



WIDER Working Paper 2018/187

## **Intergenerational mobility, human capital accumulation, and growth in India**

Roy van der Weide<sup>1</sup> and Melinda Vigh<sup>2</sup>

December 2018

**Abstract:** Productivity and socio-economic progress are inter-connected. Economic growth funds policies that promote socio-economic progress, while the latter serves as a growth engine. A society with high mobility is one where individual achievements are influenced less by the individual's parents and more by her potential. In an application to 54 state-regions in India (1983-2011), we identify the causal effect of intergenerational mobility on household expenditure's future growth. We find that intergenerational mobility has a positive effect on growth, particularly among the poor. Furthermore, mobility is found to be positive for human capital accumulation except for individuals with highly educated parents for whom it matters less whether the playing field is level or not.

**Keywords:** intergenerational mobility, inequality, growth, poor, India

**JEL classification:** D31, D63, E24, O15, O40

**Acknowledgements:** The authors are grateful to Himanshu and Rinku Murgai for their help with the India National Sample Survey data, to Rinku Murgai and Peter Lanjouw for their help with constructing the state-region identifiers, and to Abhiroop Mukhopadhyay for pointing us to the sub-national data that is available for India. The authors would like to thank Hai-Anh Dang, Chris Elbers, David Garcés Urzainqui, Himanshu, Peter Lanjouw, Abhiroop Mukhopadhyay, Rinku Murgai, and Gerton Rongen for their comments on earlier versions of the paper. The authors also gratefully acknowledge financial support received from UNU-WIDER.

---

<sup>1</sup> World Bank, Washington, DC, corresponding author: [rvanderweide@worldbank.org](mailto:rvanderweide@worldbank.org); <sup>2</sup> Vrije Universiteit, Amsterdam, The Netherlands.

This study has been prepared within the UNU-WIDER project on '[Inequality in the giants](#)'.

Copyright © UNU-WIDER 2018

Information and requests: [publications@wider.unu.edu](mailto:publications@wider.unu.edu)

ISSN 1798-7237 ISBN 978-92-9256-629-6 <https://doi.org/10.35188/UNU-WIDER/2018/629-6>

Typescript prepared by Ayesha Chari.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

## 1 Introduction

India ranks among the world's largest as well as most densely populated countries. While labour is abundant in India, skilled workers continue to be in short supply despite considerable investments aimed at improving the skill level of the labour force. Despite the shortage of skilled labour, growth in real gross domestic product (GDP) has averaged over 7 per cent over the last 20 years. The question is whether the present growth model is sustainable.

Inequality too has been increasing in India. Economic growth is leaving behind low- and middle-income households, which may have consequences for human capital accumulation: 'high levels of inequality impede sustained development and growth through numerous channels' (World Bank 2018: 6). Human capital accumulation arguably denotes one of these channels and may well be the most important channel. 'Inclusion is therefore needed not just to address the wide inequalities in India, but to ensure that growth can be sustained over decades' (World Bank 2018: 2).

This highlights the fact that productivity and socio-economic progress are interconnected. Economic growth helps fund policies that bring about socio-economic progress. Equally important, socio-economic progress serves as an engine for productivity growth. When socio-economic mobility (or equality of opportunity) is low, perfectly talented individuals may not be given the opportunity to reach their full potential due to circumstances that are entirely beyond their control (such as the family and neighbourhood they are born into). This leads to a waste of human potential. Public policies that aim to level the playing field will arguably raise the aggregate stock of human capital and thereby stimulate economic growth. Furthermore, since the waste of human potential tends to be concentrated towards the bottom of the income distribution, the growth brought about by these public interventions will in all likelihood be of an inclusive nature.

Public investments in education and health have helped improve education outcomes, including for individuals who are socio-economically disadvantaged. Public policies vary at the state level, and the degree of success has arguably been uneven. There continues to be a significant variation in education outcomes between states and presumably also within states. Important gaps in outcomes between individuals from different socio-economic backgrounds still remain: 'The task of upgrading India's human capital is therefore urgent to ensure that a more productive workforce is available to meet the demand for labor in globally competitive industries' (World Bank 2018: 8).

Does the lack of human capital accumulation in part stem from low levels of socio-economic mobility in India? If indeed, how does this impact on the income growth prospects for low-, middle-, and upper-income households? In a recent global study, Narayan et al. (2018) identify India as a country with some of the lowest rates of intergenerational mobility in the world. Given the size of the country, however, it is conceivable that economic mobility exhibits considerable within-country variation. Is there a land of opportunity in India? Where are the most and least economically mobile municipalities located? And what municipalities have made the most (and least) progress over time?

The objective of this study is twofold. First, we build a database for India that tracks socio-economic mobility and human capital accumulation at the sub-national (say district) level over the last 30 years, and examine the trends and patterns in mobility and human capital. Drawing from a variety of different data sources, the following variables are also added to this district level panel dataset: (a) household expenditure growth for the low-, middle-, and upper-income class, (b) inequality in household expenditure per capita, (c) demographics, (d) employment variables, (e) domestic infrastructure connectivity (capturing domestic market integration), (f) financial

inclusion, and (g) selected political variables (i.e. voter turnout, political competition). Second, we use these data to address the questions listed in the preceding paragraph. Specifically, we will verify whether intergenerational mobility impacts on future growth, affects the ‘inclusivity’ of this growth, and whether human capital accumulation denotes a plausible channel.

Our primary source of data is the National Sample Surveys (NSS) of India for the years 1983, 1987, 1993, 1999, 2004, and 2011. This allows us to track socio-economic progress for roughly 60 state-regions over time. Specifically, we use the state-regions constructed by Lanjouw and Murgai (2009) that accounts for merging and splitting of the sub-national units over this extended time period.

The remainder of the paper is organized as follows. Sections 2 and 3 introduce the measures of intergenerational mobility and the sources of data that we will be working with, respectively. Section 4 presents an overview of the trends and patterns in intergenerational mobility, inequality, and growth in India. The main results that document the relationship between intergenerational mobility, human capital accumulation, and inclusive growth are presented in Section 5. Finally, Section 6 concludes.

## 2 Measures of intergenerational mobility

Socio-economic mobility has been interpreted in several ways, both in the economic and in the sociological literature. This paper focuses on measures of relative mobility between generations in terms of educational attainment. Relative intergenerational mobility is the extent to which an individual’s position on the socio-economic scale is independent of the position of his or her parents. Socio-economic mobility defined this way is consistent with the interpretation of mobility as ‘origin independence’ applied to an intergenerational context. A society with high relative intergenerational mobility is one where an individual’s achievements is influenced less by the individual’s origin (his or her parents) and more by his or her human potential. Origin independence is closely related to the idea of equality of opportunity (Roemer 1998).

Three different measures of mobility will be considered: (a) the expected rank of a child (in the child education distribution) whose parents are in the bottom 50 per cent of the parent education rank distribution [a measure that is advocated by Asher et al. (2018)], (b) the statistical correlation between the years of schooling of parents and the years of schooling of their children (Van der Weide and Milanovic 2018), and (c) the share of inequality in years of schooling that is due to differences in parental education background (i.e. the share of ‘between inequality’ over total inequality). Ideally, a subgroup decomposable inequality measure is used, such as the ‘mean log deviation’ (or other members of the general entropy family). The latter measure is borrowed from the literature on inequality of opportunity (Ferreira and Gignoux 2011), and would potentially allow one to expand the set of circumstances that are beyond the control of the individual.

The first of these three measures will serve as our preferred measure [see Asher et al. (2018) for a discussion]. We will be referring to this measure as ‘upward mobility’.

Focusing on intergenerational mobility in education has pros and cons. One advantage is that it does not matter at what age we measure an adult’s level of education; after completing your education it stays with you for life. The same is not true for income, for example, which varies across a person’s life-cycle. Unlike income, however, education data tend to be relatively coarse. This has implications for identifying the individuals whose parents are in the bottom half of the parent education distribution. This requires that one is able to unambiguously rank parents from

the 1st to the 99th percentile. But if 60 per cent of parents do not have any education for example, then it is not clear what parents define the bottom half exactly. One pragmatic solution is to randomly break ties. If all individuals with uneducated parents have identical odds of completing a given level of education, then this pragmatic decision should have no bearing on our estimate of upward mobility. This decision is less innocent, however, if there is some latent ordering in the human potential among these individuals. Under this assumption, Asher et al. (2018) derive upper and lower bounds for upward mobility that can be computed on the basis of the observed coarse education data. The larger the margin by which the share of least-educated parents exceed 50 per cent, the larger the degree of ambiguity, and the larger will be the bounds on upward mobility. This logical prediction is confirmed by our data.

We will use the measure that randomly breaks ties between parents as a point of departure, and then apply the smallest possible correction that forces it within the bounds. In effect this means replacing the measure of upward mobility with the upper bound if it otherwise exceeds the upper bound, and replacing it with the lower bound if it otherwise falls below the lower bound. A graphical illustration of this correction is presented in the next section.

### **3 Data**

The primary data on education and living standards come from thick rounds of the Household Consumer Expenditure schedule of the National Sample Survey (NSS) of India. We use the NSS rounds collected in 1983, 1987–88, 1993–94, 1999–2000, 2004–05, and 2011–12. The 1983 survey round was carried out between January and December, while in the other rounds data was collected from July to June the following year (NSSO 1983, 1988, 1994, 2000, 2005, 2012). These surveys also provide a rich set of socio-economic variables including the industry and occupation type of the primary economic activity of households.

Each NSS round constitutes a population representative survey for India with over 100,000 interviewed households. Because we are interested in intergenerational mobility and the distribution of living standards and education, our unit of observation is at the state-region level, which is an administrative unit between districts and states. The definition of states and regions changed during the study period due to the creation of new states. We follow Lanjouw and Murgai (2009) in harmonizing the state-regions to the state and region definitions used in the 1999–2000 NSS round, which included the newly created states of Jharkhand, Chhattisgarh, and Uttaranchal. Some of the resulting 77 state-regions have a sample size of less than 1,000 households. We merge these following geographic considerations to arrive at 54 state-region units. Appendix A1 describes the construction of the unit of observations in detail. Hence, our core sample includes 324 observations.

In the following, the construction of the variables used in the analysis is described in detail. For some of the variables describing the economic and political conditions during the survey period, we complement the information from NSS with a number of other data sources. This is indicated in the descriptions of each of the following sub-sections.

### 3.1 Consumption expenditure

We capture the changes in living standards using the monthly per capita consumption expenditure (MPCE) of the households. We use the MPCE calculated using mixed reference periods following the recommendation of the Tendulkar Committee. To account for spatial and temporal differences in consumption expenditure, we calculate real consumption expenditure at 2004 rural all-India prices using a series of consumer price indices.<sup>1</sup>

Following established practice, we use the consumer price index for agricultural labourers (CPI-AL) for rural areas, and the consumer price index for industrial workers (CPI-IW) for rural areas. These indices are available for the 16 major states in India.<sup>2</sup> Values for the smaller states are imputed based on spatial proximity. Additionally, we use the poverty lines of 2004–05 to harmonize price levels between states and urban–rural sectors (see Appendix A2 for further details).

Using the resulting real per capita consumption expenditure, we calculate different percentiles (10, 25, 50, 75, 90, and 95 per cent) at each state-region and survey year using population weights from the NSS. Annualized consumption expenditure growth is then calculated for each of these percentiles, taking into account the different intervals between the NSS rounds used in the study.

To measure inequality in consumption expenditure, we use the mean logarithmic deviation (MLD). For a consumption distribution  $y$ , MLD is defined as

$$MLD(y) = \ln(\sum_{i=1}^n w_i y_i) - \sum_{i=1}^n w_i \ln(y_i) \quad (1)$$

where  $w_i$  denotes the population weight of individuals in the sample with  $\sum_{i=1}^n w_i = 1$ , and  $n$  is the sample size. We calculate MLD at each state-region and survey year.

### 3.2 Human capital

The main variable of interest is intergenerational mobility in education. Therefore, we focus on the educational attainment of boys in the age group 20–25 years, who still reside with their father. This group is old enough to have mostly completed their education. Figure 1a shows that about 19 per cent of boys were still enrolled in education at the age of 20 years in 1999 (13 per cent in 1983). At the same time, this group is young enough to have a significant share still living with their parents. According to Figure 1b, 73 and 82 per cent of 20-year-old boys and 49 and 56 per cent of 25-year-old boys still lived at home in 1983 and 1999, respectively. The figure also reveals the gap between boys and girls in terms of enrolment in education and home-leaving age. As a result, our data do not allow us to draw inference on the intergenerational mobility of girls.

