Mass Education or a Minority Well Educated Elite in the Process of Development: the Case of India*

Amparo Castelló-Climent† and Abhiroop Mukhopadhyay‡

† *University of Valencia* ‡ *Indian Statistical Institute*

October 2010

Abstract

This paper analyses whether in developing countries mass education is the key or a highly well educated elite should be more beneficial for growth. Using the Indian census data as a benchmark and enrollment rates of different levels of schooling we compute annual attainment levels for a panel of 16 Indian states from 1961 to 2001. Results show that one standard deviation increment in the share of population with tertiary education is 3 times more beneficial for growth than a one standard deviation increment in literacy. Using simulations we consider two alternate policies: one that doubles the increments to the literacy rates (relative to its baseline rate of increase) and another that doubles the annual increments to the share of adult population with tertiary education. We show that at the end of 35 years, the state following the latter policy has a per capita GDP 1.5 time more than the state that emphasizes the former.

JEL classification: I28, O11, O50

Key words: Distribution of education, attainment levels, economic growth, panel data

*Amparo Castelló gratefully acknowledges financial support from the Spanish Ministry of Education through ECO2008-02675 project and the Ramón y Cajal Programme. Contact: amparo.castello@uv.es, abhiroop@isid.ac.in.

1 Introduction

According to United Nations, in 2007 about 72 million children in the world did not have access to education.[1] This striking number highlights that in poorer countries a large mass of population work as unskilled labor in low productive activities. In such economies, governments face the dilemma of whether to focus on policies that extend education to those that are illiterates or in policies that increase the share of well educated workers, who could specialize in high-skill sectors with large productivity and large economic growth rates. In order to better understand whether in developing countries mass education is more growth enhancing than to have a minority well educated elite, this paper focuses on a particular developing economy and estimates the effect of different measures of human capital that capture the distribution of education as well as the influence of each level of schooling.

The conventional wisdom about the relevance of each level of schooling is that mass education is the key. A possible explanation for this believe is that the marginal return to schooling is found to be decreasing with the level of schooling (e.g. Psacharopoulos and Patrinos, 2004). Nevertheless, these studies do not take into account the fact that in many countries the majority of university graduates are employed in the public sector, in which the wages do not reflect its market value. Moreover, recent estimates with cross-country data have also challenged this traditional view by showing an estimated rate of return to an additional year of schooling being higher at the secondary and tertiary levels than at the primary level (e.g. Barro and Lee, 2010). On the other hand, private and social returns of education may differ as well. As noted by Pissarides (2000), educated laborers can be engaged in activities with high private returns but located in sectors that are not growth-enhacing.

The goal of this paper is to analyze the importance of the composition of human capital in the less developed countries, which are characterized by a large share of population with no education and, therefore, a trade-off between literacy and high skill education may arise. Among the developing countries, India stands out for a priority of its governments in extending primary schooling among the illiterates as well as in increasing the share

[1] The second United Nations Millennium Development Goal is to achieve universal primary education by 2015.

of population with tertiary education. High quality engineering and technology-oriented institutions of higher education have been the aim of all the Indian governments since its independence in 1947. The high mass of illiterates joint with a non negligible number of a highly educated elite makes India an important case study for how the shape of the distribution of education may affect the economic performance of an economy.

The case of India is also convenient since it is one of the few developing countries with good statistics on relevant variables. Data on real GDP and other determinants of growth for the main Indian states are available on a year basis for the period 1960-2000 (e.g. Besley and Burgess, 2000, 2002, 2004; Ghate and Wright, 2010). The advantage of using the cross-sectional and the temporal dimension of the data is that they can be used to estimate a panel data model that controls for state specific effects and, therefore, minimize any omitted variable bias in the analysis. On the other hand, the use of sub-national level data also has the advantage that the different levels of education are more comparable across states within a country than across economies.

Whereas data on income measures and other relevant variables are available for the states of India on a year basis, there are not data on attainments at different levels of schooling. We fill this gap by computing yearly data on educational attainments across the states of India. Specifically, we use the Indian Census as a benchmark, which contains decade information on the educational levels across the states, and estimate annual observations using a perpetual inventory method and data on enrollment rates for different education levels. We compute the share of population 15 years and above with no schooling, some primary, completed primary, completed middle, completed secondary and completed tertiary for 16 Indian States from 1961 to 2001. We also use these data to compute distributional measures, such as the Gini coefficient and the distribution of education by percentiles.

From a methodological point of view, this paper shows that in developing countries, measures commonly used in the literature such as the average years of education or the human capital Gini coefficient are not sufficient to assess the effect of the level of education on the economic growth rates, since they are determined to a large extent by the huge mass of people with no education. For instance, in the case of India the correlation between the average years of schooling and the share of illiterates is above 0.9. These measures, therefore, mainly pick up the influence of illiterates on the economic performance of the economies. Alternate specifications that include the share of illiterates, the average years of education among the literates and the Gini coefficient among literates also fail in extricating the effect. This is because these distributional statistics are driven by large proportions of people with low levels of education. Hence the average educational attainment is very collinear to these popularly used distributional statistics.

We first show that when the shares of educational attainment vary a lot in degrees of magnitude, using shares of attainment of each education level does a much better job in bringing out the effect of the distribution of education. Moreover, controlling for fixed effects and using instrumental variables our results show that the effect of lowering illiteracy on growth is much smaller than the effect of increasing the share of tertiary education attainment. In quantitative terms, a one standard deviation decrease in the share of illiteracy increases growth by 0.046 percentage points whereas a one standard deviation increase in the share of tertiary education increases growth by 0.141 percentage points. This is in contrast to micro-studies like Psacharapoulus and Patrinos (2004), who report the returns of education are usually higher for low levels of schooling and decline for higher levels of education. However, in line with our macro findings, Bosworth et al. (2007) also find that in India the returns of primary education are relatively lower than the average returns. This finding is consistent with a low quality educational system at the primary level which may lead to higher literacy but does not contribute to skill accumulation. In fact, teachers absenteeism and teacher negligence is common in many Indian schools (e.g. Kremer et al., 2005).

As an extension, we examine if the effects of greater illiteracy rate are different for states with different levels of tertiary education. Our results show that there is no differential impact and that greater illiteracy always lowers growth. Using these estimated results as a benchmark, we simulate the effect of alternative educational policies. Overall, results show that economies might be better off in the medium term (35 years) following a strategy that emphasizes tertiary education attainment. We suggest one such policy that increases enrollment in universities (that includes diplomas) of the existing share of 19 years olds with higher secondary education. We find that at the end of 35 years, such a policy would yield higher returns than one where the cohort of 6 year old is targeted to increase literacy rates (as is aimed at under universalization of primary education). The latter policy takes time to affect the labour force (when they turn 15) and if the tertiary rate completion of these cohorts are low, the economy is never able to accelerate. With one such set of policies, we find that the former policy targeting tertiary education leads to a per capita NSDP 1.5 times that achieved under a policy that focuses exclusively on literacy.

In another set of simulations, we investigate if the states that belong to the poorest quartile can catch up with the next richest quartile through these education policies in 35 years. Our results show that the catch up is possible if tertiary rate completion rises annually by 0.45 percentage points or if the literacy rates rise by 5 percentage points as soon as (and subsequent to) the first wave of more literate cohorts reach the age of 15. We however point out that while it is possible to grow if the literacy rate declines are substantial, such a policy would not have an effect on the economy for many years. Even with such a drastic decline of illiteracy, the catch up is only possible after 35 years. While an optimal policy would require push both on literacy and school/university completion, often it is not possible to target both. Our paper points out that while lowering illiteracy is good, enough focus should also be given on completion and there should be attempts to endow the literate population with skills and education that only university and graduate diplomas institutes can give.

So far the traditional literature that empirically investigates the influence of human capital on economic growth has not emphasized the role of the composition of human capital. Instead, the most common approach has been the use of the average years of schooling of the adult population as a proxy of the stock of human capital (Benhabib and Spiegel, 1994; Bils and Klenow, 2000; de la Fuente and Domenech, 2006; Cohen and Soto, 2007). However, this paper shows that an aggregate measure of education is not sufficient to asses the effect of education on growth in countries characterized by a high number of illiterates. We show that an increase in the average years of education may be the result of an increase in the share of individuals with primary education, secondary schooling or an increase in the share of population with a university degree, each of them with a different effect on the growth rates.

