 1997. A trickle-down theory of growth and development with debt overhang. Review of Economic Studies 64 (2), 151–172.

Atkinson, T., Brandolini, A., 2001. Promise and pitfalls in the use of "secondary" data-sets: income inequality in the OECD countries as a case study. Journal of Economic Literature 39, 771–779.

Banerjee, A., Duflo, E., 2000. Inequality and growth: what can the data say? Working Paper, vol. 7793. National Bureau of Economic Research July.

Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. Journal of Political Economy 101 (2), 274–298.

Barro, R., i Martin, X.S., 1995. Economic Growth, McGraw-Hill, New York, London, Montreal.

Bertrand, M., Duflo, E., Mullainathan, S., 2001. How much should we trust differences in differences estimates. Mimeo, MIT.

Bils, M., Klenow, P., 2000. Does schooling cause growth or the other way around? American Economic Review 90 (5), 1160–1183.

Blanchard, O., 1997. The medium run. Brookings Papers on Economic Activity 2, 89–158.

Blanchard, O., Wolfers, J., 2000. The role of shocks and institutions in the rise of European unemployment: the aggregate evidence. Economic Journal, 1–33.

Bound, J., Kzedi, G., Turner, S., 2000. Trade in university training: cross state variation in the production and use of college educated labor. MIMEO, University of Michigan.

Caballero, R., Hammour, M., 1998. The macroeconomics of specificity. Journal of Political Economy 106 (4), 724–767.

Caballero, R., Hammour, M., 2000. Creative destruction and development: institutions, crises, and restructuringWorking Paper, vol. 7849. National Bureau of Economic Research. August.

Card, D., Lemieux, T., 2001. Can falling supply explain the rising return to college for younger men? A cohort-based analysis. Quarterly Journal of Economics 116 (2), 705–746.

Duflo, E., 2001. Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. American Economic Review 91 (4), 795–813.

Foster, A.D., Rosenzweig, M.R., 1996. Technical change and human-capital returns and investments: evidence from the green revolution. American Economic Review 86 (4), 931–953.

Frankenberg, E., Thomas, D., Beegle, K., 1999. The real costs of Indonesia's economic crisis: preliminary findings from the Indonesia family life surveys. Working Paper 99-04, RAND Labor and Population Program.

Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. Review of Macroeconomic Studies 60, 35–52.

Glewwe, P., Jacoby, H., 2000. Economic growth and the demand for education: is there a wealth effect? MIMEO, World Bank.

Harris, J., Todaro, M., 1970. Migration, unemployment and development: a two-sector analysis. American Economic Review 40, 126–142.

Heckman, J.J., Lochner, L., Taber, C., 1998. General equilibrium treatment effects: a study of tuition policy. Working Paper, vol. 6426. National Bureau of Economic Research.

Hill, H., 1996. The Indonesian Economy Since 1966: Southeast Asia's Emerging Giant Cambridge Univ. Press, Cambridge.

Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963–1987: supply and demand factors. Quarterly Journal of Economics 107 (1), 35–78.

Kremer, M., Thomson, J., 1998. Why isn't convergence instantaneous? Young workers, old workers, and gradual adjustment. Journal of Economic Growth 3, 5–28.

Krueger, A., Lindahl, M., 2001. Education for growth: why and for whom? Journal of Economic Literature 39 (4) 1101–1136.

Lucas, R., 1988. On the mechanics of economic development. Journal of Monetary Economics 22, 3–42.

Miguel, E., Gertler, P., Levine, D., 2001. Did industrialization destroy social capital in Indonesia. Mimeo, U.C. Berkeley.

Moretti, E., 1999. Estimating the social return to education: evidence from repeated cross-sectional and longitudinal data. Working Paper 22 (UC Berkeley).

Solow, R.M., 2000. Toward a macroeconomics of the medium run. Journal of Economic Perspectives 14 (1), 151–158.

Welch, F., 1979. Effects of cohort size on earnings: the baby boom babies' financial bust. Journal of Political Economy 87 (5), S65–S97 (part 2).