We measure educational attainment as years of schooling. In the NSS, information on the general education level was collected for all household members, which we used to calculate the years of education using the following conversion rates: 0.1 years for illiterates, 0.5 years for literates without formal schooling, 2.5 years for literates below primary level, 5 years for primary school, 8

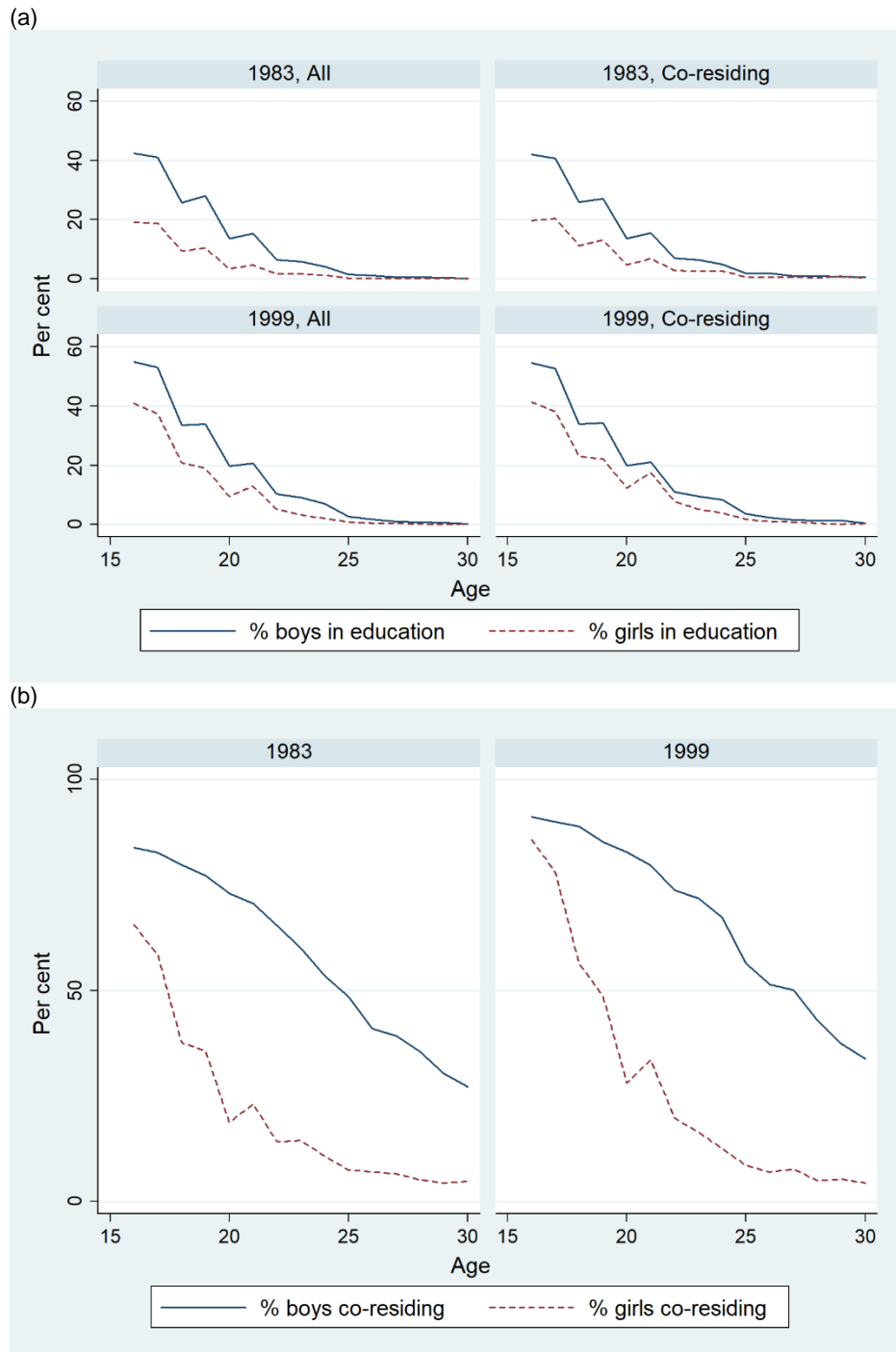
---

<sup>1</sup> We observe that between 1983 and 2011 prices increased almost eightfold.

<sup>2</sup> These 16 states are Andhra Pradesh, Assam, Bihar, Gujarat, Haryana, Himanchal Pradesh, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Orissa, Punjab, Rajasthan, Tamil Nadu, Uttar Pradesh, and West Bengal.

years for middle school, 12 years for secondary and higher secondary graduates, and 16 years for graduates and above.<sup>3</sup>

Figure 1: (a) School enrolment and (b) co-residence of young adults aged 20–25 years



Source: Authors' compilation based on data from NSSO (1983, 2000).

<sup>3</sup> For the 1983 round, we use the individual-level data from the Employment and Unemployment schedule of the NSS due to data availability (see NSSO 1983). The Employment schedule is collected from different households than the Consumption expenditure schedule. Both schedules are population representative, and we use data aggregated at the state-region level in our analysis.

For the educational attainment of fathers, we distinguish between five categories: illiterates, below primary level, primary school, middle school, and secondary and higher education, which are derived from the above-mentioned seven categories.

In the analysis of human capital accumulation, we condition the son's years of education on the father's level of education. Hence, we calculate changes in the years of education of boys by their father's education level.

### 3.3 Intergenerational mobility

We calculate different intergenerational educational mobility measures for boys in the age group 20–25 years, who still reside with their father. Our main measure of intergenerational mobility is upward mobility: the expected outcome of a son born into the bottom half of the parent outcome distribution. This rank-based approach has the advantage that is comparable across time. However, the coarse measure of educational attainment makes it difficult to provide a point estimate (Asher et al. 2018). We follow the indicator proposed by Asher et al. (2018), which overcomes this problem by putting bounds on the upward mobility measure. We refer to these as Asher–Novosad–Rafkin (ANR) bounds.<sup>4</sup>

One of the requirements of upward mobility measures is the monotonicity of the rank distribution. For some of the lower education categories, we find that monotonicity is violated in some of the state-regions. Therefore, we merge the lower three categories, and use the following five education categories when calculating upward mobility: below primary, primary, middle school, secondary and higher secondary school, and graduates and above. In addition, we calculate the ANR bounds only when at least 50 son–father pairs are available in the state-region and year. This condition is satisfied for all but one observation (state-region 4 in 1983). We impute the values from the 1987 survey for this observation.

The ANR bounds are less precise when a substantial proportion of the population falls into the lowest education level. In the context of India, especially in the 1980s and 1990s, more than half of the fathers were still illiterate in many state-regions. Therefore, we also calculate point estimates of upward mobility using a random tie-breaking algorithm to assign unique ranks to both fathers and sons. We use this point estimate for state-regions and years where the estimate falls into the ANR bounds, and apply the smallest possible modification that forces it within the ANR bounds. Figure 2 demonstrates the construction of our upward mobility measure for one of the state-regions.

Another measure of intergenerational educational persistence is the share of education inequality explained by the father's education level. We use the son's years of education (as described) to calculate total inequality and inequality by the father's level of education. For the father's education level we use the following five categories: illiterate, below primary, primary school, middle school, and secondary and higher education. These categories were selected with the consideration of having a reasonably balanced share of fathers in each category.

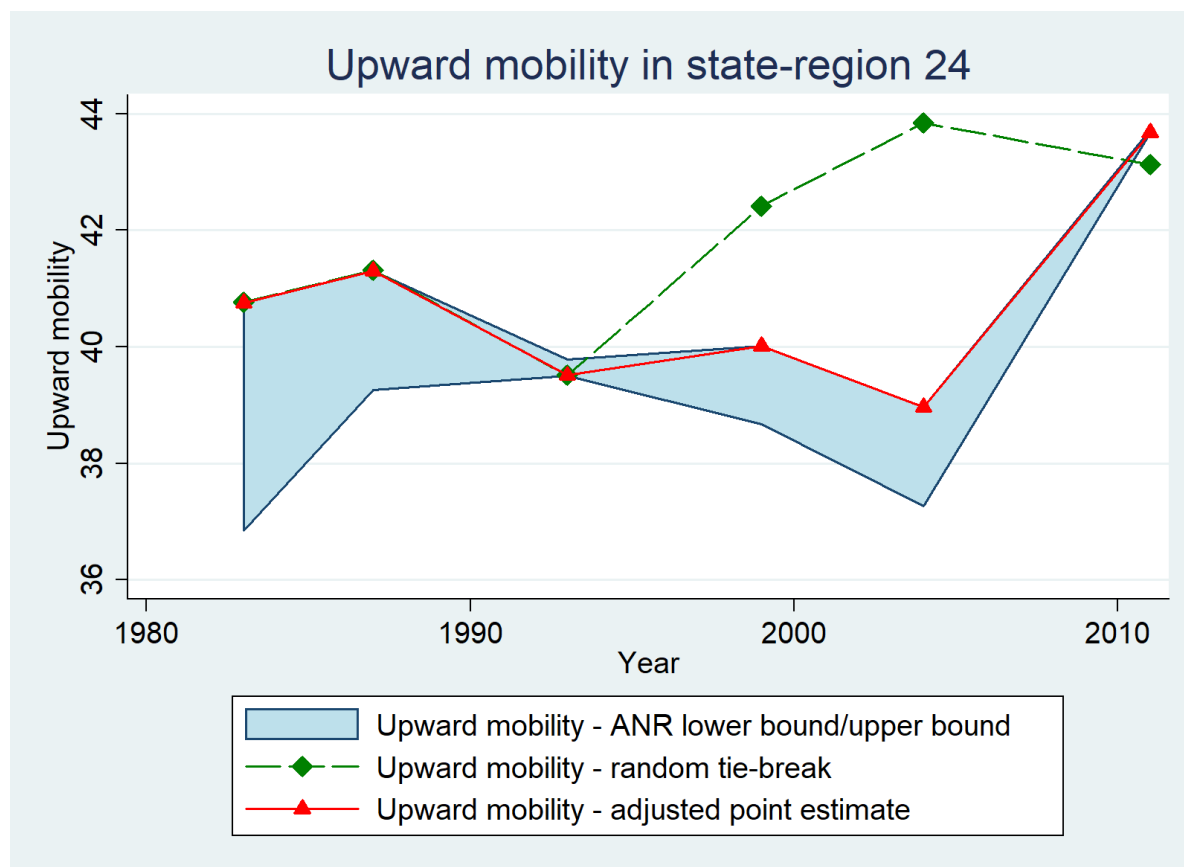
When constructing the correlation coefficient between the father's and son's years of education, we use the years of education as described earlier. In addition, we calculate the correlation coefficient only when at least ten son–father pairs are available in the state-region and year.

---

<sup>4</sup> We are grateful to the authors who provided us with their code to calculate their proposed upward mobility measure.



Figure 2: Construction of the upward mobility measure



Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012).

### 3.4 Demographics

Information on the share of working age population (ages 16–64 years), the share of population adhering to different religions, and the share of urban population comes from the NSS. Urban–rural classification is from the NSSO.<sup>5</sup>

Historical data on the share of Brahman (in 1909 and 1931) and scheduled tribes and scheduled castes (between 1951 and 2001) in the population are provided by Castelló-Climent et al. (2017). The dataset contains information at the district level for 500 districts in 20 states. We match the districts to the state-regions following the definition of regions and their compositions according to NSSO (2000), and aggregate the variables to the state-region level.

### 3.5 Economic conditions

Variables on economic conditions come from a number of different data sources. Information on the share of agriculture, manufacturing, and services sector is calculated using the industry and occupation classification of the main household activity from the NSS. Information was provided using three digit industry and occupation codes following the National Industrial Classification and National Classification of Occupations. Changes in these registers were harmonized at the

<sup>5</sup> The NSS has followed the census definition of urban areas, which is based on several criteria, including a population of at least 5,000, a density of at least 4,000 people per square kilometre, and three-fourths of the male workers engaged in non-agricultural activities (Datt and Ravallion 2011).

two-digit level, and households were grouped into one-digit industry and occupation categories, which are detailed in Appendix A3.

We calculate access to electricity using information on the lighting source in the NSS. We define access as having at least one household in the primary sampling unit (consisting of about ten households) that uses electricity as the source of light.

Data on access to banking come from statistical tables published by the Reserve Bank of India (see RBI 1998, 2000, 2005, 2012). Information is available on the number of bank offices, officers, deposits, and credit for all the NSS years at the state level for all states except for the newly created states of Jharkhand, Chhattisgarh, and Uttaranchal. For these states, we use the values from the states where they were carved out of. In the analysis, we use data on the number of bank offices per capita. In order to calculate this variable, we use state-level population data from the census of 1991, 2001, and 2011. We use log-linear extrapolation/interpolation to calculate the population of states for the survey years.

We follow Allen and Atkin (2016) to construct a measure of market access using the (inverse) travel-time weighted size of market that can be accessed from a location. We proxy the size of market using population size. Allen and Atkin (2016) provide data over the bilateral highway travel times between districts over time for 308 districts in 19 states. District-level population data are provided by the ICRISAT Village Dynamics in South Asia Macro-Meso Database (VDSA), which can be matched to the travel times data for all the NSS years. Using these sources, we calculate market access at the district level as

$$MA_{it} = \sum_{j \neq i} \left( \frac{1}{\text{traveltime}_{ijt}} \right) P_{jt} \quad (2)$$

where  $P_{jt}$  is the population of district  $j$  in period  $t$ . For our analysis, we aggregate the resulting district-level variable to the state-region level.

### 3.6 Politics and public finance

In order to control for public policy during the survey period, we use data on the total state government expenditure and state expenditure on education as a share of the state's GDP. Information on public finance variables is available at the Economic Organisation and Public Policy Programme's (EOPP) Indian States Database compiled by the Suntory and Toyota International Centre for Economics and Related Disciplines (STICERD) at the London School of Economics using the Public Finance Statistics published by the Ministry of Finance. It covers annual public expenditure data in 16 states of India for the period of 1957–2000. Data on state-wise GDP come from the Ministry of Statistics and Program Implementation of India for the period between 1980 and 1993, and RBI's Handbook of Statistics on Indian States for the period between 1990 and 2015.

We measure the political climate using voter turnout and political competition. The latter is measured as the difference between the vote share of the largest and second largest parties following Besley and Burgess (2002). These data are also available in the EOPP Indian States Database compiled by STICERD using statistics from Butler et al. (1995) and the Electoral Commission of India, covering voter turnout and the number of seats won by parties in 16 states of India for the period of 1949–2002.

For all the variables discussed in this section, Table 1 summarizes the annual means and the total sample size.