One of the few attempts to analyze the role of the composition of human capital is Vandenbussche, Aghion and Meghir (2006), which focuses on the relevance of tertiary education in innovation activities in a sample of OECD countries that are close to the technological frontier. Our results show that even in the less developed countries, tertiary education may be crucial in shaping the economic performance of a country. The channels through which human capital influences growth are not analyzed in the paper. Nevertheless, as noted by Kocchar et al. (2006), in India both manufactures and services are relatively concentrated in skill-intensive output. Thus, a possible explanation for the relevant influence of tertiary education is that highly educated workers have been employed in sectors with high productivity.

The organization of the paper is as follows. In the next section we present the data used and the methodology to estimate annual educational attainments across the Indian states. In Section 3 we discuss some specifications and display the econometric models to be estimated. Section 4 shows the results. Taking the estimates of different specifications as a benchmark, in Section 5 we show some simulated results. Finally, Section 6 discusses the conclusions reached.

2 Data and estimation of educational attainments

2.1 Data

The Indian Census is the most credible source of information on educational attainments across the states of India. It contains decade information on the educational levels of the population classified by age and sex. The educational categories include illiterates, incomplete primary, primary, middle, matriculation, higher secondary, non-technical diploma, technical diploma and graduates and above.[2] We take the five available data points in census ranging from 1961 to 2001 as a benchmark and compute the annual attainment levels with enrollment figures. The total number of students annually enrolled in primary (classes I-V), middle (classes VI-VIII) and secondary/higher secondary (classes IX-XII) are taken from *Growth of Enrolment in School Education 1950-51 to 1993-94* (*Planning, Monitoring & Statistics Division* at the *Ministry of Human Resource Development).* The number of enrollees in classes I-V, VI-VIII, IX-XII for the years 1994-2000 are sourced from annual publications *of Education in India (Government of India, Ministry of Hu-*

[2] Some of these categories are grouped differently for earlier census years.

man Resource Development, Department of Education, Planning, Monitoring & Statistics Division). We use the population by age groups from the census to compute the gross enrollment ratios.[3] Data on enrollments in tertiary education (including, among other things, professional education and diplomas) are sourced from annual publications of *Education in India* (1965-1979) and *Selected Education Statistics* (1980s onwards).

Data on real net domestic product and standard determinants of the economic growth rates for the main Indian states for the period 1960-2000 are taken from Besley and Burgess (2000, 2002, 2004) and updated by Ghate and Wright (2010).

2.2 Estimation methodology of annual educational attainments

We follow a perpetual inventory method in line with Barro and Lee (2001) to estimate annual attainments. The procedure consists in taking data on educational attainments as benchmark stocks and using enrollment data, with appropriate lags, to measure the new entrants as flows that add to the stock. Annual observations on school attainment for the population 15 and above are computed as follows. Let t only refer to the decade year where census data is available. In our dataset, $t = 1961, 1971, 1981, 1991$ & 2001. Let $H_{j,t}$ denote the number of population 15 years and above for whom j is the highest level of education attained; $j = 5$ refers to complete tertiary, 4 complete secondary, 3 complete middle, 2 complete primary, 1 incomplete primary and 0 no schooling. *HIGH*, *SEC*, *MDL* and *PRI* are the gross enrollment ratios in tertiary, secondary, middle and primary, respectively. The variable L_m refers to the total population aged m years old. For example, $L_{t,15}$ is the total population aged 15 years old at time t, $L_{t,20-24}$ is the total population ranging between 20 and 24 years old at time t, and so on. The variable δ_t is the mortality rate for the population 15 years and above between year t and $t-10$ and has been estimated by using the formula

$$\delta_t = 1 - \left(\frac{L_{t,15+} - L_{t,15-24}}{L_{t-10,15+}} \right)^{0.1}$$

We estimate annual completion ratios ($CR_{t,j}$) for each educational level using census data (see Appendix for the estimation method used to calculate them). Given the

[3] The age groups for each educational level are 6-11 years old for primary, 11-14 years old for middle and 14-18 years old for matriculation and higher secondary.

completion rates, the implicit annual stock for tertiary education is give by

$$H_{5,t+i} = H_{5,t} * (1-\delta)^i + \sum_{j=1}^{i} CR_{t,5} * HIGH_{t+j} * L_{t+j,21} * (1-\delta)^{i-j} \quad (1)$$

where the subscript i refers to each year within the decade for which census data is not available, with $i = 1, ..., 10$. Similarly,

$$H_{4,t+i} = H_{4,t} *(1-\delta)^i + \sum_{j=1}^{i} [(CR_{t,4} * SEC_{t+j} * L_{t+j,16}) - (CR_{t,5} * HIGH_{t+j} * L_{t+j,21})]*(1-\delta)^{i-j} \quad (2)$$

$$H_{3,t+i} = H_{3,t} *(1-\delta)^i + \sum_{j=1}^{i} [(CR_{3,t} * MDL_{t+j-1} * L_{t+j,15}) - (CR_{4,t} * SEC_{t+j} * L_{t+j,16})]*(1-\delta)^{i-j} \quad (3)$$

$$H_{2,t+i} = H_{2,t} *(1-\delta)^i + \sum_{j=1}^{i} [(CR_{2,t} * PRI_{t+j-5} * L_{t+j,15}) - (CR_{3,t} * MDL_{t+j-1} * L_{t+j,15})]*(1-\delta)^{i-j} \quad (4)$$

$$H_{0,t+i} = H_{0,t} *(1-\delta)^i + \sum_{j=1}^{i} (1 - PRI_{t+j-5}) * L_{t+j,15} * (1-\delta)^{i-j} \quad (5)$$

$$H_{1,t+i} = 100 - H_{0,t+i} - H_{2,t+i} - H_{3,t+i} - H_{4,t+i} - H_{5,t+i}$$

In census tables reporting age, population are not available for each age. We handle this by apportioning equally the population in an age band to each age.

Our algorithm ensures, by construction that we match the actual data attainments in the census years. Our method is also relevant for scenarios where Enrollment figures are overstated (as is often the case with developing countries). Given the fact we force the completion ratios to be such that it has to be consistent with the initial attainment and the final attainment of the population every ten years means that it adjusts the possible biases in reporting enrollment decade wise for each state.

Using the attainment data for each level of schooling we also compute a human capital Gini coefficient:

$$Gini = \frac{1}{2\overline{H}} \sum_{i=0}^{5} \sum_{j=0}^{5} \mid \widehat{x}_i - \widehat{x}_j \mid n_i n_j$$

where $\overline{H}$ are the average years of education of the population 15 years and above, i and j stand for the different levels of education: no schooling (0), incomplete primary (1), complete primary (2), complete middle (3), complete secondary (4) and complete tertiary (5); n_i and n_j are the shares of population with a given level of education, and $\widehat{x}_i$ and $\widehat{x}_j$ are the cumulative duration in years of of each education level. We take 0 years for no schooling, 3 years for incomplete primary, 5 years for complete primary, 8 years for complete middle, 10 years for complete secondary and 15 years for complete tertiary.

Figure 1 plots the different measures of education for the 16 states at the beginning and at the end of the sample year. As preliminary evidence, it is worth noting that the scatter plot of the average years of schooling, and mainly that of the Gini coefficient highly resembles the evolution of the share of illiterates. It seems that aggregate measures of education, such as the average years of schooling and the Gini coefficient, are mainly picking up the fact that the majority of the population in 1961 were illiterates and the effort the states have made in reducing this share over the years. The bottom part of the figure shows the evolution of the higher levels of schooling in each state. Overall, all the states have experienced extraordinary rates of growth in secondary and tertiary education, given the extremely low starting levels. In the case of tertiary education, the figure shows that in 1961 most of the states concentrated around a value of 0.4 percent. Among these states Gujarat, Andhra Pradesh and Tamil Nadu display values of tertiary education around 6 percent in 2001. At the bottom of the distribution in 2001 are Bihar, Rajastan and Assam with a share of tertiary schooling lower than 5 percent. In all figures Kerala stands out as an outlier state with much lower levels of illiteracy and higher share of population with secondary and tertiary schooling.