Table 1: Mean of variables by year and sample size

	1983	1987	1993	1999	2004	2011	N observations
Log real consumption expenditure							
10th percentile	5.48	5.60	5.64	5.78	5.79	6.01	324
25th percentile	5.76	5.83	5.86	6.00	6.01	6.22	324
50th percentile	6.07	6.11	6.12	6.26	6.27	6.49	324
75th percentile	6.42	6.41	6.40	6.56	6.59	6.81	324
90th percentile	6.78	6.73	6.71	6.88	6.94	7.17	324
95th percentile	7.03	6.95	6.91	7.10	7.19	7.42	324
Education years							
Boys, father illiterate	3.46	3.68	4.68	5.39	6.05	7.18	324
Boys, father below primary school	6.18	6.36	7.28	7.33	8.01	8.96	324
Boys, father primary school	7.86	7.76	8.54	8.55	8.58	9.70	324
Boys, father middle school	9.45	9.35	9.93	9.96	10.02	11.16	324
Boys, father high school and above	11.83	11.92	11.96	12.20	12.23	12.80	324
Share							
Boys illiterate	0.29	0.26	0.20	0.15	0.11	0.06	324
Boys below high school	0.49	0.46	0.44	0.45	0.47	0.38	324
Boys high school and above	0.23	0.27	0.36	0.40	0.42	0.56	324
Father illiterate	0.52	0.50	0.46	0.42	0.39	0.36	324
Father below primary school	0.41	0.41	0.41	0.41	0.43	0.42	324
Father high school and above	0.07	0.10	0.13	0.17	0.19	0.23	324
IGM boys							
Upward mobility	41.39	40.91	40.43	39.57	38.37	38.55	324
Correlation coefficient	0.55	0.55	0.53	0.54	0.52	0.50	324
Share of father's education in MLD	0.12	0.11	0.10	0.10	0.11	0.13	324
Inequality in years of education for boys	1.00	0.96	0.78	0.62	0.45	0.29	324
Inequality in real consumption expenditure	0.16	0.12	0.11	0.12	0.13	0.14	324
Share of age 16–64 years	0.55	0.56	0.58	0.59	0.61	0.64	324
Share of urban	0.25	0.24	0.25	0.25	0.26	0.29	324
Share of Christians	0.06	0.05	0.06	0.06	0.06	0.06	324
Education years: population	3.35	3.70	4.39	5.07	5.78	6.88	324
Services industry	0.22	0.26	0.26	0.28	0.29	0.33	324
Manufacturing industry	0.15	0.17	0.15	0.18	0.20	0.24	324
Access to electricity	0.58	0.67	0.79	0.84	0.80	0.88	324
Log bank per capita	-9.72	-9.53	-9.64	-9.59	-9.65	-9.48	324
Market access	10.11	10.23	10.40	10.59	10.78	10.95	276
Government expenditure per GDP	0.14	0.17	0.17	0.17	0.20	0.17	246
Education expenditure per GDP	0.03	0.03	0.03	0.04	0.04	0.03	246
Voter turnout last election	57.95	62.30	62.25	66.09	63.61	—	205
Political competition	-0.47	-0.44	-0.34	-0.32	-0.25	—	205
Share of Brahman, 1931	0.05	0.05	0.05	0.05	0.05	0.05	288
Share of scheduled tribes, 1961	0.08	0.08	0.08	0.08	0.08	0.08	288

Notes: MLD, mean logarithmic deviation; GDP, gross domestic product. Population weighted averages of state-regions.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005 and 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Clement et al. (2017), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and STICERD and LSE (2005–19).

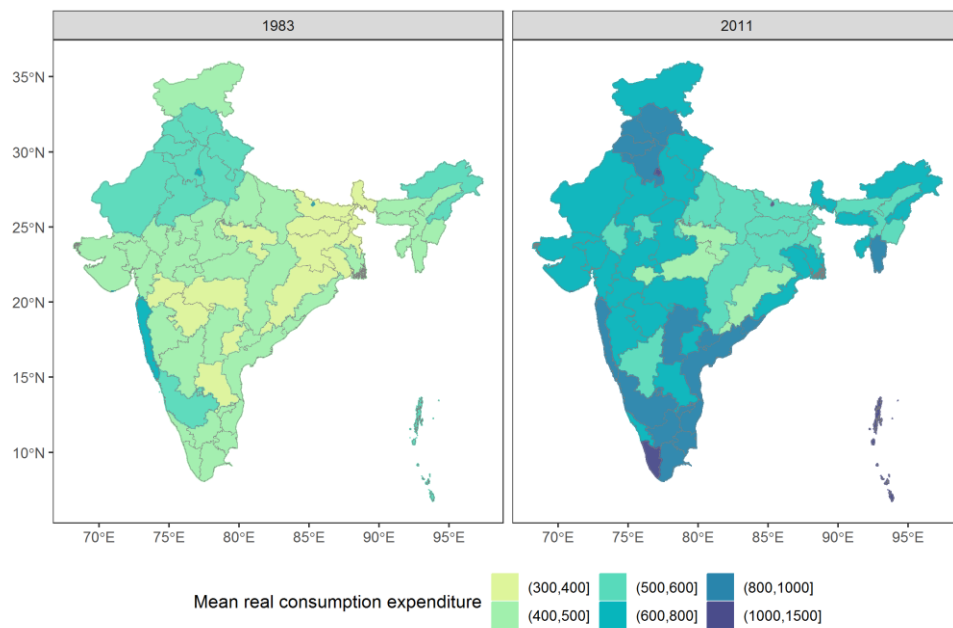
#### 4 A first look at the data

Between 1983 and 2011 the per capita GDP in India grew by an average of 4.3 per cent. At the same time, the average annual growth rate of per capita real consumption expenditure was 1.5 per cent. Figure 3a reveals signs of the structural changes that the Indian economy underwent during

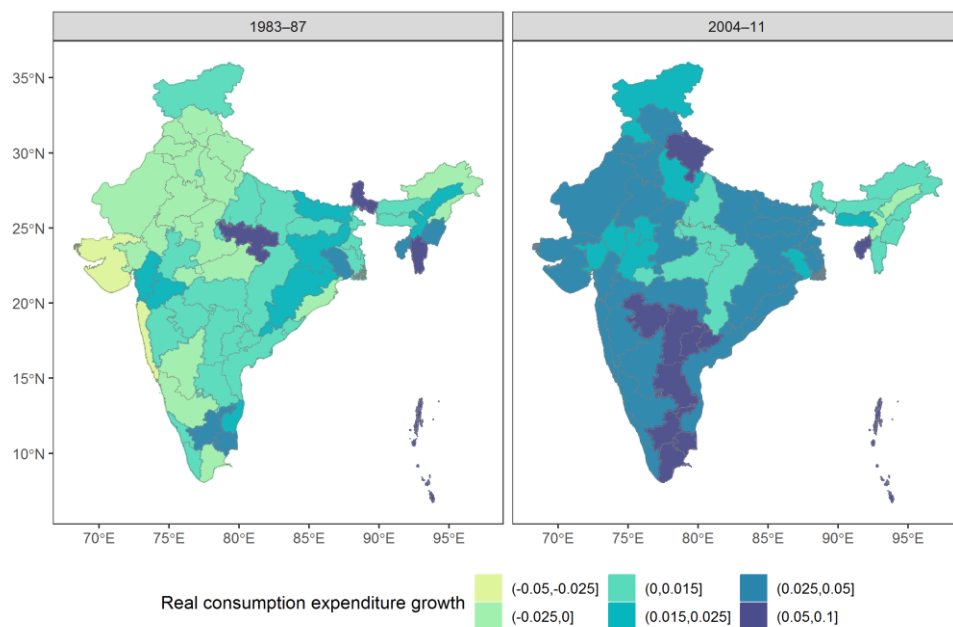
the three decades. In 1983, the poorest state-regions especially in the north-east showed the strongest economic performance as indicated in Figure 3b. By 2011, the pattern reversed and economic growth was concentrated in the relatively better-off areas, especially in the southern zone and also in the western and northern zones. The figure also highlights that there was substantial variation in the growth rates among the state-regions in 1983 and 2011.

Figure 3: (a) Mean monthly per capita real consumption expenditure and (b) its annualized growth between 1983 and 2011 in the survey of state-regions

(a)



(b)

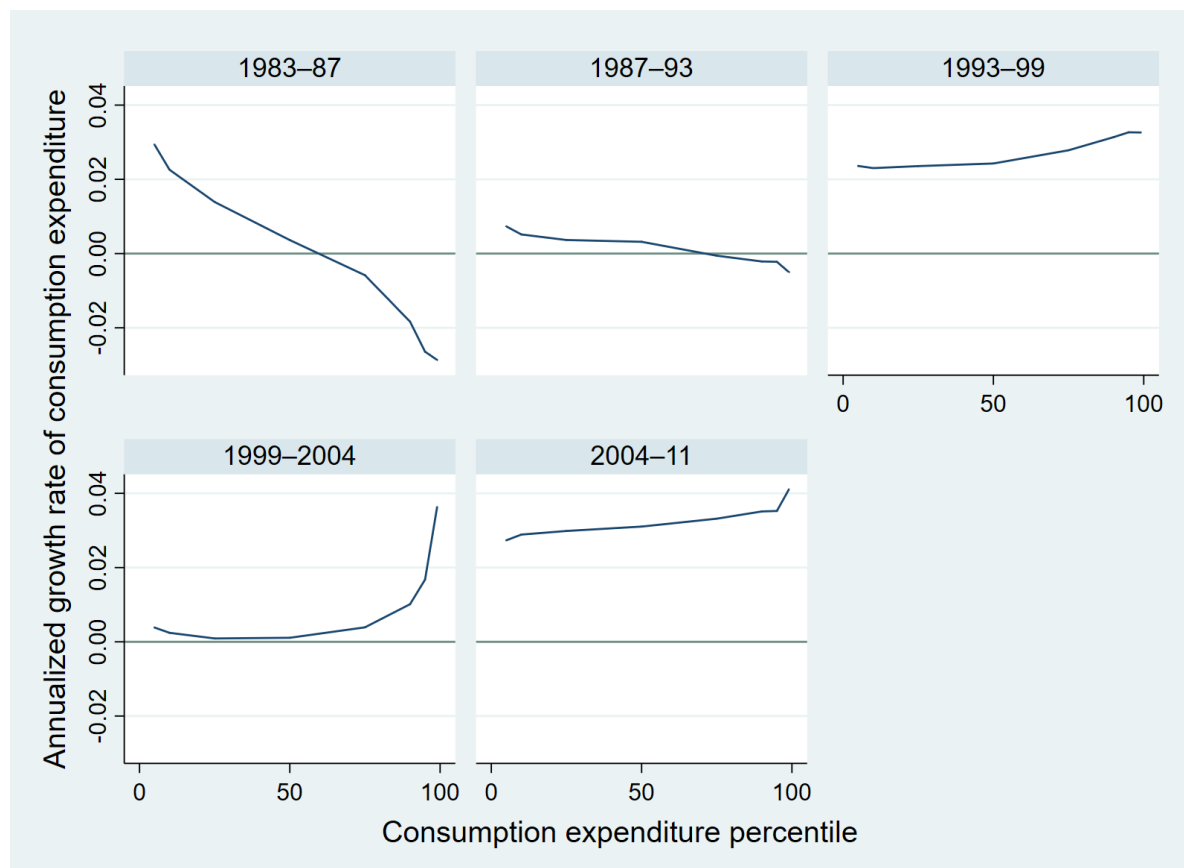


Note: Consumption expenditure is reported in 2004 rural all-India prices.

Source: Authors' compilation based on data from NSSO (1983, 1988, 2005, 2012) and Government of India (2012-15a, 2012-15b).

Until the early 1990s, economic growth was mainly driven by the agricultural sector as a result of the green revolution. This had a poverty reducing effect (see Figure 5) and reached rural areas in particular while also benefiting the urban poor (Datt and Ravallion 2011). Figure 4 reveals that consumption expenditure growth was indeed the highest among the poorest (lowest consumption percentiles) between 1983 and 1987. The macroeconomic crisis that hit the Indian economy in the early 1990s is also visible in the figure for the period 1987–93, when consumption expenditure growth was close to zero for all consumption percentiles.

Figure 4: Real consumption expenditure growth by percentile, 1983–2011

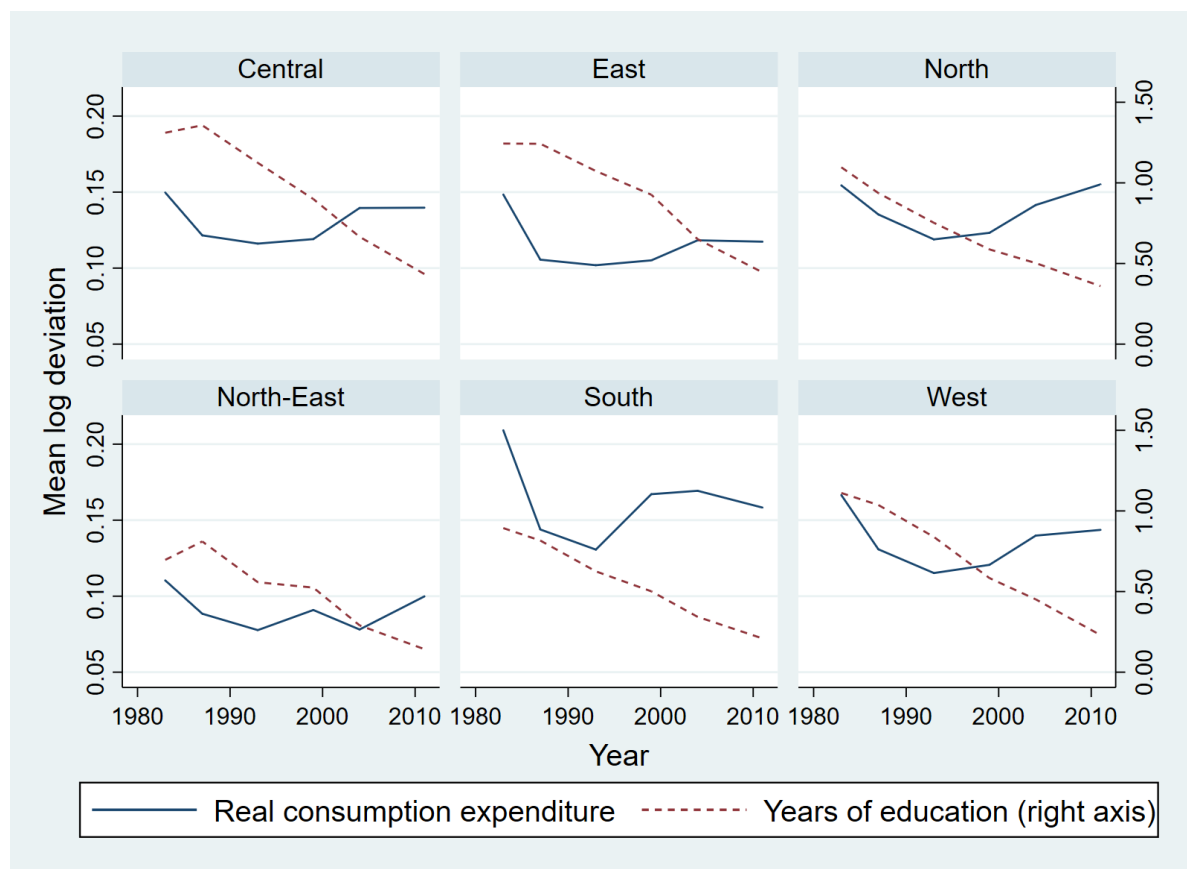


Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012) and Government of India (2012–15a, 2012–15b).

The crisis triggered economic reforms to restart the Indian economy and to control inflation and the balance of payments deficit. The reforms focused on liberalization of trade (by reducing import and export taxes) and industrial activities (by abolishing licensing and through privatization), and encouraged foreign direct investment into high technology and high-investment sectors. As a result, the Indian economy boomed in the past decades.<sup>6</sup> Economic growth has been mainly driven by the manufacturing and construction industries, hence the economic opportunities shifted to urban areas and mostly bypassed many of India's poor (Datt and Ravallion 2011). As a consequence, after the 1990s consumption growth was higher among the higher expenditure percentiles compared with the lower percentiles (Figure 4) leading to growing consumption inequality (Figure 5) both within and between state-regions.

<sup>6</sup> The high consumption expenditure growth in the period 1993–99 may be overestimated due to changes in the consumption expenditure survey instrument in the 1999 NSS. For a detailed discussion, see Deaton (2003).

Figure 5: Inequality in consumption expenditure and years of schooling



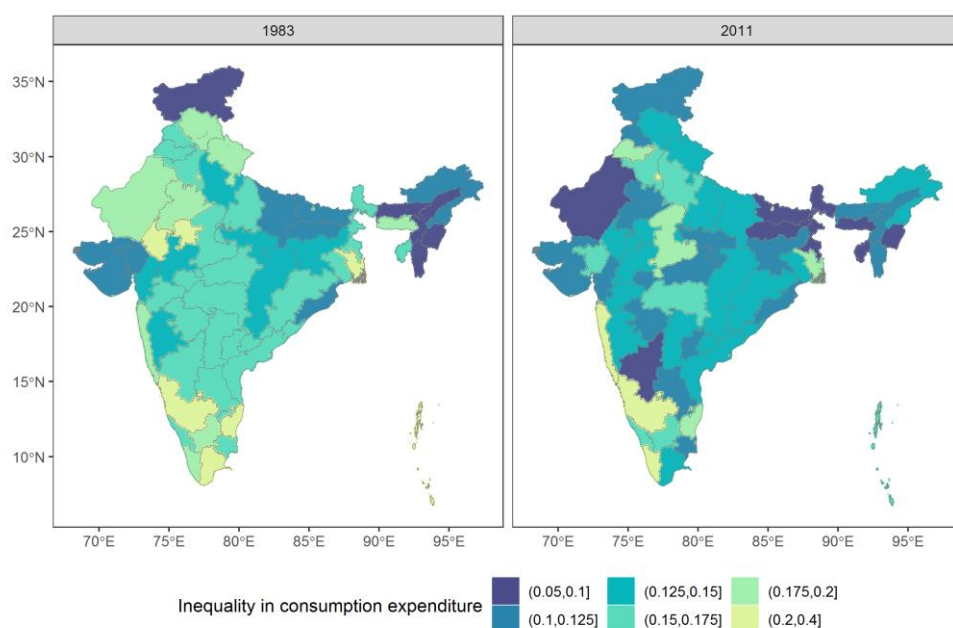
Notes: Inequality is measured by mean log deviation. Sample for years of schooling includes boys between 20 and 25 years living in the same household as their father.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012) and Government of India (2012–15a, 2012–15b).

As discussed, Figure 5 shows that in all zones of India consumption expenditure inequality (solid line) decreased in the 1980s and early 1990s, but increased in the late 1990s and 2000s. Both Figures 5 and 6 show that consumption expenditure inequality nonetheless improved somewhat between 1983 and 2011 in many state-regions. In Figure 6, darker areas are associated with lower consumption inequality. We can observe a substantial variation in inequality among the state-regions, with the southern zone displaying the highest inequality and the north-eastern zone having the lowest inequality.

At the same time, inequality in education reduced substantially between 1983 and 2011 as shown in Figure 5 (dashed line). This is a very welcome achievement, as still around one in three boys between 20 and 25 years of age was illiterate in 1983. The literacy rate and educational attainment increased considerably over the past decades in large part due to affirmative action programmes targeting the scheduled castes, scheduled tribes, and other backward castes (Desai and Kulkarni 2008; Azam and Bhatt 2015).

Figure 6: Inequality in per capita real consumption expenditure in the survey state-regions in 1983 and 2011



Note: Inequality is measured by mean log deviation.

Source: Authors' compilation based on data from NSSO (1983, 2012) and Government of India (2012–15a, 2012–15b).

The stratification of Indian society by castes, religion, and ethnic boundaries is reflected in inequalities of educational attainment of the different groups (Desai and Kulkarni 2008). Affirmative action policies were put in place in the form of quotas for political representation, public employment, and (higher) education since independence in 1947. In addition, a number of schemes were introduced to promote education among the above-mentioned disadvantaged groups in the form of midday meals, scholarships, free secondary schooling, uniforms, and books (Desai and Kulkarni 2008; Cassan 2019). Cassan (2019) finds that these measures contributed to a 10 percentage point increase in literacy and a 7 percentage point increase in secondary school attainment driven exclusively by the increased educational attainment of boys. Despite these improvements, our data show that, in 2011, 10 per cent of boys in the age group 20–25 years remained illiterate, and the rate was 21 per cent among girls of the same age.

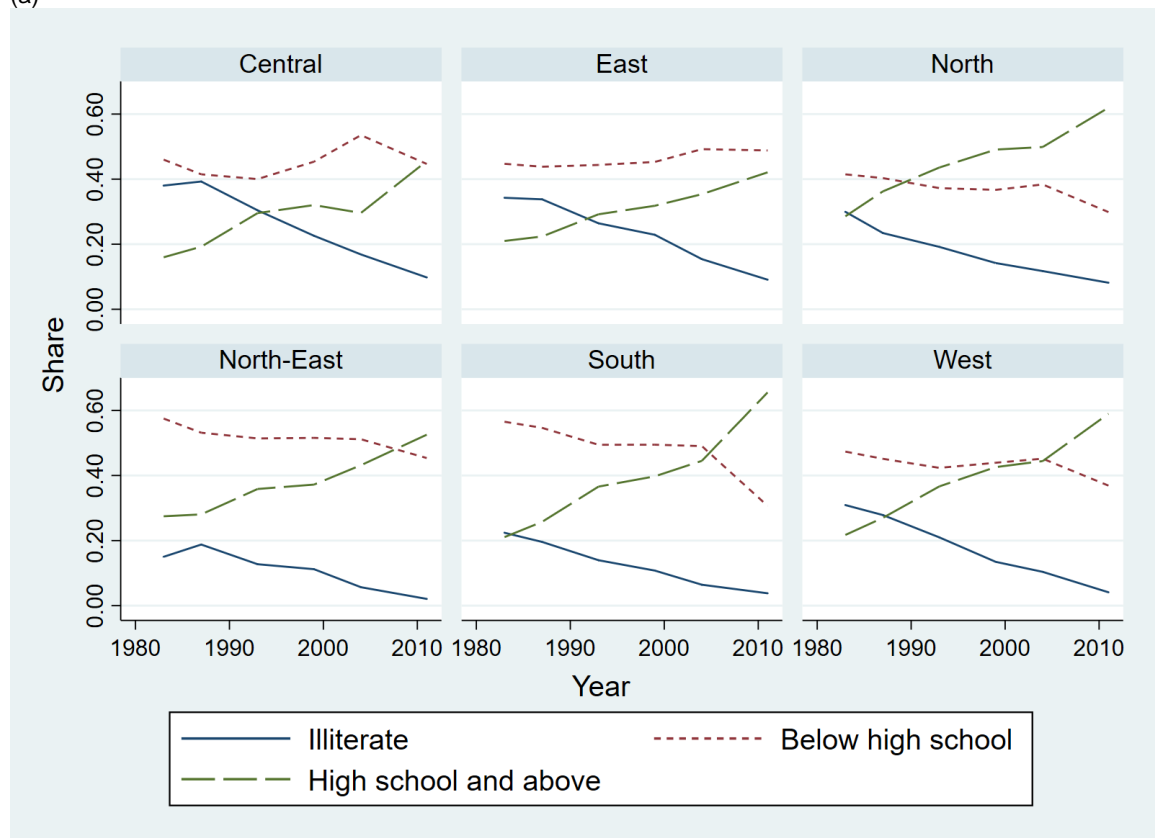
We focus on the human capital accumulation of boys aged 20–25 years who still live with their father. In line with Asher et al. (2018), we find a positive bias (between 0.35 and 0.58 years of schooling or between 4 and 9 per cent) for the educational attainment of co-resident boys compared with all boys in the same age group.

Figure 7 shows the distribution of the education level for boys (a) and their father (b) in the six zones of India. The most pronounced development is the increase in the proportion of boys with at least high school education from 23 per cent in 1983 to 55 per cent in 2011. As the figure shows, the central and eastern zones lag behind somewhat in the share of high school graduates as a result of a larger share of boys quitting school before high school.

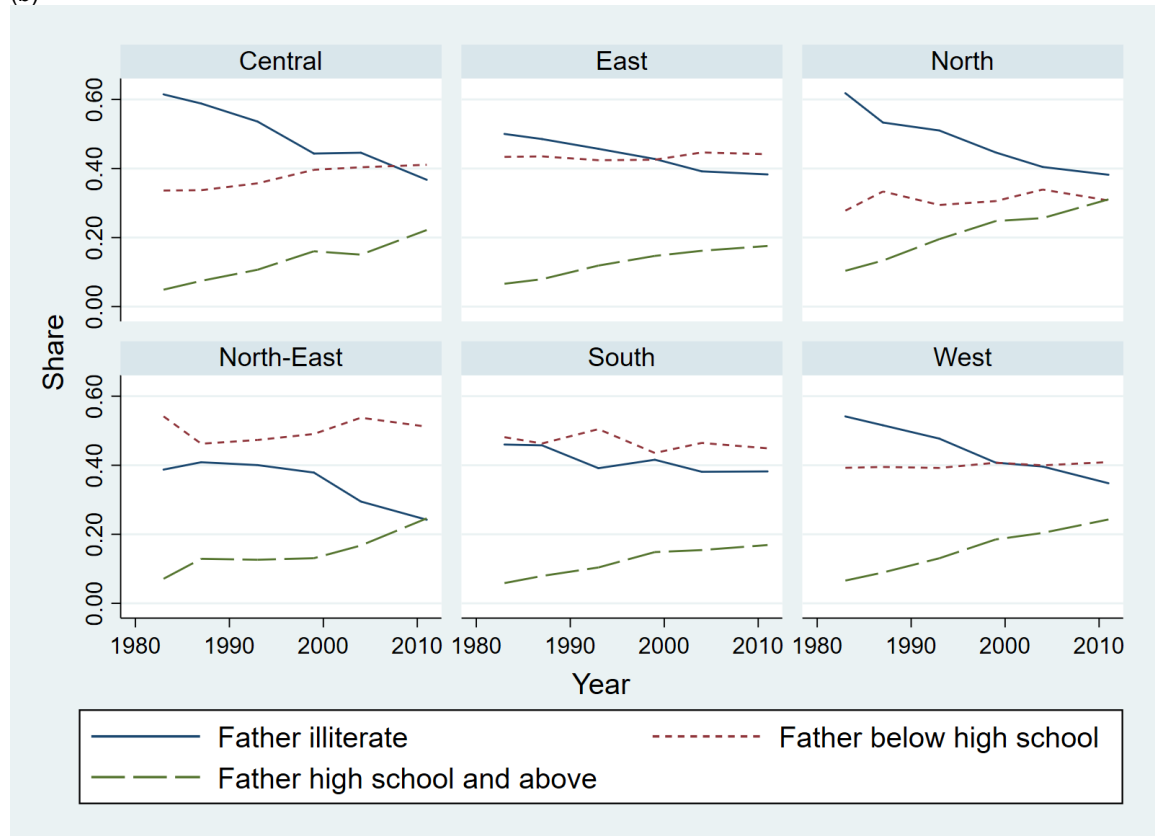
The increase in father's education follows behind the boy's with a generation lag, and we also observe a significant reduction in the illiteracy rate for this group from 54 per cent in 1983 to 39 per cent in 2011. Again, we also observe regional variation in literacy rates and their rate of change, with the north-eastern and southern zones showing the highest literacy rates and the central, northern, and western zones booking the largest gains in literacy between 1983 and 2011.

Figure 7: Level of education for (a) boys aged 20–25 years and (b) their father

(a)



(b)



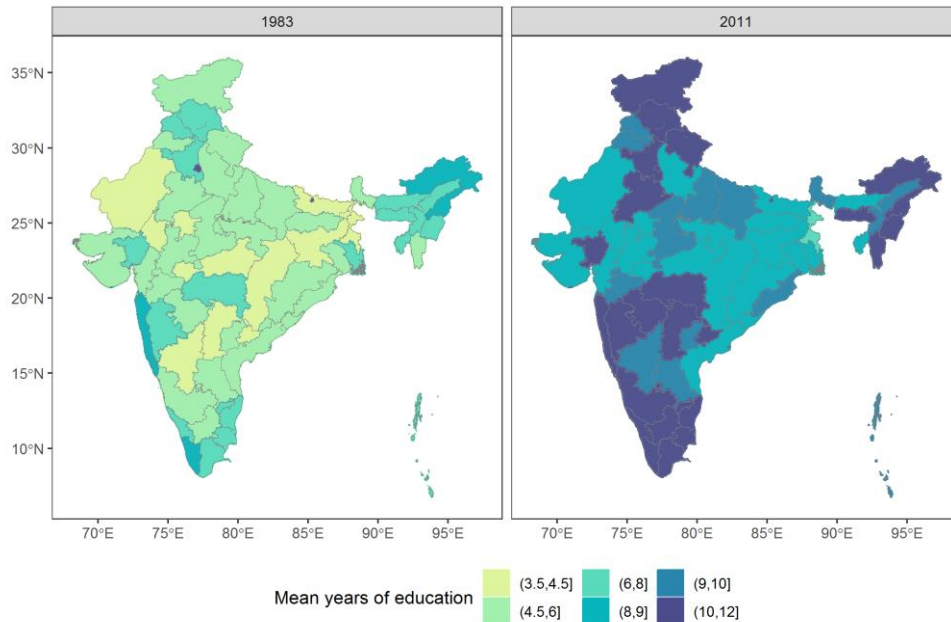
Note: Sample includes boys between 20 and 25 years and their father who live in the same household.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012).



In the analysis, we measure human capital using years of education. Figure 8 shows the distribution of the average years of education attained by boys in the above-described study cohort. In 1983, the average years of education ranged between 3.5 and 8.9 years (excluding Delhi), with a mean of 5.6 years. By 2011, it increased to between 7 and 11.9 years, with a mean of 9.6 years, which is a sizeable improvement as the darkening of the map also indicates.