3 Empirical Model

In this section, we wish to specify models that allow us to test whether a highly educated labour force, albeit in an otherwise low literacy state can grow faster than a state where literacy is more widespread but where the average education level among the literates is not very high. In other words, we would like to see the effect of the distribution of educational achievement in the population on economic growth after controlling for other covariates. In particular, we would like to investigate, as a thought experiment, whether a state can afford to have a higher illiteracy rate and yet grow faster because a larger share of its literate population has completed tertiary education.

A specification that has been used before (e.g. Castello and Domenech, 2002) includes the average years of education and the Gini of education as additional regressors in an otherwise standard econometric model of growth. This standard model, often used in the context of cross country growth regressions, would regress growth rates on the usual covariates like capital stock, fertility rate, initial GDP and other variables, the choice of which differ depending on the focus of the paper. The inclusion of the Gini, controlling for average years of schooling, measures the distributional impact of education on growth. However, as noted above, given high illiteracy rates, the average years of education and the Gini capture the same idea. Since a very small fraction of the population are educated, the total stock of education is concentrated. Thus, in countries with high illiteracy rates, all we can capture with this specification is that illiteracy is bad. But this is well known and not the focus of the paper. Here, to begin with, we would like to see the effect of the distribution of education among the literates, after controlling for, among other things, illiteracy rates and the average years of schooling among the literate. The inclusion of these controls keep the size of the pie constant. Thus, the first empirical model we estimate in this paper uses this specification.

Let g_{it} denote the growth rate of per capita GDP, y, of the ith state, between years $t+1$ and t; denote the literacy rates by S_{it}^{ILL}, the average years of schooling among the literate by $\overline{Edu}_{it}^{Lit}$ and the Gini coeffient of education among the literate by $Gini_{it}^{Lit}$. Let μ_i capture all state i specific time invariant heterogeneity and Z_{it} be the other observables which determine growth (which we discuss later). Then

$$g_{it} = \alpha + \mu_i + \beta_1 y_{it} + \beta_2 S_{it}^{ILL} + \beta_3 \overline{Edu}_{it}^{Lit} + \beta_4 Gini_{it}^{Lit} + \Pi Z_{it} + \xi_{it} \qquad (6)$$

Note, however, that in developing countries, where the education structure of society is very concentrated, even the use of the Gini among the literate may not be very informative. This is because the share of population who complete and stop at lower levels of education will be enormous as compared to those who complete higher studies. Given the high correlation $Gini_{it}^{Lit}$ and $\overline{Edu}_{it}^{Lit}$, it is very probable that the specification will not be able to identify the impact of the two seperately. However, we start with this specification to illustrate a methodological issue in model specification while dealing with countries/states where the distribution of education is very uneven and where one is interested in the effect of the education level held by very few. We also estimate other spefications that substitute $Gini_{it}^{Lit}$ with other attributes of the distribution that are commonly used in the literature: the share of the top 10 percent and top 1 percent of the population. In fact, in the context of income inequality, Voitchovsky (2005) states that aggregate indicators of inequality, as measured for example by the Gini coefficient, could mask the different effect that the lower and upper part of the income distribution have on growth. We include only indicators of the upper part of the distribution since the large number of illiterates gives values equal to zero to the bottom percentiles.

We also suggest a specification that may do a better job in extricating the impact of tertiary education: one that is held by very few. Consistent with the notation before, let us denote the share of the labor force with completed education level j by S^j, where $j = PRI, SEC$ and $TERT$ where PRI denotes primary schooling (and includes incomplete primary), SEC represents complete and incomplete secodary (includes middle and higher secondary), and $TERT$ stands for complete tertiary education[4]. Thus the second empirical model we estimate is

$$g_{it} = \alpha + \mu_i + \delta_1 y_{it} + \delta_2 S_{it}^{PRI} + \delta_3 S_{it}^{SEC} + \delta_4 S_{it}^{TERT} + \Pi Z_{it} + \xi_{it} \tag{7}$$

We omit the share of illiterates in the labor force. Hence δ_i measures what would be the effect on growth if a unit share of the illiterates in the labor force were to acquire i th level of education. Hence it is expected that, if education is useful, δ_i would always be greater than 0.

[4] Due to classification problems, we cannot treat higher secondary and middle as a separate level. We do not want to include higher secondary as incomplete tertiary as the effect of university/professional degrees may be disproportianely large as compared to higher secondary.

The parameters in (2) show the merit of having populations with different levels of education. However, they show the tradeoffs with respect to the omitted category, i.e., illiteracy. The results of the estimation of this specification would essentially establish that literacy is good (if all parameters are positive) and which education level is most productive for growth for a given literacy rate (if $\delta_4 > \delta_3 > \delta_2$). In this paper, we want to build on these results. Our objective is to compare two alteratives: a state at a given literacy levels can have two alternatives: It could concentrate on bringing more people into education, that is lower literacy levels. Some, but not all of these people would complete tertiary education. The other alternative would be to concentrate on those who are literate and to ensure larger proportion of them complete tertiary education.

Given the objective of this paper, we would like to respecify (2) so as to make transparent on what is trade off between greater illiteracy and higher attainment of any one particular level of education (at the cost of other types). Thus we estimate separately the following two equations

$$g_{it} = \alpha^k + \mu_i^k + \beta_1^k y_{it} + \beta_2^k S_{it}^{ILL} + \beta_3^k S_{it}^k + \beta_4^k S_{it}^{ILL} * S_{it}^k + \Pi^k Z_{it} + \varepsilon_{it}^k;\ k = PRI, TERT \quad (8)$$

This particular specification is motivated by the exercise we are interested in. As an illustration, consider the case of the equation with $TERT$. In this equation, we would like to investigate how much higher the share of population with tertiary education would have to be to compensate for a higher illiteracy rate. In other words, can an economy with lower literacy rate grow as fast by having a higher proportion of people with complete tertiary education? This would involve looking at two comparative statics. First we look at the derivative with respect to share of illiteracy:

$$\frac{\partial g}{\partial S^{ILL}} = \beta_2^{TERT} + \beta_4^{TERT} S_{it}^{TERT} \quad (9)$$

This comparitive static gives us the impact of higher illiteracy rates. We allow the impact to be different depending on the existing tertiary level attainment of the population. Next we look at the derivative with respect to share of tertiary education.

$$\frac{\partial g}{\partial S^{TERT}} = \beta_3^{TERT} + \beta_4^{TERT} S_{it}^{ILL} \quad (10)$$

Given these two partial effects, the aim is then to calculate the change in the share of population with tertiary education required to offset a unit (or x percent) higher share of illiteracy.[5][6]

There are two major econometric issues that arise in the estimation of our specifications. The first one relates to the methods used to estimate the dynamic panel model. The usual problems are dealt with using Arellano and Bond (1991) and Blundell and Bond (1998) estimators in the growth literature. These estimators involve using all available lagged values as instruments to take care of endogeneity that spring out of the dynamic panel structure. However, in our case, these methods are not useful as these estimators have been devised for problems where i is large and t is small. Indeed the endogeneity problem is substantial mostly in the case of large i and small t. Since our data set has 40 years of data on 16 states, we have long t and small i. The inconsistency, if any, in our estimators are likely to be small (e.g. Nickell, 1981). Thus we can run fixed effects methods to estimate all of our specifications.

What is potentially more problematic for our exercise is endogeneity even after we take into account fixed effects. While we can use rainfall to instrument for initial income y_t, what could be critical for us, especially since we use state level data, is endogenous choice of where to locate. While migration rates within India have been found to be low especially during the period of our sample, it could be contended that migrants with tertiary education are more mobile and locate themselves in urban centres that are rich in the first place. Therefore what we pick up as the effect of growth of tertiary education in our within estimator is the effect of income growth trends. Notice though, that controlling for intial income already factors out this confounding factor and the effect of tertiary education is over and above that. However, to purge our estimators of any additonal endogeneity, we look for an instrument for tertiary education. Using an auxillary regression (results available on request), we find that controlling for y, Z and the different shares,

[5] We also test models with other nonlinearities using square terms.

[6] We deliberately do not run the specification with SEC. A unit increase in SEC may lead to a lower number of people with PRI among the literates. This will imply higher attainment for the population. On the other hand, higher SEC could also be at the cost of $TERT$ which reduces the attainment of the literates. Hence it is not possible to interpret the coefficent.

the fifth lag of the share of tertiary rate has no significance in explaining growth rates. However, S_{t-5}^{TERT} is correlated with S_t^{TERT} controlling for all exogenous regressors[7]. Thus we use S_{t-5}^{TERT} to instrument for S_t^{TERT} in all regressions. To further validate our results, we investigate if growth of S_t^{TERT} depends on past growth rates of states. For example, it can be contended that people with tertiary education do not just settle down in states with higher income but in states that have grown faster in the previous periods. To investigate this, we regress S_t^{TERT} on S_{t-1}^{TERT} and the lagged growth rate. If the coefficient of lagged growth rate is insignificant, this would further substantiate our result that growth of share of tertiary education is not picking up growth trends.

3.1 Main Results

In all regressions, in addition to the variables that capture the impact of education, we use rainfall during the year $t+1$, $Rainfall_{i,t+1}$, to proxy for agricultural shocks; the intial per capita income, y_t, to control for convergence in income across states; the rural population of the state, POP_t^R, to proxy for the size of the agriculture sector; the proportion of total roads unsurfaced, $Infraest_t$, to proxy for the level of infrastructure. Following previous work by Besley and Burguess (2000, 2004) and others, we use total expenditures, $TEXP_t$, and development expenditures, $DEXP_t$, to measure fiscal policy at the state level. As noted in the previous section, we control for state level heterogeneity by running fixed effects regression.

We also report the results using instrumental variables. It is possible that individuals with tertiary education migrate to states that are rich or that have grown faster in the previous periods. We explore this possibility in Table 1, which displays the effect of lagged growth and lagged per capita income on the current share of population with tertiary education. Results show that the share of tertiary education is not determined by the previous growth rates, the coefficient of lagged growth is not statistically significant in any specification. However, lagged per capita income has a positive and significant influence on the current level of tertiary education. Thus, we factor out this effect by including initial per capita income as an additional control and by using rainfall as an

[7] We get a similar result for S_{t-2}^{TERT} but we use S_{t-5}^{TERT} to be sure that we are not picking up the effect of the lagged share on growth.

instrument for income (e.g. Miguel et al., 2004, Ciccone, 2008). Moreover, the fifth lag of the share of tertiary rate is used as an instrument for the attainment levels in higher schooling.[8]

We start by analyzing the influence that inequality in the distribution of education may have on the growth rates. In most developing countries the large number of illiterates among the adult population suggests that typical distributional measures, such as the Gini coefficient, will not be informative. As noted in Table A1, the correlation between the Gini coeffcient and the share of illiterates across the Indian states is 0.994 and that of illiterates and the average years of schooling is 0.943. Thus, while the high collinearity among these variables prevents to disentagle its individual effect, it is also likely that the Gini coefficient will not pick up a distributional effect but the negative influence of the large share of illiterates.

As an alternative, we compute the Gini coefficient and the average years of education among the literates. The results of their effect on the growth rates are displayed in Table 2. We find that controlling for the aggregate education in society (given by the share of the population illiterate and the average years of schooling among the literate), the Gini coefficient among the literates is not statistically significant in both the instrumented and non instrumented versions (columns 1 and 4). Measures of the share of education attained by the top end of the distribution- the share of the top 1% (columns 2 and 5) and top 10% (columns 3 and 6)- are neither statistically significant. This result points out that, in the case of the developing countries, these distributional measures do not convey any more information than that conveyed by the averages years or the share of illiteracy. Predictably, we find that the effect of illiteracy is negative and its effect on growth rates range from 0.45 to 0.83. Similarly, the effect of average years of education among the literates is positive and its coeffcient ranges from 0.052 to 0.087.[9] The results in Table 2

[8] After controlling for all the explanatory variables, S_{t-5}^{TERT} has no significance in explaining growth rates but it is highly correlated with S_t^{TERT}.

[9] It can be contended that the $\overline{EDU}_{it}^{LIT}$, controlling for the share of illiterates mirrors the distribution of education among the literates in the case of developing countries. However, one needs to make assumptions on how many years of education to impute to those who have incomplete education in any category. This is especially true for categories of larger mass, for example, incomplete primary where drop out rates are high. Thus the average will be very sensitive to the choice of imputation. We therefore do not use this measure to look at the distributional impact.

also show that the effect on growth of the additional controls is as expected. We find that greater rainfall has a positive impact on economic growth. An increase in the total revenue expenditure discourages the growth rates whereas greater development expenditure has a beneficial influence. Results also show that a greater share of population living in rural areas is negatively related to economic growth and better infrastructure, measured as the share of surfaced roads, promotes development. Finally, in line with Nagaraj et al. (2000), we also find evidence of conditional convergence across States; the coefficient of initial per capita income is negative and significant at 1 percent level.

To extricate the distributional impact, we look at the specification with the shares of labor force with different levels of education (eq. [2]). For the rest of this section, to avoid confusion, we discuss parameters from the instrumented version unless there is a difference in verdicts between the results with instrumental variables and the one without instruments. In Table 3 the share of all schooling levels are inlcuded and the share of illiterates is the omitted category. Therefore the coefficient of each share is its trade-off with S^{ILL}. It can be seen that the returns to a state from 1 percent more people with complete tertiary education is more than 10 times the return to just entering primary school. This points out to a huge return to a state if it can ensure more labor force with tertiary education. This could be done in one of two ways. One by attracting more manpower with tertiary education to its state or by ensuring that its students who enter school complete tertiary education and stay back. The coefficient of SEC is surprisingly insignificant. This is because of collinearity between S^{SEC} and S^{TERT} (the correlation between both variables is 0.883). This can be seen in the specification when we drop S^{TERT} (columns 2 and 5). S^{SEC} now becomes positive and significant. That the coefficient of S^{TERT} primary picks up the effect of tertiary education and not the intermediate level is immediately apparent from an inspection of columns 1 and 4.

The results so far show that education is good and higher education is better. This points out to the usefulness of using the shares of each level of schooling in specifications without making assumptions about the years of education of each person. Next, we want to look at the trade-offs of better completion of higher education with a lower proportion of literates. Out next empirical exercise, displayed in Table 4, looks at the estimation results of eq. [3].

Columns (1) and (3) include both the share of illiterates and the share of population with tertiary education in the set of controls. As expected, the coefficient of the proportion of illiterates is negative and statistically significant at the 1 percent level and that of the share of tertiary is positive and also significant at 1 percent. In quantitative terms the results show that both reducing the number of illiterates and increasing the proportion of highly educated individuals have a large impact on the economic growth rates. Nevertheless, the impact of increasing the share of tertiary almost triples that of reducing the share of illiterates. According to column (3), a one standard deviation increase in the share of illiterates (0.151) reduces the growth rate by 0.046 percentage points, whereas a one standard deviation increase in the share of tertiary education (0.018) increases growth by 0.141 percentage points.

In columns (2) and (4) we examine whether a larger share of illiterates have a lesser impact when the level of tertiary education is very high. To do so, we add an interaction term between both variables. Results show that the interaction coefficient is positive but not statistically significant in any specification, suggesting that raising the share of illiterates always has a negative impact on growth and this effect is independent on the level of tertiary education.

3.1.1 Simulations

In the previous section, we have quantified how share of tertiary education and share of illiteracy affect growth rates. Using these estimates, we now address the implications of two different education policies using simulations. The baseline scenario, labelled *business as usual* (*BAU*) is one where all independent variables grow at their post 1990s trend rate. We start with the initial per capita NSDP (averaged across all states) in 2001 and simulate its predicted future path using our parameter estimates. For illustration, we exposit the scenarios for 35 years (2001-2035). Given this baseline, we conduct two counterfactual policies: The first policy is an emphasis on completing tertiary education, i.e. increasing the share of tertiary education in the economy. Such a policy would focus on both schools and universities to ensure there is low drop out among those who are enrolled. In our simulation, we model this policy as one that allows share of literate labour force with tertiary rates to grow faster than the baseline. According to our baseline, tertiary rates grow by 0.18 percentage points a year. We look at a policy that would lead to an increase at twice that rate (therefore an increase of 0.36 percentage point a year). Implicit in raising share of tertiary education is the need to not only reduce drop outs of the existing literates from schools but also increase enrolment and reduce drop outs in universities. We focus on increasing enrolment in the universities by assuming that a university enrolment push shows effect in the third year of the policy (the average duration of undergraduation). Given the levels and shares in 2001, this would imply that instead of the current 27 percent of the cohort that has just finished higher secondary schooling joining university, the policy would require that about 45 percent of such students pursue university education. The required enrolment could be even lower at around 40 percent if there are no dropouts from university. We label this policy as *tertiary push*.