Figure 8: Mean years of education for boys aged 20–25 years in the survey state-regions in 1983 and 2011



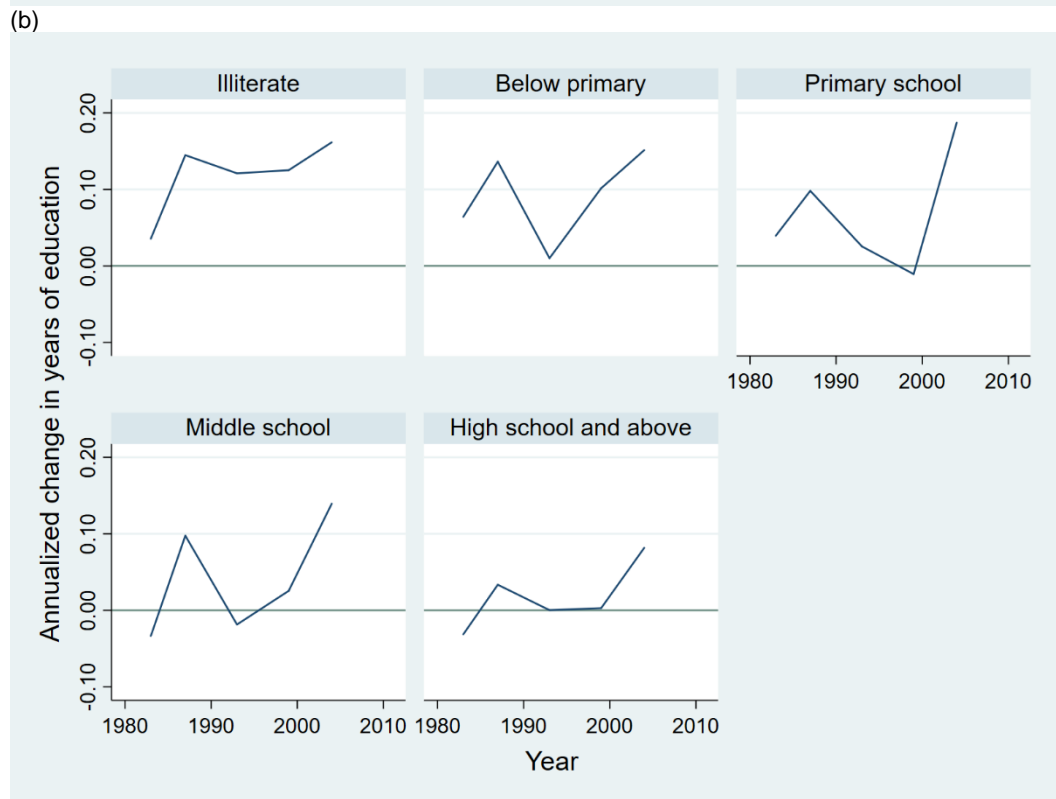
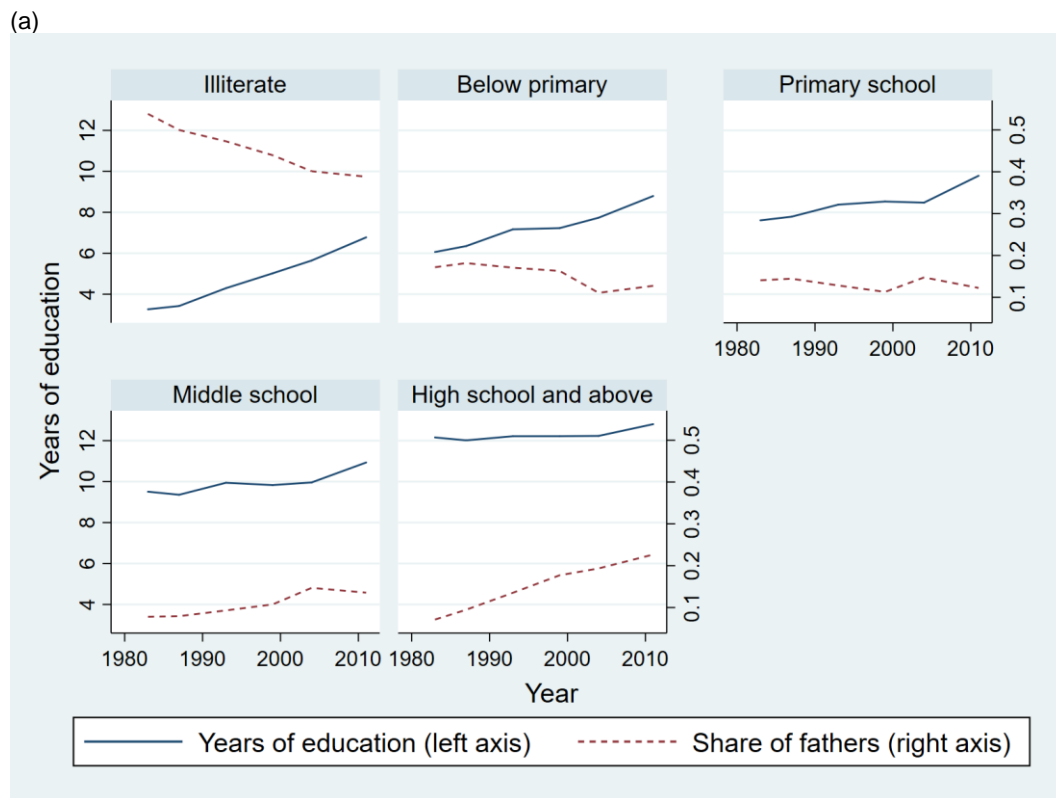
Note: Sample includes boys between 20 and 25 years living in the same household as their father.

Source: Authors' compilation based on data from NSSO (1983, 2012).

Is this improvement in years of education universally distributed in society or does parental education determine the educational attainment of their children? Looking at Figure 9a, which shows the average years of education of boys conditional on their father's education (solid line), we observe a positive correlation between the father's and son's education level. The dashed line also reveals the change in the share of fathers in each education category. Particularly, the share of illiterate and higher educated fathers changed significantly over time.

As a result, we analyse the human capital accumulation of boys conditional on their father's level of education. Figure 9b shows a similar pattern for human capital accumulation at all education level of fathers. Gains in human capital were the highest in the late 1980s and 2000s. In the 1990s, gains in educational attainment were mostly achieved by the least privileged groups with illiterate fathers.

Figure 9: (a) Years of education for boys aged 20–25 years by father’s education and their share in the sample. (b) Annualized change in years of education for boys aged 20–25 years by father’s education



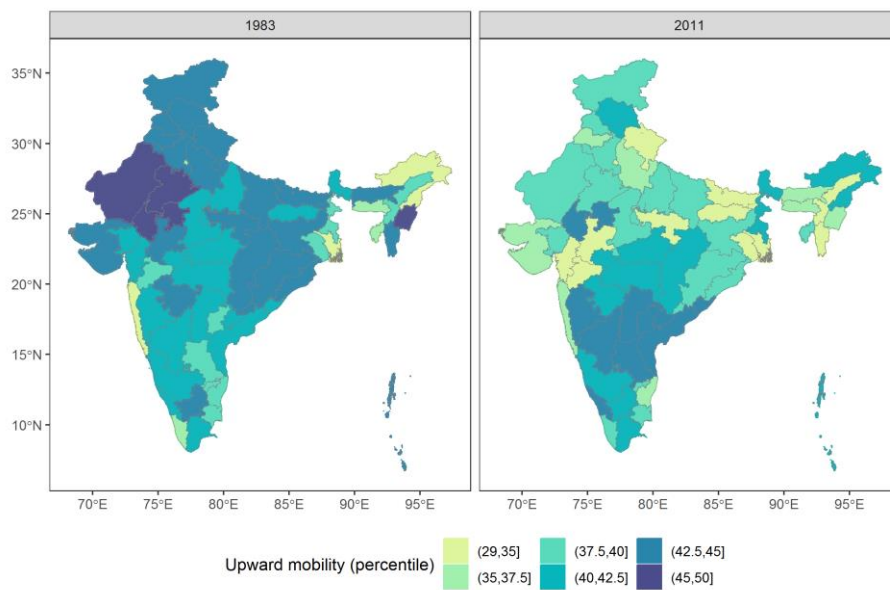
Notes: Sample includes boys between 20 and 25 years living in the same household as their father. Categories in the figures indicate the education level of fathers.

Source: Authors’ compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012).

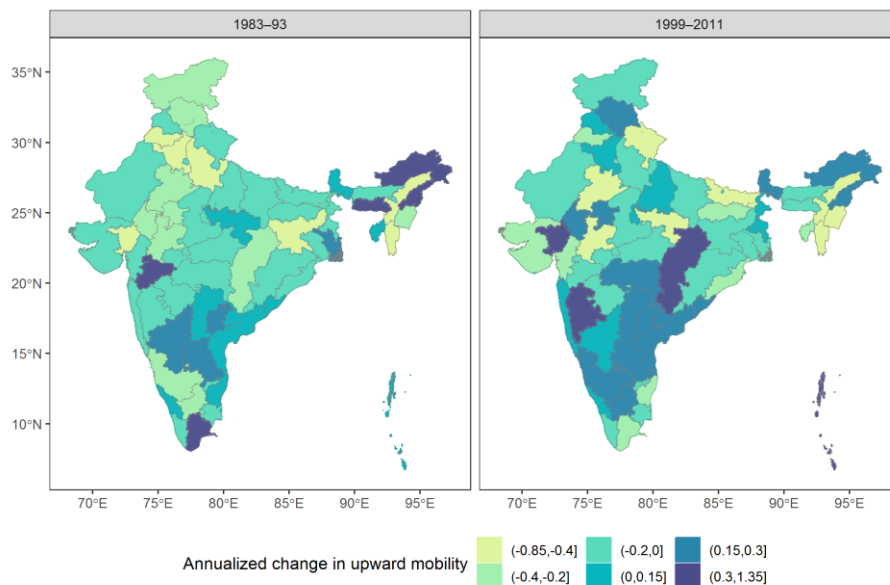
Turning to intergenerational mobility in education, it is not surprising that the period with high illiteracy rate but also large improvements in literacy exhibits the highest levels of upward mobility. Comparing the map of upward mobility in 1983 in Figure 10a with that of years of education in Figure 8, the negative correlation between the two indicators is visible in 1983 (correlation= $-0.38$ ). As the level of education increased in the population, this negative correlation became weaker and changed sign. In 2011, the correlation between the two indicators was 0.22. Figure 11a indeed indicates that the zones with the highest education level (north-eastern and southern) display the most stable upward mobility pattern over the study period, while we observe a reduction in upward mobility in the other zones.

Figure 10: (a) Intergenerational upward mobility in education and (b) annualized change in upward mobility for boys aged 20–25 in the survey state-regions

(a)



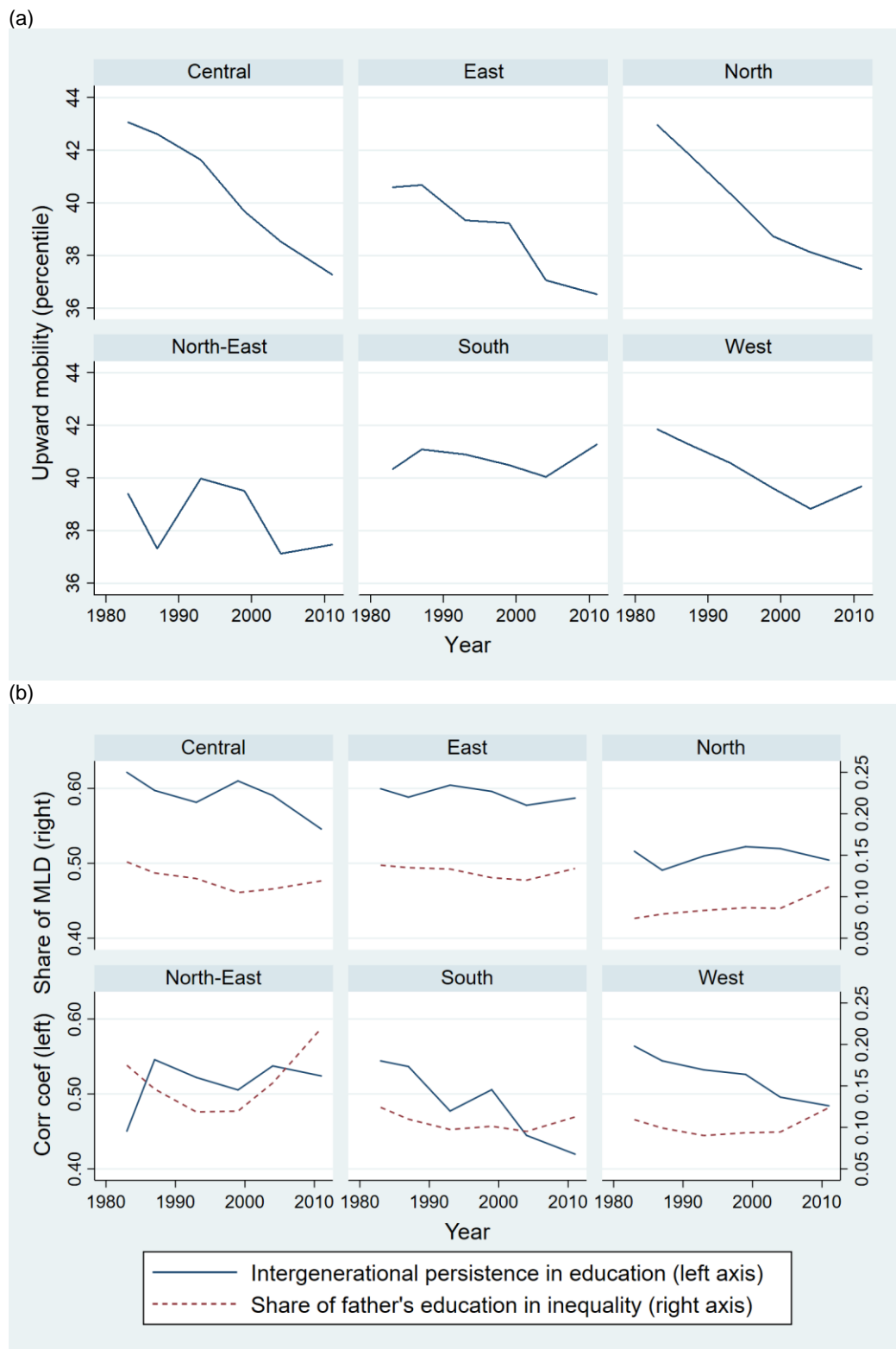
(b)



Note: Sample includes boys between 20 and 25 years living in the same household as their father.