An alternate policy would be to focus on raising literacy, i.e. providing a large mass of people basic schooling. In line with recent attempts for universalization of primary education, this would lead to a rise in literacy rates of children. Given that children join school at the age of 6, this will only show an effect as a rise in the literacy of population aged 15 and above 9 years after the policy is enforced. We simulate the effect of one such policy, labelled here as *literacy push*. The baseline scenario would lead to an increase in literacy rates at 1 percentage point a year. As a part of *literacy push*, literacy rates of the labour force rise by 2 percentage points 9 years after implementation. Subsequently, the literacy rate rises by an additional 2 percentage points each year as new waves of the more literate cohort reach the age of 15. As a part of this scenario, we also recognize that the greater base of literate people would also lead to an increase in the share of tertiary education. However, this effect would be felt only when the cohorts reach the age of 21. Thus in our simulations, the share of population with tertiary education picks up after the 15th year of the policy. We need to make an assumption here on what share of the literate population would complete tertiary. For the 15th year and beyond, we use the following method: First, for each year after the 15th year, we calculate the share of literate population who complete tertiary education in the baseline. The growth of this share over time in the baseline reflects improvements in other unmodelled factors that increase the tertiary completion rates when there is no push. Thus, if x is the proportion of literate population that completes tertiary education in 2016, we assume that share of tertiary education rises by $0.02{*}x$ in that year. We follow the same method for all subsequent years.

Our first simulation illustration constrasts the effect of the two policies mentioned above to BAU for the average economy. Figure 2 illustrates the result of the three policies. Clearly, either education policy is better than BAU confirming that they indeed represent better education policies than those in the 1990s. For example, at the end of 2035, the per capita NSDP under *tertiary push* is 1.8 times that under BAU. What is more interesting is the comparison of *tertiary push* to *literacy push*. The *tertiary push* will definitely yield results faster as compared to a policy that concentrates on raising literacy by focusing at initial schooling. It is equally true that in the long run, raising literacy and slow secular changes in tertiary completion will ultimately lead to better outcomes. However here we undertake the comparison in between the immediate and the very long run. We find that even at the end of 35 years, the *tertiary push* yields a per capita NSDP 1.5 times that under *literacy push*. Notice also that while the economy under *literacy push* kicks off after 9 years, it is still much lower after 15 years, even after the policy has started an effect on the share of labour force with tertiary education. Lower tertiary conversion rates imply that the economy can never quite take advantage of the higher number of literates, at least within the medium term that we look at.

In the second simulation, we ask whether its possible for the poorest states to catch up with richer ones by using these alternate education policies. Like before, we look at a 35 year horizon. We divide states into four quartiles based on their per capita NSDP in 2000. As an illustration, we consider the first ($Q1$) and second quartile ($Q2$) and look at the two education policies in the context of the poorest quartile of states. We follow a very similar method as above except that we allow different trend rates for each quartile (we average across states in each quartile). First we calculate the BAU scenario for second quartile. The question we want to answer is if it is possible for states in the first quartile to catch up using education policies. Figure 3 plots the scenario for both the policies. Our calculations show us that for the $Q1$ states to catch up using *tertiary push* , they need to increase their tertiary rates annually by 0.45 percentage points. On the other hand, if the states follow a *literacy push* strategy, they have to increase literacy rates by 5.8 percentage points each year. As Figure *** shows, what is interesting in the *literacy push* strategy is that there will be a big gap in the state performance for the first 15 years. However, there after the economy speeds up as the twin effects of big annual decreases in illiteracy as well as a larger mass going to tertiary education together create a big impact. This points out that in the case of states that lag behind in literacy levels, *literacy push* may well be the policy to follow. However this push has to be large every year to create these effects.

4 Conclusions

The link between human capital and growth is a well established one. However, there has been less emphasis on the distribution of education among the literates and its effect on growth. Most of the work has been done with data from OECD countries and there has been less emphasis on developing coutries. Among developing countries, India has at various points in its history emphasized both setting up of tertiary institutions as well as reducing illiteracy by attempting universal access to primary education. Given variations between states and over time in illiteracy rates and share of labor force with tertiary education, we seek to investigate the distributional impact of education on growth in this context. Data for developing countries, especially time series, are difficult to get for most countries and are many a times non comparable. For example a particular level of education across countries may reflect very different quality of knowledge that is accumulated. Thus working with data on Indian states is an attempt to look within a comparable data set. We investigate the link between the distributional impacts of education and growth using data from 16 major states of India for the period 1961-2000.

Many developing countries are characterized by very skewed distributions with large proportion of the labor force who are illiterate and a relatively tiny proportion of people who complete tertiary education. In this paper, we show that the usual measures like Gini, education level of the top 1% (or top 10%) that are usually advocated are not able to extricate the important effect of tertiary education. The reason is a statistical one: given the large proportion of illiterate people, the average education that controls for the size of the pie is collinear with most commonly used distribution descriptives like the Gini, education level of the top 1%. We find that in such scenarios, using shares of the adult population with different levels of education yield better results. Our result lays a case for its use for other similar contexts in developing countries.

While data from India is better than from most other developing countries, annual data on educational attainments are not available. We adopt a perpetual inventory method, which uses census data on educational attainment every ten years (from 1961) and enrolment figures reported annually, to construct the annual series for educational attainment. Our method deals with often reported biases in enrollments by using an algorithm such that it fits consecutive census data on attainments. We use this data to estimate a fixed effects model of growth. In the process we also take into account some endogeniety issues. In particular, we are aware of the possibility of endogenous location of tertiary educated people. We take care of this using appopriate instruments.

Our results show that one standard deviation decrease in the share of illiteracy increases growth by 0.046 percentage points where as one stardard deviation increase in the share of tertiary education increases growth by 0.141 percentage points. This result should be viewed in light of the strategy of many developing countries that stress on getting children to school to make them literate. A relevant example is that of India which has recently enforced the "Right to Education" that lays the ground for universal access to primary education. While this is important, our results point out that there are equally important and higher economic gains from focussing on school and tertiary education completion. This is not to understate the role of illiteracy which we find has a negative impact irrespective of the existing level of tertiary education in the economy. However empowering the existing share of literates with skills and education that a tertiary education brings can lead to faster growth in the short as well as long term.

We show this result is equally true in a dynamic context with covariates following their 1990s trends as the baseline. Using simulations derived from our estimates, we show the effect of two kinds of policies. One that focuses on tertiary education (*tertiary push*) the other that focus on increasing literacy (*literacy push*). In both policies, we allow the relevant rates to move up at twice their exisiting rate of increase. We assume the timing of *tertiary push* to be such as to simulate the effect of an increased proportion of high school graduates joining universities. We find that even at the end of 35 years, the *tertiary push* yields a per capita NSDP 1.5 times that under *literacy push*. In another simulation exercise we also show that the poorest states can catch up with the next richest quartile of states in 35 years if share of tertiary attainment rises annually by 0.45 percentage points or if the literacy rates rise by 5 percentage points as soon as (and subsequent to) the first wave of more literate cohorts reach the age of 15. This points out that growth through increasing literacy rate is possible if the annual rate of increase is large. However, in medium term, even these poorest states can grow faster if they focus on tertiary education.

These simulations are of course stark and are constructed as illustrations to bring out the point that tertiary education attainment is important even for societies with high illiteracy. The ideal world is of course one where both aspects are targetted equally but this is unlikely given that developing countries have limited resources. In so far as people with high school education are already a self selected group with lesser inclination to drop out, it may be easier to target them in the short run by ensuring they complete tertiary education. Our paper highlights the fact that it may indeed be profitable to do so.

References

[1] Arellano, M. and S. R. Bond (1991). "Some Test of Specification for Panel Data: Monte Carlo Evidence and and Application to Employment Equations". *Review of Economic Studies*, 58 (2), 277-97.

[2] Barro, R. and J-W Lee (2001). "International Data on Educational Attainment Updates and Implications," *Oxford Economic Papers*, III, 541-63.