Source: Authors' compilation based on data from NSSO (1983, 1994, 2000, 2012).

Figure 11: (a) Upward mobility in education and (b) intergenerational educational persistence (i.e. correlation coefficient and share of parental education in inequality) for boys aged 20–25 years



Notes: Sample includes boys between 20 and 25 years living in the same household as their father. Intergenerational persistence in education is measured by the correlation coefficient between the years of education of sons and fathers. The dashed line in Figure 11b shows the share of education inequality that is explained by the education level of the father, where education inequality is measured by the mean log deviation.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012).

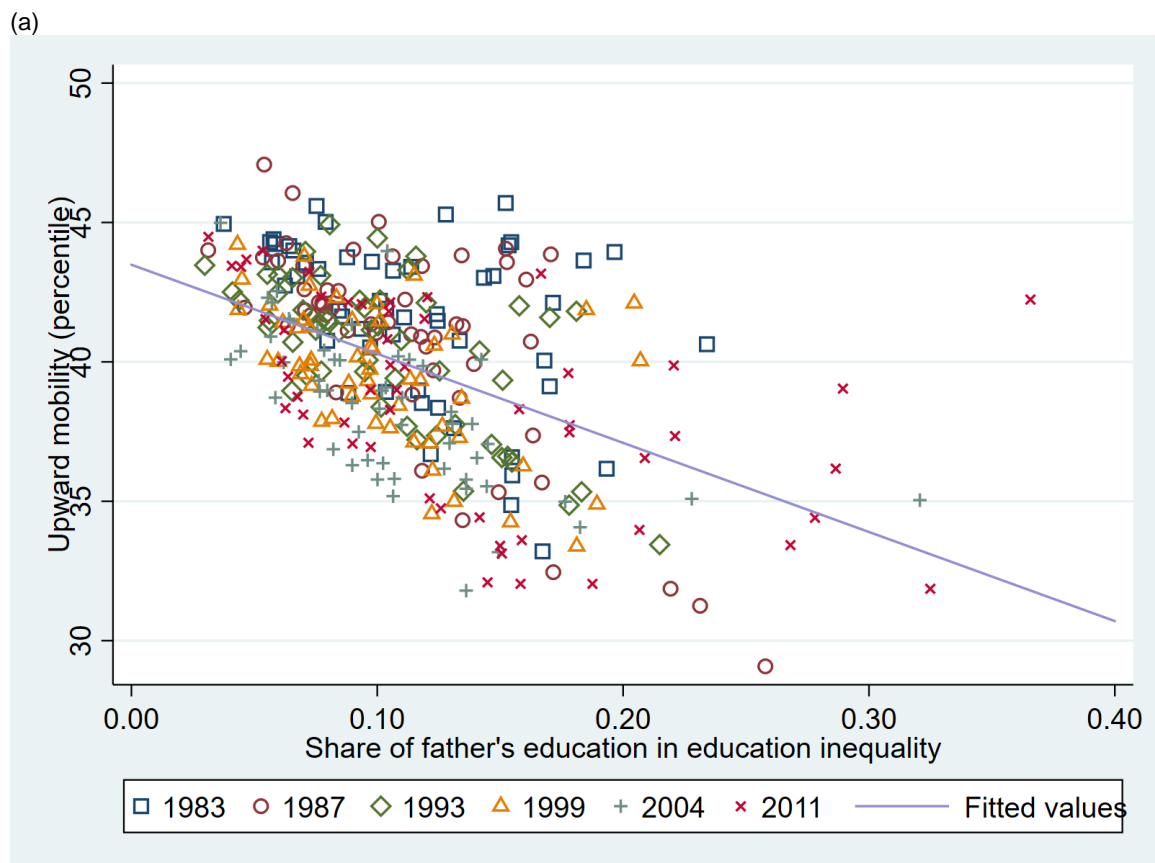
Other measures of intergenerational educational persistence can provide additional insight for assessing socio-economic mobility in India. Azam and Bhatt (2015) find that educational persistence measured by the regression coefficient of father’s education as a predictor of son’s education also declined over time, which supports our observation that increases in average educational attainment are driven primarily by increases among sons of less-educated fathers.

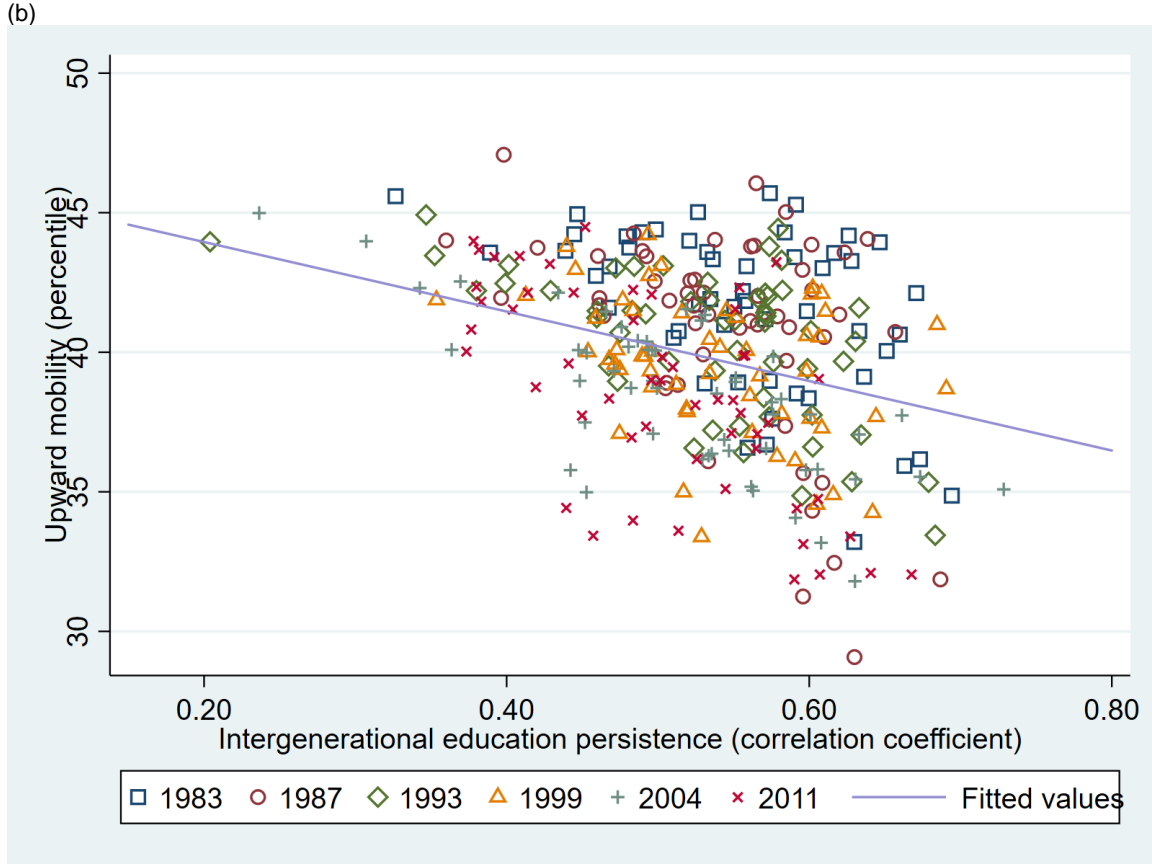
The solid line in Figure 11b shows the correlation coefficient between the years of education of sons and fathers based on the same regression specification. Similar to Azam and Bhatt (2015), we do not find the same declining trend for the correlation coefficient as for upward mobility, which indicates that while persistence declined at the lower end of the father’s educational distribution, it increased at the top end as high school graduation is becoming more universal in India.

We can capture a different aspect of intergenerational educational persistence using the share of education inequality that is explained by the father’s education level. The dashed line in Figure 11b suggests that the relative importance of parental education in education inequality did not change much in the past decades. It remained about 10 per cent in most zones despite the declining overall inequality in education.

Based on these figures, it is difficult to find a common pattern regarding educational mobility in India. While Figure 12 shows the expected negative correlation between intergenerational mobility and persistence measures, the trends in Figure 11 often appear rather flat. And in other cases, where we do observe some changes, the different measures do not agree on the direction of the trend. Therefore, we do not find conclusive evidence in the data that socio-economic mobility is improving in India. At the same time, the data indicate that education outcomes improved at least at the lowest end of the distribution.

Figure 12: (a) Intergenerational upward mobility and share of parental education in inequality and (b) upward mobility and intergenerational educational persistence for boys aged 20–25 years





Note: Sample includes boys between 20 and 25 years living in the same household as their father.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012).

Before turning to the regression analysis, it is worthwhile to take a look at some of the trends in the Indian economy, as displayed in Table 1. Over the survey period, we observe an increase in the share of services and manufacturing industries in the economy (from 22 to 33 per cent and from 15 to 24 per cent, respectively), and an improvement in the access to electricity (from 58 to 88 per cent), banking, and markets. It is important to control for these and other general socio-economic development trends as the past decades in India were characterized by significant developments in many aspects. This makes the study of India in the context of economic growth, human capital accumulation, and intergenerational mobility an interesting case.

## 5 Results

### 5.1 Identification strategy

Taking state-regions as the unit of analysis, we attempt to identify the causal effect of intergenerational mobility on future growth in household expenditure per capita at different rungs of the socio-economic ladder. Our main specification takes the following form:

$$\Delta \ln y_{it}(p) = \gamma(p) \ln y_{it-1}(p) + \alpha(p) M_{it-1} + \beta^T(p) x_{it-1} + \lambda_i(p) + \delta_t(p) + \varepsilon_{it}(p), \quad (3)$$

with  $\Delta \ln y_{it}(p) = \ln y_{it}(p) - \ln y_{it-1}(p)$ , where  $\ln y_{it}(p)$  is real household expenditure per capita at the  $p$ -th percentile for state-region  $i$  in year  $t$ ,  $M_{it-1}$  is a measure of relative intergenerational mobility,  $x_{it}$  is

a vector of control variables for state-region  $i$  that may vary with time  $t$ ,  $\lambda_i(p)$  denotes zone fixed effects (FEs), and  $\delta_t(p)$  denotes time FEs.

The zone FEs control for local features that are time-invariant such as local climate, geographic conditions, and local culture. The year effects control for time-varying conditions at the all-India level, including changes in global food and commodity prices, terms of trade, and shifts in public policy. To account for time-varying features at the state-region level that could affect both socio-economic mobility and household expenditure growth, we also control for total inequality in household expenditure per capita, the share of working age individuals (between the age of 16 and 65 years), the average years of education among working age individuals, sectoral composition of the labour market (share working in manufacturing and share working in services), share of households with access to electricity, financial inclusion (log of number of banks per capita), share of urban population, and share of Christians. All controls are lagged by one period.

In the regression analysis, we drop between one and three observations (depending on the choice of percentile) that are identified as outliers based on standard regression diagnostics; the studentized residual exceeds four and/or the Cook's distance exceeds 0.1. All outliers correspond to small state-regions and the survey round from 1983. The final sample used in the ordinary least-square (OLS) regressions (with lagged independent variables) ranges between 267 and 270 observations. The number of observations used in our instrumental variables regressions reduces the sample to 240 observations. Standard errors are robust to heteroskedasticity.

The parameter of interest is  $\alpha(p)$ , which measures the effect of intergenerational mobility on future growth in household expenditure per capita at the  $p$ -th percentile. OLS estimation of  $\alpha(p)$  may be subject to three potential sources of bias: Nickell bias (Nickell 1981), omitted variables bias, and reverse causality bias. Nickell bias emerges as controlling for the time-invariant location FE introduces a correlation between the residuals and the lagged dependent variable (and selected independent variables). We minimize this bias by using grouped FE (i.e. zone instead of state-region FE) combined with using an inclusive set of control variables that includes both time-variant and time-invariant variables. This strategy also helps limit bias due to omitted variables. We further address the concern of omitted variables bias by considering an expansion to the set of control variables (at the expense of losing a number of observations as the additional controls are not available for the full set of state-regions).

The third potential source of bias, due to reverse causality, may emerge when intergenerational mobility is itself driven by innovations in the standard of living in the state-region measured by household expenditure per capita. This concern is partially mitigated by using lagged independent variables; it is less likely that socio-economic mobility today is due to changes in the standard of living over the next five years (controlling for today's standard of living). However, when intergenerational mobility is highly persistent, reverse causality cannot be ruled out entirely. We address this concern by constructing an instrument for intergenerational mobility. The resulting instrumental variables regression serves as our preferred method of estimation.

## 5.2 Intergenerational mobility: First-stage results

Our instrument for intergenerational mobility essentially consists of the local share of the Brahman caste in 1931 and the local share of scheduled tribes in 1961, both interacted with national trends in intergenerational mobility. Specifically, we regress intergenerational mobility on the local shares of Brahman and scheduled tribes interacted with the six time-period dummy variables to allow for non-linear time-trends, and use the predicted values from this regression as our instrument. The Brahman caste are among the first to take up Western education and arguably have played an

important role in spreading education in India. In contrast, the scheduled tribes rank among the most disadvantaged in India and historically have had limited access to education. On the basis of this one would expect socio-economic mobility to be positively correlated with the share of Brahman and negatively correlated with the share of scheduled tribes. This is confirmed in Table 2 that shows the results from the first-stage regression. The correlation with the share of Brahman is found to be stronger.

Table 2: First-stage regression for upward mobility

	Upward mobility
Share of Brahmans	
1931×Year 1983	24.3944*** (7.1040)
1931×Year 1987	25.0136** (9.6999)
1931×Year 1993	14.5063* (8.2947)
1931×Year 1999	-2.9133 (6.9821)
1931×Year 2004	-3.6136 (8.0743)
1931×Year 2011	-35.4712** (14.0767)
Share of scheduled tribes	
1961×Year 1983	4.3067 (2.9879)
1961×Year 1987	5.2319** (2.6420)
1961×Year 1993	2.9937 (2.8835)
1961×Year 1999	-2.3170 (2.7376)
1961×Year 2004	-4.5596 (3.3002)
1961×Year 2011	-4.3157 (4.3311)
<i>F</i>	6.4437
Adjusted <i>R</i> <sup>2</sup>	0.1769

Notes: Heteroskedasticity robust standard errors reported (\**P*=0.10, \*\**P*=0.05, \*\*\**P*=0.01). Year fixed effects (FEs) not reported.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005 and 2012) and Castelló-Climont et al. (2017).

The motivation for using the share of Brahman and share of scheduled tribes in 1931 and 1961, respectively, is that the resulting instrument is plausibly orthogonal to changes in household welfare between 1983 and 2011 conditional on initial household welfare. While this benefits the validity of the instrument, it may weaken its strength. Indeed, the correlation between the instrument and observed mobility for any given year ranges between 0.11 and 0.33.

### 5.3 Intergenerational mobility and inclusive growth

The estimates of Equation 3 obtained from the instrumental variables regression are presented in Table 3. The regression coefficients corresponding to intergenerational mobility are also plotted in Figure 13. The following observations stand out. First, intergenerational mobility is found to have a positive effect on growth for all percentiles, although the effects are not statistically significant. Second, the effect is visibly larger (and almost significant) at lower percentiles. This is consistent with the hypothesis that higher intergenerational mobility is good for growth, particularly for inclusive growth as those held back by an uneven playing field tend to be concentrated towards the bottom of the income distribution. Note that this finding also predicts a negative relationship between intergenerational mobility and inequality: higher mobility is associated with inclusive growth which in turn is associated with lower inequality. Estimates obtained using OLS (shown in Appendix B1) are qualitatively similar.



Table 3: Determinants of real per capita consumption expenditure growth at different percentiles (IV regression: Model 1)

	10%	25%	50%	75%	90%	95%
AR(1)	-0.0985*** (0.0135)	-0.0849*** (0.0123)	-0.0772*** (0.0105)	-0.0694*** (0.0091)	-0.0671*** (0.0099)	-0.0741*** (0.0115)
Upward mobility	0.0038 (0.0026)	0.0033 (0.0025)	0.0021 (0.0024)	0.0016 (0.0024)	0.0012 (0.0020)	0.0019 (0.0022)
Inequality in RMPCE	-0.0582 (0.0476)	-0.0371 (0.0397)	-0.0281 (0.0362)	-0.0281 (0.0332)	-0.0638* (0.0361)	-0.0773* (0.0464)
Share of age 16–64 years	0.0699 (0.0578)	0.0516 (0.0532)	0.0798 (0.0501)	0.1417*** (0.0532)	0.1943*** (0.0537)	0.2087*** (0.0602)
Years of education	0.0025 (0.0035)	0.0031 (0.0033)	0.0020 (0.0032)	0.0018 (0.0032)	0.0025 (0.0028)	0.0038 (0.0032)
Services industry	0.1246*** (0.0323)	0.1041*** (0.0334)	0.1023*** (0.0369)	0.0692** (0.0352)	0.0407 (0.0350)	0.0385 (0.0394)
Manufacturing industry	0.0683*** (0.0216)	0.0612*** (0.0201)	0.0581*** (0.0195)	0.0490** (0.0197)	0.0565*** (0.0209)	0.0673*** (0.0226)
Access to electricity	-0.0063 (0.0094)	-0.0076 (0.0087)	-0.0061 (0.0080)	-0.0072 (0.0084)	-0.0094 (0.0081)	-0.0094 (0.0088)
Log bank per capita	0.0032 (0.0037)	0.0030 (0.0029)	0.0032 (0.0023)	0.0036* (0.0020)	0.0025 (0.0018)	0.0027 (0.0018)
Share of urban	-0.0257 (0.0227)	-0.0158 (0.0212)	-0.0191 (0.0212)	-0.0020 (0.0208)	0.0159 (0.0190)	0.0319 (0.0209)
Share of Christians	-0.0178 (0.0389)	-0.0146 (0.0359)	-0.0118 (0.0347)	0.0045 (0.0321)	0.0093 (0.0339)	0.0203 (0.0346)
Observations	240	240	240	240	240	240
Adjusted $R^2$	0.361	0.393	0.491	0.559	0.612	0.619

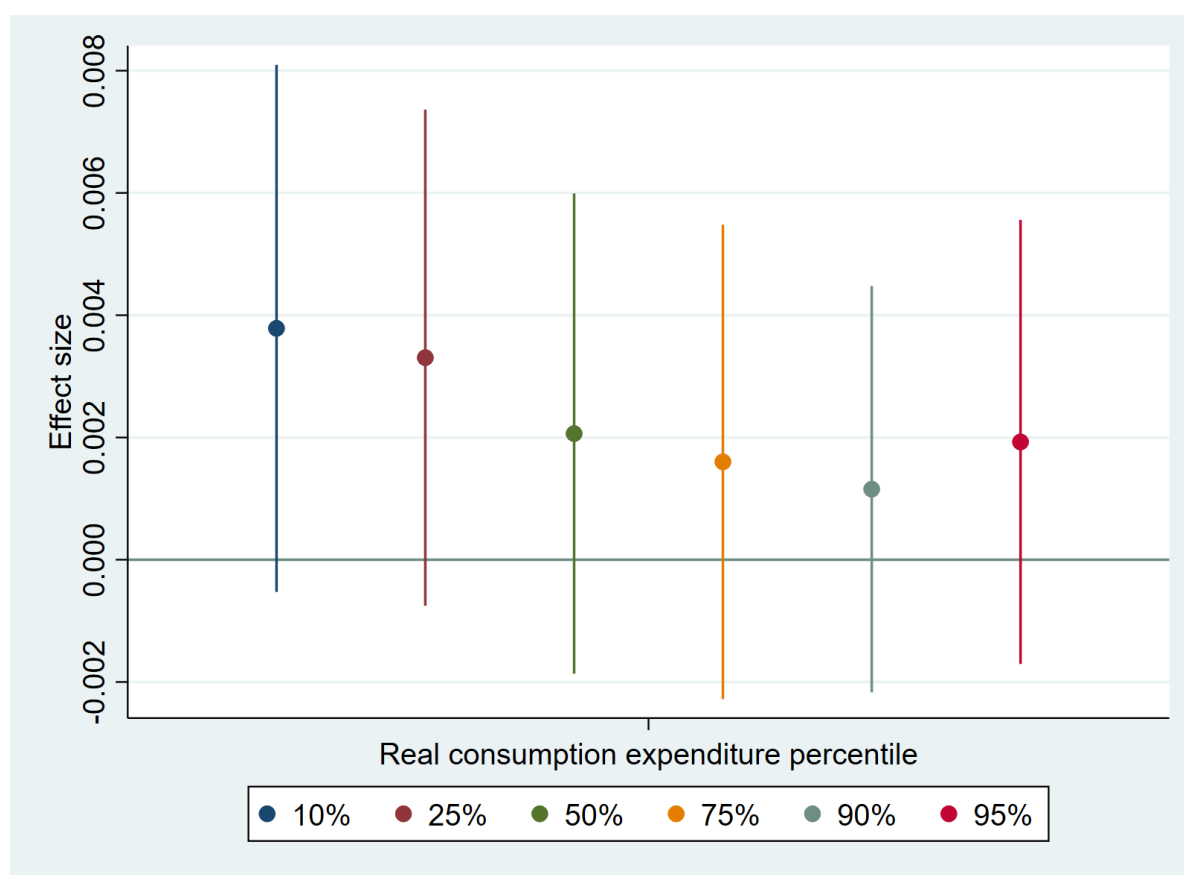
Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years. Instruments for upward mobility: share of Brahman (1931) and share of scheduled tribes (1961) interacted with the survey year.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), and Government of India (2012–15a, 2012–15b).

Let us also briefly comment on the effects of the control variables.<sup>7</sup> A high share of working age population is found to be a positive force for growth, particularly for growth at the high end of the distribution (among the better-off). Growth is also found to be higher in state-regions where there is more employment in manufacturing and services and less employment in agriculture. Access to electricity and the share of individuals residing in urban areas is either insignificant or is negatively associated with growth. This might suggest that an important share of the growth may be originating in rural areas. Financial inclusion measured by (the log of) the number of banks per capita is found to have a positive effect on growth, although the coefficient is not significant except for the 75th percentile. Education measured by the average years of schooling and the share of Christians also have a positive yet insignificant effect. The lack of significance of education is perhaps somewhat surprising. Finally, inequality in household expenditure per capita is found to have a negative effect on future growth. The effect is significant at high percentiles but not at lower percentiles. This effect, however, could be in part mechanical; it can be verified that if household expenditures follow an autoregressive process (for each household) where state-region inequality does not feature as an independent variable, then inequality would be negatively correlated with anonymous growth at the top end of the distribution and positively correlated with growth at the lower end.

<sup>7</sup> Note that we do not use instruments for these variables.

Figure 13: Effects of upward mobility on consumption expenditure growth at different percentiles (IV regression: Model 1)



Note: Ninety per cent confidence interval plotted.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), and Government of India (2012–15a, 2012–15b).

The estimates for the control variables obtained using OLS are mostly similar with two notable exceptions: (a) the positive effect of financial inclusion and its significance is increased, and (b) the coefficient associated with the share of Christians is now positive and significant for all percentiles. Next, we expand the set of control variables in an effort to further address concerns of omitted variables bias. It should be noted, however, that doing so we reduce the number of observations to 198–200 as the additional controls are not available for the state-regions of small states. The added controls include: (a) log of market access, which captures the extent of domestic market integration, (b) public expenditures as a share of local GDP, (c) voter turnout in the most recent state-election, and (d) political competition measured by the share of votes going to the second largest political party minus the share going to the largest party. The total number of observations used in the regression analysis is reduced from 240 to 198–200. We use the exact same instrument for our measure of intergenerational mobility. (See Table 2 for the first-stage regression results.)

Table 4 presents the instrumental variables estimates for the growth regressions using the expanded set of controls, and Figure 14 plots the coefficients corresponding to intergenerational mobility. While the relationship between intergenerational mobility and growth at the different percentiles is similar to the one obtained using the smaller set of controls, the significance of this relationship is notably stronger when the expanded set of controls is used. The effect of mobility is now positive and strongly significant for growth at low percentiles while it is small and

insignificant at higher percentiles. This confirms that higher socio-economic mobility is good for growth, particularly for inclusive growth.

While the coefficients corresponding to the added controls are not statistically significant, the effects are economically significant (and nearly statistically significant) in two notable cases: (a) market access has a positive effect on growth at top end of the distribution, while (b) public expenditures is found to have a positive effect on growth at the bottom end of the distribution.

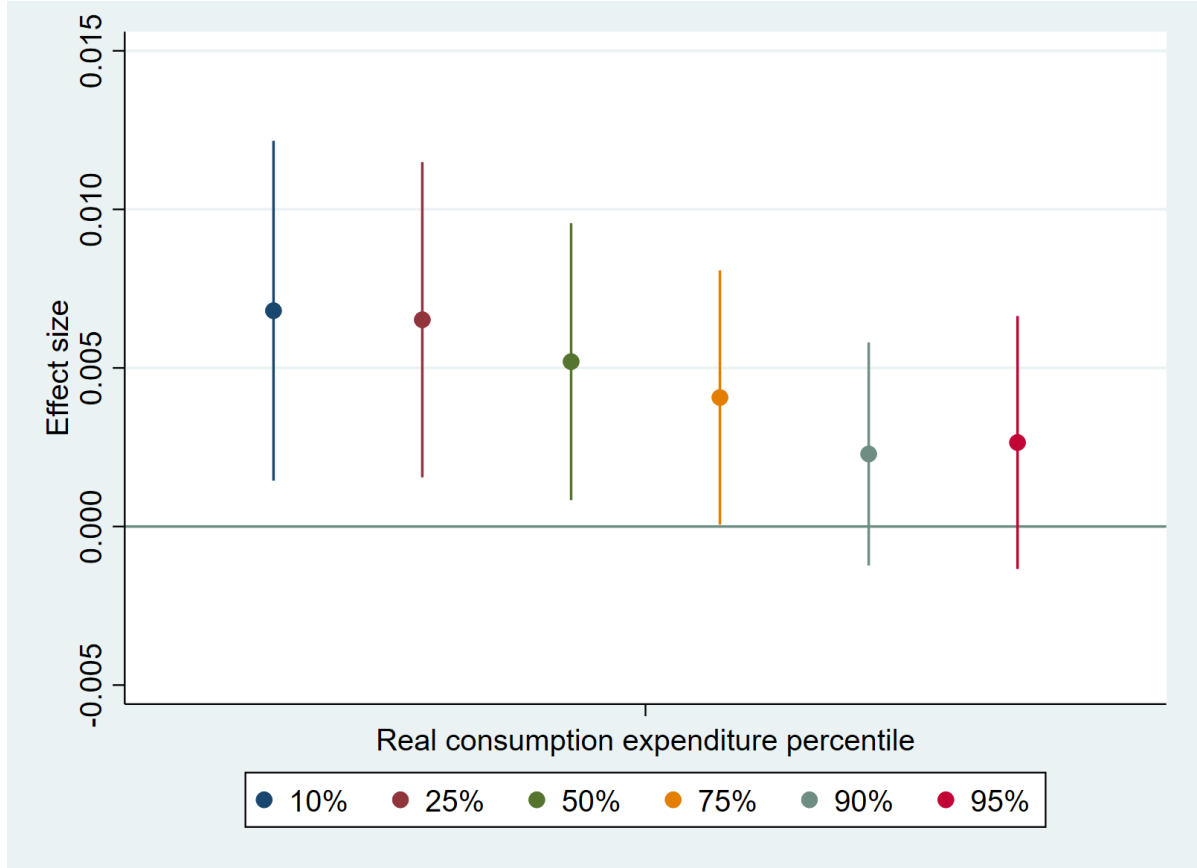
Table 4: Determinants of real per capita consumption expenditure growth at different percentiles (IV regression: Model 2)