[3] Barro, R. and J-W Lee (2010). "A New Dataset of Educational Attainment in the World, 1950-2010," NBER Working Paper N^o 15902.

[4] Benhabib, J. and M. Spiegel (1994). "The Role of Human Capital in Economic Development." Journal of Monetary Economics, 34 (2), pp. 143-174.

[5] Besley, T. and R. Burgess (2000). "Land Reform, poverty Reduction, and Growth: evidence from India," *Quarterly Journal of Economics*, 115(2), pp. 389-430.

[6] Besley, T. and T. Burgess (2004). "Can Labor Regulation Hinder Economic Performance? Evidence From India," *Quarterly Journal of Economics*, 19(1), pp. 91-134.

[7] Bils, M and P. Klenow (2000). "Does Schooling Cause Growth." American Economic Review, 90(5), pp.1160-1183.

[8] Blundell, R. and S. Bond (1998). “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models”. *Journal of Econometrics*, 87, 115-43.

[9] Bosworth, B., S. Collins and A. Virmani (2007). “Sources of Growth in the Indian Economy,” NBER working Paper N^o 12901.

[10] Castelló, A. and R. Doménech (2002). “Human Capital Inequality and Economic Growth: Some New Evidence”. *Economic Journal*, 112 (478), 187-200.

[11] Cohen, D. and M. Soto (2007). “Growth and Human Capital: Good Data, Good Results,” *Journal of Economic Growth*, 12(1), pp.51-76.

[12] Ciccone, A. (2008). “Transitoy Economic Schocks and Civil Conflict,” Mimeo ICREA-Universidad Pompeu Fabra.

[13] de la Fuente, A. and R. Doménech (2006). “Human Capital in Growth Regressions: How Much Difference Does Data Quality Make,” *Journal of the European Economic Association*, 4(1), pp.1-36.

[14] Ghate, Ch. and S. Wright (2009). “The "V-Factor": Distribution, Timing and Correlates of the Great Indian Growth Turnaround,” mimeo Indian Statistical Institute.

[15] Kochhar, K., U. Kumar, R. Rajan, A. Subramanian and I. Tokatlidis (2006). “India’s Pattern of Development: What Happend, What Follows?” *Journal of Monetary Economics*, 53, pp. 981-1019.

[16] Kremer, M, K. Muralidharan, N. Chaudhury, J. Hammer and F. H. Rogers (2005). “Teacher Absence in India,” *Journal of the European Economic Association*, 3, pp. 658-67.

[17] Miguel, E., S. Satyanath and E. Sargenti (2004). “Economic Shocks and Civil Conflict: An Instrumental Variable Approach,” *Journal of Political Economy*, 112(4), pp.25-53.

[18] Nagaraj, A., A. Varoudakis and M. A. Veganzones (2000). "Long-Run Growth Trends and Convergence Across Indian States," *Journal of International Development*, vol. 12, pp. 45-70.

[19] Nickell, S. (1981). "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, pp. 1417-26.

[20] Pissarides, Ch. (2000). "Human Capital and Growth: A Synthesis Report," OECD Development Centre Working Paper N^o 168.

[21] Psacharopoulos and H. Patrinos (2004). "Returns to Investment in Education: A Further Update," Education Economics, Vol. 12 (2), pp. 111-34.

[22] Vandenbussche, J., P. Aghion and C. Meghir (2006). "Growth, Distance to Frontier and Composition of Human Capital." *Journal of Economic Growth*,11, pp. 97-127.

[23] Voitchovsky, S. (2005) "Does the Profile of Income Inequality Matter for Economic Growth?: Distinguishing between the effect of inequality in different parts of the income distribution," *Journal of Economic Growth*, 10, 273-96.

5 Appendix

5.1 Estimation of Completion Ratios

We use the census data and assume they hold constant over the decade. As an illustration, let us explain how the completion ratio for tertiary education is calculated. If we want to compute the completion ratio between census years t and $t-10$ ($CR_{5,t}$), we start with the stock of people with tertiary education in the year $t-10$ ($H_{5,t-10}$). Given mortality rates δ, $(1-\delta)^{10}$ proportion of them will survive by the year t. Next, we look at how many new entrants are added to this stock. We assume that the average student completes tertiary education by the age of 21. Thus we need to account for how many people in the cohort $21-30$ in the year t have completed tertiary education[10]. Let $\overline{HIGH}_{t,t-10}$ be the average enrollment ratios over the decade. Thus, the new stock of tertiary educated

[10]One advantage of looking at the 21-30 cohort in the year t. is that it already takes into account mortality rates over the period.

people over the decade would be $CR_{5,t} * \overline{HIGH}_{t,t-10} * L_{t,21-30}$. Given everything else, the completion ratios for tertiary educational level for the decade between t and $t-10$ is then given my the following formula:

$$CR_{5,t} = \frac{H_{5,t} - (1-\delta)^{10} * H_{5,t-10}}{L_{t,21-30} * \overline{HIGH}_{t,t-10}} \tag{11}$$

For calculating the other completion ratios, we use the following assumptions. Student finish primary schooling by the age of 11, middle schooling by the age of 14, secondary school by 16[11]. While $\overline{SEC}_{t,t-10}$ is the average enrollment rate for secondary schooling, $\overline{MDL}_{t-1,t-11}$ is the average middle school enrollment between the years $t-1$ and $t-11$. Similarly $\overline{PRI}_{t-5,t-15}$ is the average enrollment rates in primary education for each decade. The last two enrollment rates are constructed based on lags with different start and end years because the minimum age to be counted in our attainment figures it 15. For example, it takes people who pass out middle schooling one year to be counted. Thus for the cohort of age 15, what matters for completion is what was the enrollment rate among them one year back. Arguments analogous to the one made for tertiary education (and to those made in Barro %%%%) yield the following

$$CR_{4,t} = \frac{H_{4,t} - (1-\delta)^{10} * H_{4,t-10} + L_{t,21-30} * \overline{HIGH}_{t,t-10} * CR_{5,t}}{L_{t,16-27} * \overline{SEC}_{t,t-10}} \tag{12}$$

$$CR_{3,t} = \frac{H_{3,t} - (1-\delta)^{10} * H_{3,t-10} + L_{t,21-30} * \overline{SEC}_{t,t-10} * CR_{4,t}}{L_{t,15-24} * \overline{MDL}_{t-1,t-11}} \tag{13}$$

$$CR_{2,t} = \frac{H_{2,t} - (1-\delta)^{10} * H_{2,t-10} + L_{t,15-24} * MDL_{t,t-10} * CR_{3,t}}{L_{t,15-24} * \overline{PRI}_{t-4,t-14}} \tag{14}$$

[11]Due to variations over time and across states in higher secondary, we do not look at that education category seperately

Appendix

Table A1

Summary Statistics

	Obs	Mean	Std. Dev	Min	Max
$\Delta \ln y$	641	0.023	0.073	-0.224	0.383
lny	642	8.650	0.385	7.691	9.636
S^{years}	656	3.044	1.223	0.777	7.090
Gini_h	656	0.670	0.116	0.311	0.904
$\mathrm{Gini}_h^{Literates}$	656	0.250	0.027	0.164	0.296
$\mathrm{S}^{Illiterates}$	656	0.561	0.151	0.101	0.872
$\mathrm{S}^{\Pr imary}$	656	0.222	0.083	0.068	0.547
$\mathrm{S}^{Secondary}$	656	0.189	0.090	0.033	0.490
$\mathrm{S}^{Tertiary}$	656	0.029	0.018	0.002	0.088
Rainfall	680	1.295	0.678	0.214	4.003
Total Expenditure	603	0.263	0.420	0.002	2.954
Development Expenditure	603	0.168	0.257	0.001	1.644
Rural population	679	32.906	23.484	2.965	140.308
Unsurfaced Roads	502	47.682	21.944	2.470	90.320

.

Table A2

Correlations among the main variables

	$\Delta \ln y$	lny	S_{years}	$Gini_h$	$S^{Literates}_{Years}$	$Gini^{Literates}_h$	$Sh^{Illiterates}$	$Sh^{\mathrm{Pr}imary}$	$Sh^{Secondary}$	$Sh^{Tertiary}$	top_h_1	top_h_10
$\Delta \ln y$	1											
lny	0.011	1										
S^{years}	0.089	0.726	1									
$Gini_h$	-0.071	-0.636	-0.961	1								
$S^{Literates}_{Years}$	0.074	0.577	0.599	-0.380	1							
$Gini^{Literates}_h$	-0.005	-0.351	-0.367	0.268	-0.649	1						
$Sh^{Illiterates}$	-0.071	-0.615	-0.943	0.994	-0.321	0.169	1					
$Sh^{\mathrm{Pr}imary}$	0.018	0.209	0.475	-0.683	-0.354	0.306	-0.737	1				
$Sh^{Secondary}$	0.081	0.686	0.970	-0.892	0.712	-0.498	-0.856	0.284	1			
$Sh^{Tertiary}$	0.112	0.774	0.883	-0.737	0.784	-0.355	-0.714	0.130	0.882	1		
top_h_1	-0.089	-0.657	-0.900	0.870	-0.612	0.352	0.859	-0.467	-0.852	-0.795	1	
top_h_10	-0.059	-0.659	-0.874	0.890	-0.369	0.310	0.877	-0.565	-0.854	-0.731	0.807	1

Note: 641 Observations.

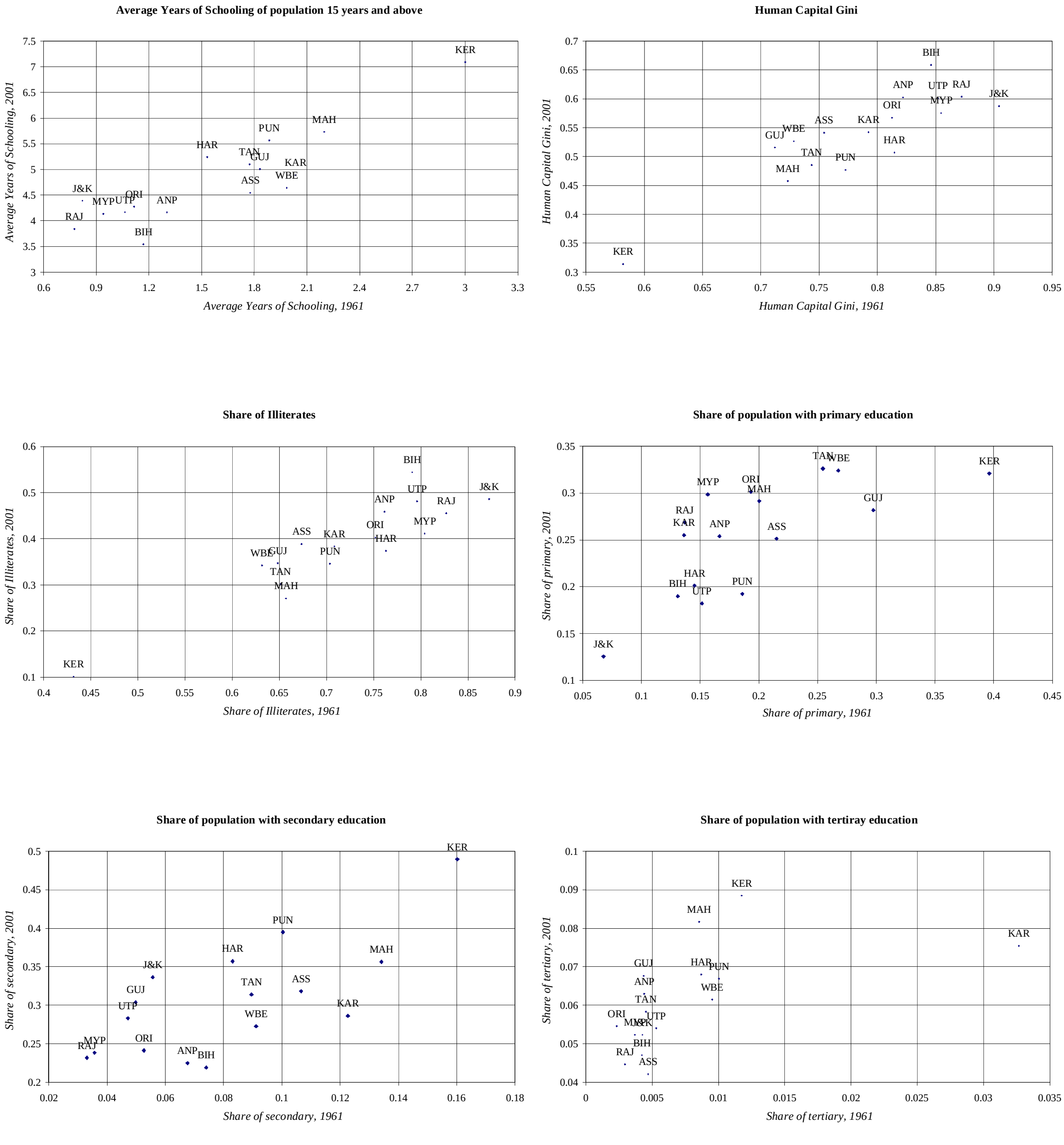

Figure 1: Different measures of schooling in 1961 and in 2001.

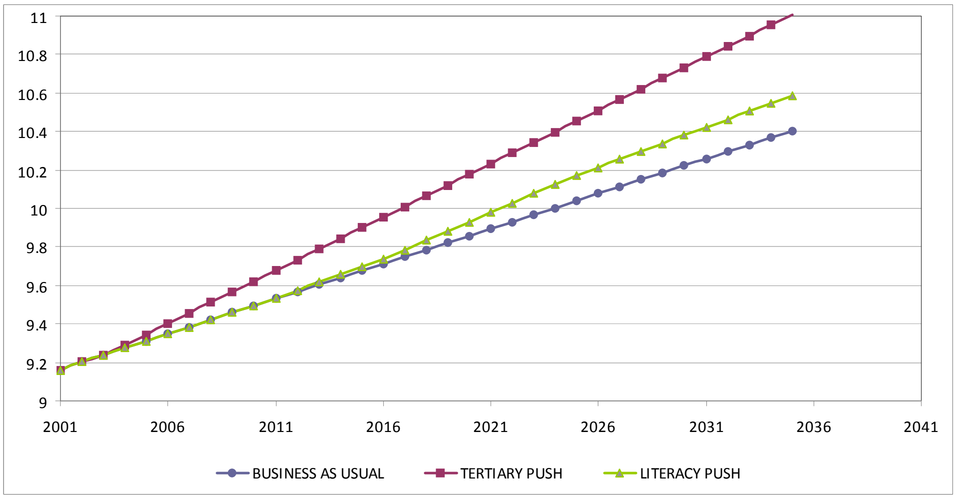

Figure 2: Simulated per capita NSDP. Tertiary push vs literacy push

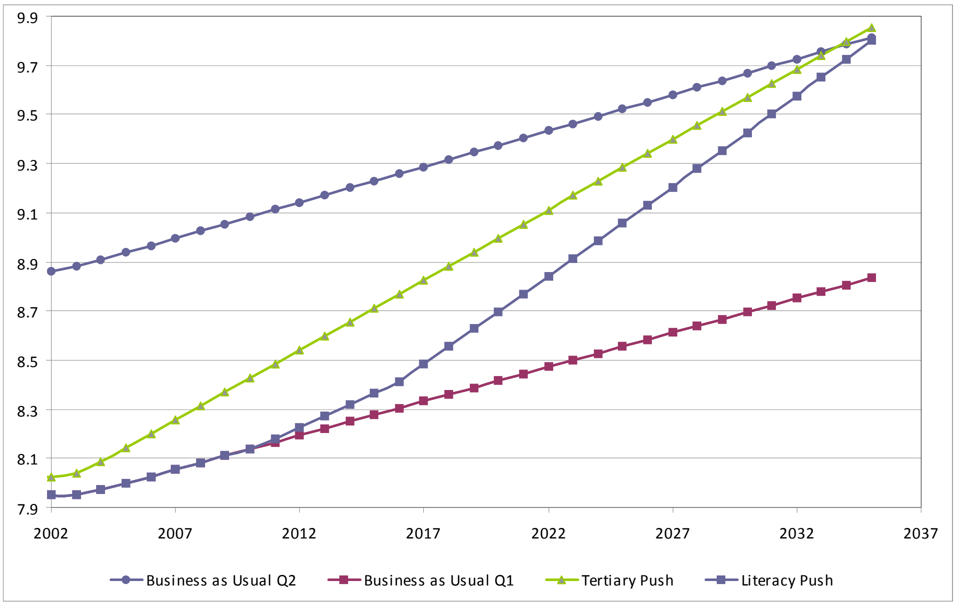

Figure 3: Simulated per capita NSDP. Convergence across poor states

Table 1

Dependent variable: Population with tertiary education, $S_{i,t}^{TERT}$

	FE (1)	FE (2)
$\Delta \ln y_{i,t-2}$	0.00001	-0.00048
	(0.0049)	(0.00052)
$S_{i,t-1}^{TERT}$	1.02631^{a}	1.01101^{a}
	(0.0023)	(0.00601)
$lny_{i,t-1}$		0.00106^{a}
		(0.00039)
$Constant$	0.00060^{a}	-0.00816^{b}
	(0.00607)	(0.00300)
R^2	0.9971	0.9971
States	16	16
Observ.	609	609

Note: Fixed Effects estimation. Robust standard errors in parenthesis. a, b, c stand for significance level at 1, 5 and 10 per cent respectively.

Table 2

Dependent variable: Growth rate of per capita income, $\Delta \ln y_{i,t,t+1}$

	FE			FE-IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln y_{i,t}$	-0.397[a]	-0.397[a]	-0.394[a]	-0.692[a]	-0.688[a]	-0.687[a]
	(0.038)	(0.038)	(0.038)	(0.133)	(0.133)	(0.134)
$\mathrm{S}_{i,t}^{ILL}$	-0.452[a]	-0.495[a]	-0.523[a]	-0.773[a]	-0.827[a]	-0.809[a]
	(0.074)	(0.109)	(0.088)	(0.159)	(0.186)	(0.157)
$Gini_{i,t}^{Lit}$	0.225			0.203		
	(0.273)			(0.291)		
$Edu_{i,t}^{Lit}$	0.057[a]	0.052[a]	0.054[a]	0.087[a]	0.084[a]	0.082[a]
	(0.012)	(0.012)	(0.010)	(0.019)	(0.019)	(0.016)
$S_{i,t}^{TOP1}$		0.279			0.472	
		(0.879)			(0.940)	
$S_{i,t}^{TOP10}$			1.411			0.680
			(1.367)			(1.493)
$Rainfall_{i,t+1}$	0.091[a]	0.091[a]	0.091[a]	0.100[a]	0.100[a]	0.100[a]
	(0.012)	(0.013)	(0.012)	(0.014)	(0.014)	(0.014)
$TEXP_{i,t}$	-0.124[b]	-0.148[b]	-0.136[b]	-0.159[b]	-0.183[a]	-0.172[a]
	(0.062)	(0.059)	(0.058)	(0.068)	(0.064)	(0.063)
$DEXP_{i,t}$	0.361[a]	0.399[a]	0.383[a]	0.563[a]	0.597[a]	0.585[a]
	(0.108)	(0.100)	(0.100)	(0.144)	(0.137)	(0.139)
$POP_{i,t}^{R}$	-0.003[a]	-0.003[b]	-0.003[a]	-0.006[a]	-0.005[a]	-0.005[a]
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
$Infraest_{i,t}$	-0.001[a]	-0.001[b]	-0.001[b]	-0.002[a]	-0.002[a]	-0.002[a]
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Constant	3.258[a]	3.348[a]	3.294[a]	5.871[a]	5.890[a]	5.900[a]
	(0.371)	(0.353)	(0.353)	(1.193)	(1.172)	(1.202)
R^2	0.280	0.270	0.280			
States	16	16	16	16	16	16
Observ.	454	454	454	454	454	454

Note: Fixed Effects estimation. Robust standard errors in parenthesis. a, b, c stand for significance level at 1, 5 and 10 per cent respectively.

Table 3

Dependent variable: Growth rate of per capita income, $\Delta \ln y_{i,t,t+1}$

	FE			FE-IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln y_{i,t}$	-0.423[a]	-0.360[a]	-0.422[a]	-0.809[a]	-0.699[a]	-0.795[a]
	(0.038)	(0.037)	(0.038)	(0.163)	(0.140)	(0.157)
$\mathrm{S}_{i,t}^{PRI}$	0.220[a]	0.203[b]	0.216[a]	0.415[a]	0.397[a]	0.427[a]
	(0.081)	(0.083)	(0.080)	(0.122)	(0.119)	(0.121)
$\mathrm{S}_{i,t}^{SEC}$	0.072	0.785[a]		-0.458	1.431[a]	
	(0.184)	(0.112)		(0.312)	(0.284)	
$\mathrm{S}_{i,t}^{TERT}$	4.998[a]		5.326[a]	12.608[a]		10.129[a]
	(1.038)		(0.614)	(2.988)		(1.968)
$Rainfall_{i,t+1}$	0.087[a]	0.090[a]	0.087[a]	0.092[a]	0.100[a]	0.093[a]
	(0.012)	(0.013)	(0.012)	(0.014)	(0.014)	(0.014)
$TEXP_{i,t}$	-0.107[c]	-0.196[a]	-0.101[c]	-0.043	-0.277[a]	-0.088
	(0.059)	(0.058)	(0.057)	(0.073)	(0.071)	(0.063)
$DEXP_{i,t}$	0.281[a]	0.473[a]	0.266[a]	0.253[c]	0.781[a]	0.351[a]
	(0.106)	(0.101)	(0.099)	(0.136)	(0.165)	(0.119)
$POP_{i,t}^{R}$	-0.001[c]	-0.001	-0.001[c]	-0.004[a]	-0.003[b]	-0.004[a]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$Infraest_{i,t}$	-0.001[b]	-0.001[b]	-0.001[b]	-0.002[b]	-0.002[a]	-0.002[a]
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Constant	3.432[a]	2.880[a]	3.426[a]	6.700[a]	5.700[a]	6.560[a]
	(0.325)	(0.312)	(0.324)	(1.370)	(1.171)	(1.316)
R^2	0.290	0.250	0.290			
States	16	16	16	16	16	16
Observ.	454	454	454	454	454	454

Note: Fixed Effects estimation. Robust standard errors in parenthesis.
a, b, c stand for significance level at 1, 5 and 10 per cent respectively.

Table 4

Dependent variable: Growth rate of per capita income, $\Delta \ln y_{i,t,t+1}$

	FE		FE-IV	
	(1)	(2)	(3)	(4)
$\ln y_{i,t}$	-0.422^a	-0.423^a	-0.767^a	-0.754^a
	(0.038)	(0.038)	(0.149)	(0.146)
$S_{i,t}^{ILL}$	-0.202^a	-0.222^b	-0.307^a	-0.362^a
	(0.077)	(0.086)	(0.105)	(0.118)
$S_{i,t}^{TERT}$	4.152^a	3.770^a	7.808^a	6.480^a
	(0.656)	(0.973)	(1.575)	(1.687)
$S_{i,t}^{ILL} * S_{i,t}^{TERT}$		0.939		2.909
		(1.767)		(2.519)
$Rainfall_{i,t+1}$	0.087^a	0.088^a	0.094^a	0.095^a
	(0.012)	(0.012)	(0.014)	(0.014)
$TEXP_{i,t}$	-0.120^b	-0.112^c	-0.124^b	-0.099
	(0.057)	(0.059)	(0.063)	(0.066)
$DEXP_{i,t}$	0.308^a	0.300^a	0.421^a	0.390^a
	(0.100)	(0.101)	(0.125)	(0.125)
$POP_{i,t}^{R}$	-0.001^c	-0.002^c	-0.004^a	-0.005^a
	(0.001)	(0.001)	(0.001)	(0.002)
$Infraest_{i,t}$	-0.001^b	-0.001^b	-0.002^a	-0.002^a
	(0.000)	(0.000)	(0.001)	(0.001)
$Constant$	3.623^a	3.653^a	6.630^a	6.583^a
	(0.349)	(0.354)	(1.313)	(1.297)
R^2	0.290	0.290		
States	16	16	16	16
Observ.	454	454	454	454

Note: Fixed Effects estimation. Robust standard errors in parenthesis. a, b, c stand for significance level at 1, 5 and 10 per cent respectively.