	10%	25%	50%	75%	90%	95%
AR(1)	-0.1120*** (0.0194)	-0.0936*** (0.0188)	-0.0812*** (0.0158)	-0.0646*** (0.0134)	-0.0538*** (0.0121)	-0.0662*** (0.0141)
Upward mobility	0.0068** (0.0033)	0.0065** (0.0030)	0.0052* (0.0027)	0.0041* (0.0024)	0.0023 (0.0021)	0.0026 (0.0024)
Inequality in RMPCE	-0.1387** (0.0579)	-0.1008** (0.0513)	-0.0778* (0.0461)	-0.0623 (0.0391)	-0.0950*** (0.0340)	-0.1213*** (0.0448)
Share of age 16–64 years	0.1575** (0.0725)	0.1330* (0.0712)	0.1616** (0.0662)	0.2078*** (0.0623)	0.2390*** (0.0638)	0.2658*** (0.0674)
Years of education	0.0039 (0.0044)	0.0047 (0.0040)	0.0037 (0.0036)	0.0029 (0.0033)	0.0023 (0.0028)	0.0036 (0.0032)
Services industry	0.1858*** (0.0596)	0.1660*** (0.0579)	0.1646*** (0.0573)	0.1358*** (0.0484)	0.1082** (0.0494)	0.0834 (0.0572)
Manufacturing industry	0.0619** (0.0265)	0.0507* (0.0272)	0.0419 (0.0271)	0.0254 (0.0255)	0.0236 (0.0260)	0.0451 (0.0288)
Access to electricity	-0.0266* (0.0148)	-0.0268* (0.0138)	-0.0212* (0.0122)	-0.0229* (0.0120)	-0.0166 (0.0122)	-0.0148 (0.0135)
Log bank per capita	-0.0112 (0.0157)	-0.0110 (0.0143)	-0.0073 (0.0129)	-0.0029 (0.0113)	0.0020 (0.0030)	0.0015 (0.0031)
Share of urban	0.0205 (0.0401)	0.0243 (0.0372)	0.0103 (0.0345)	0.0181 (0.0314)	0.0123 (0.0313)	0.0379 (0.0348)
Share of Christians	0.0019 (0.0496)	0.0003 (0.0465)	-0.0020 (0.0415)	0.0094 (0.0367)	0.0114 (0.0336)	0.0256 (0.0345)
Market access	-0.0038 (0.0169)	-0.0010 (0.0151)	0.0054 (0.0144)	0.0084 (0.0135)	0.0198 (0.0137)	0.0225 (0.0160)
Government expenditure per GDP	0.0125 (0.0427)	0.0173 (0.0405)	0.0297 (0.0384)	0.0170 (0.0341)	-0.0017 (0.0314)	-0.0117 (0.0386)
Voter turnout last election	0.0005 (0.0004)	0.0004 (0.0004)	0.0002 (0.0003)	0.0001 (0.0002)	-0.0002 (0.0002)	0.0000 (0.0002)
Political competition	-0.0008 (0.0080)	0.0016 (0.0073)	-0.0005 (0.0069)	-0.0000 (0.0064)	-0.0029 (0.0065)	-0.0041 (0.0068)
Observations	198	198	198	198	200	200
Adjusted $R^2$	0.147	0.077	0.269	0.425	0.581	0.584

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years. Instruments for upward mobility: share of Brahman (1931) and share of scheduled tribes (1961) interacted with the survey year. RMPCE stands for real monthly per capita consumption expenditure.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and (STICERD and LSE 2005–19).

Figure 14: Effects of upward mobility on consumption expenditure growth at different percentiles (IV regression: Model 2)



Note: Ninety per cent confidence interval plotted.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and (STICERD and LSE 2005–19).

#### 5.4 Intergenerational mobility and human capital

We conjecture that human capital accumulation denotes an important channel via which intergenerational mobility impacts on growth and on the degree of ‘inclusivity’ of growth. To explore the plausibility of this conjecture, we consider changes in years of schooling for individuals with different parental education backgrounds as dependent variable. Let  $h_{it}(k)$  denote the average years of schooling in sub-region  $i$  and time  $t$  for boys whose father have an education level equal to  $k$ . Five different levels for father education are considered: (1) illiterate, (2) below primary, (3) primary, (4) secondary, and (5) tertiary degree. Assume that human capital accumulation at different parental education backgrounds can be described by:

$$\Delta h_{it}(k) = \phi(k)h_{it-1}(k) + \alpha(k)M_{it-1} + b^T(k)x_{it-1} + \omega_i(k) + \eta_t(k) + u_{it}(k), \quad (4)$$

with  $\Delta h_{it}(k) = h_{it}(k) - h_{it-1}(k)$ , where again  $M_{it-1}$  denotes our measure of relative intergenerational mobility,  $x_{it}$  is a vector of control variables,  $\omega_i(k)$  represents zone FEs, and  $\eta_t(k)$  denotes time FEs. Our prediction is that the effect of intergenerational mobility measured by  $\alpha(k)$  is positive and declines with  $k$ . The effect of mobility on human capital accumulation is arguably zero for individuals from highly advantaged backgrounds, for whom it matters less whether or not the playing field is even.

We use the exact same instrument for intergenerational mobility and the same set of core controls. However, one variable is added to the expanded set controls, namely government expenditure on education as a share of (local) GDP.

Using years of schooling as the dependent variable, there are no observations identified as outliers (based on either the studentized residual or the Cook's distance statistic). In the OLS regressions (see Appendix B1) with either the core or expanded set of controls, the number of observations equals 270 and 200, respectively. For the instrumental variables regressions, we have a total of 240 and 200 observations, respectively. Standard errors are robust to heteroskedasticity.

The instrumental variables estimates of Equation 4 using the core set of controls are shown in Table 5. The estimates of  $\alpha(k)$  are also plotted in Figure 15. These estimates confirm that mobility has a positive and significant effect on human capital accumulation of individuals with less than highly educated parents, while the effect is insignificant for individuals with highly educated parents (for whom it matters less whether or not the playing field is level). This result is robust to the choice of controls. Using the expanded set of controls changes very little (see Table 6 and Figure 16).

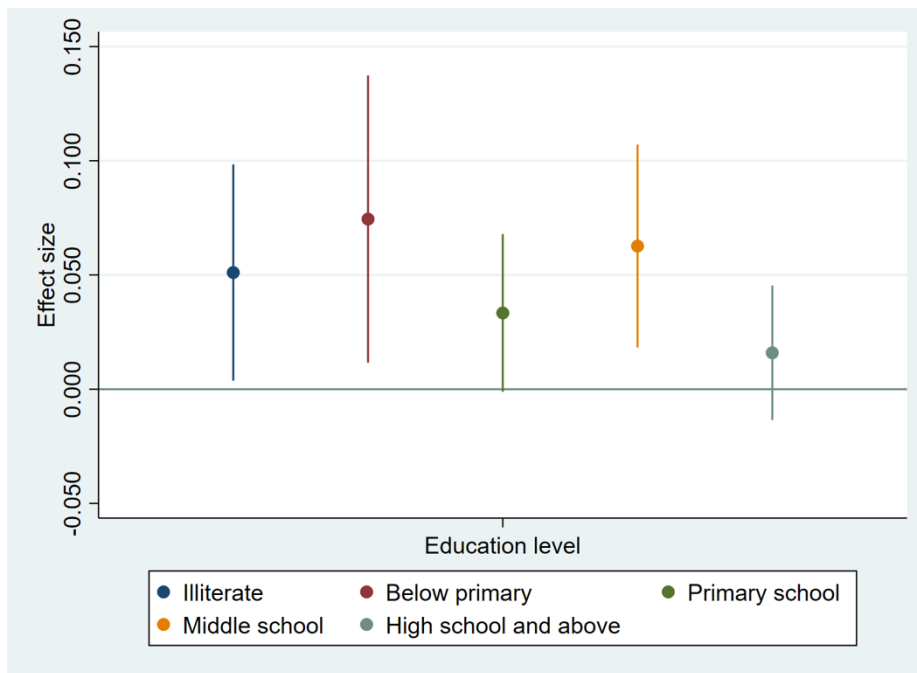
Table 5: Determinants of change in boys' years of education (20–25 years) by father's education (IV regression: Model 1)

	Illiterate	Below primary	Primary	Middle	High and above
AR(1)	-0.1672*** (0.0460)	-0.1974*** (0.0395)	-0.1430*** (0.0146)	-0.1429*** (0.0163)	-0.1602*** (0.0128)
Upward mobility	0.0511* (0.0288)	0.0745* (0.0382)	0.0334 (0.0210)	0.0627** (0.0270)	0.0159 (0.0179)
Inequality in RMPCE	-0.0435 (0.3401)	-0.4444 (0.5967)	-0.0060 (0.3785)	-0.0750 (0.4971)	-0.4918* (0.2989)
Share of age 16–64 years	0.3662 (0.5353)	-0.1230 (0.7224)	-0.5208 (0.4810)	0.3909 (0.6378)	0.0562 (0.4225)
Years of education	0.1682** (0.0703)	0.1350** (0.0640)	0.0943*** (0.0297)	0.0741* (0.0387)	0.0190 (0.0248)
Services industry	-0.0092 (0.3872)	0.4387 (0.6006)	-0.5651* (0.3070)	0.4715 (0.4432)	0.4335 (0.2934)
Manufacturing industry	-0.2428 (0.2209)	-0.2535 (0.3199)	-0.2837 (0.2617)	-0.0111 (0.3348)	0.0213 (0.1604)
Access to electricity	0.0731 (0.0705)	-0.0734 (0.1202)	0.0646 (0.0740)	-0.0798 (0.1164)	-0.0191 (0.0763)
Log bank per capita	-0.0441** (0.0200)	-0.0167 (0.0212)	0.0061 (0.0132)	-0.0543 (0.0413)	-0.0041 (0.0180)
Share of urban	-0.2642 (0.2014)	0.0616 (0.3613)	0.4126* (0.2346)	0.0212 (0.3002)	0.0804 (0.1785)
Share of Christians	-0.5696 (0.4521)	-0.3794 (0.5963)	-0.1947 (0.4033)	-0.2662 (0.5346)	-0.1173 (0.2730)
Observations	240	240	239	240	240
Adjusted $R^2$	-0.043	-0.050	0.366	0.135	0.434

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years. Instruments for upward mobility: share of Brahman (1931) and share of scheduled tribes (1961) interacted with the survey year.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005 and 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), and Government of India (2012–15a, 2012–15b).

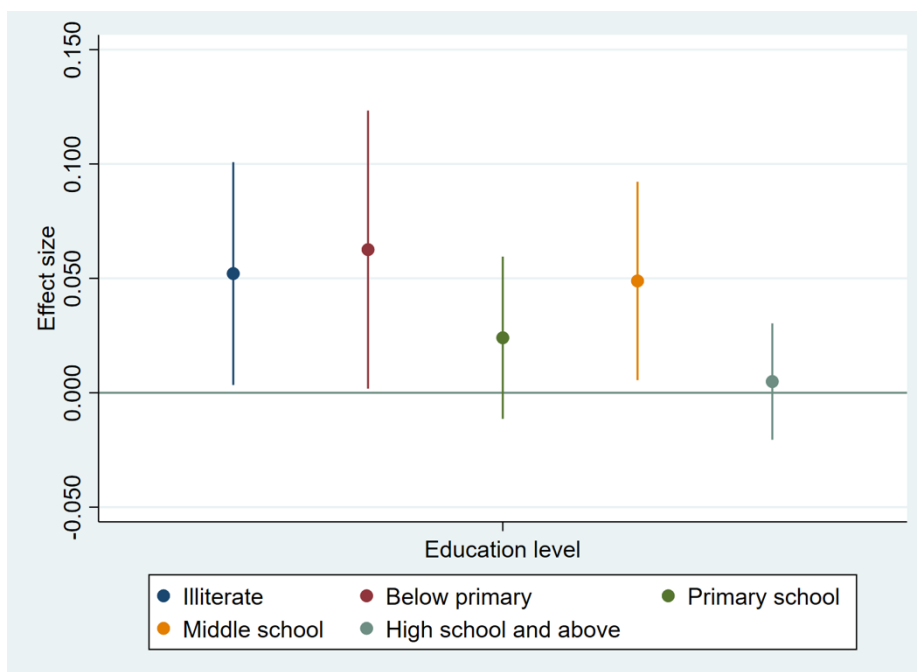
Figure 15: Effects of upward mobility on change in boys' years of education (20–25 years) by father's education (IV regression: Model 1)



Note: Ninety per cent confidence interval plotted.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), and Government of India (2012–15a, 2012–15b).

Figure 16: Effects of upward mobility on change in boys' years of education (20–25 years) by father's education (IV regression: Model 2)



Note: Ninety per cent confidence interval plotted.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and (STICERD and LSE 2005–19).

Let us also briefly highlight coefficients obtained for selected controls. Many of the controls are insignificant (when also controlling for zone FE; omitting zone FE increases the significance of the controls). The control whose significance stands out is (log of) market access, which is found to have a positive effect on human capital accumulation across all parental backgrounds (Table 6).

Table 6: Determinants of change in boys' years of education (20–25 years) by father's education (IV regression: Model 2)

	Illiterate	Below primary	Primary	Middle	High and above
AR(1)	-0.1646*** (0.0489)	-0.1900*** (0.0416)	-0.1501*** (0.0158)	-0.1513*** (0.0167)	-0.1605*** (0.0129)
Upward mobility	0.0521* (0.0296)	0.0626* (0.0370)	0.0240 (0.0216)	0.0489* (0.0264)	0.0049 (0.0155)
Inequality in RMPCE	-0.4478 (0.3546)	-1.0055* (0.5606)	0.2946 (0.4074)	-0.5910 (0.4704)	-0.4971 (0.3240)
Share of age 16–64 years	0.8898 (0.6049)	0.6941 (0.8385)	-0.6400 (0.5828)	0.9258 (0.7416)	0.4440 (0.4835)
Years of education	0.1486** (0.0709)	0.1082* (0.0600)	0.0811*** (0.0299)	0.0262 (0.0356)	-0.0039 (0.0205)
Services industry	-0.3762 (0.4771)	-0.0155 (0.6849)	-0.7553 (0.4606)	0.1860 (0.5138)	0.4467 (0.3999)
Manufacturing industry	-0.1400 (0.2550)	-0.1209 (0.3423)	-0.3857 (0.3159)	0.1229 (0.3391)	0.0253 (0.1882)
Access to electricity	0.1268 (0.1008)	-0.0153 (0.1492)	0.2370** (0.1140)	-0.0682 (0.1600)	0.0984 (0.1029)
Log bank per capita	-0.0457* (0.0260)	-0.0487** (0.0212)	0.0066 (0.0225)	-0.0235 (0.0179)	-0.0465*** (0.0109)
Share of urban	0.1093 (0.3417)	0.2034 (0.4561)	0.3563 (0.3229)	0.3851 (0.3984)	-0.1243 (0.2201)
Share of Christians	-0.3023 (0.4851)	-0.1000 (0.5301)	0.1344 (0.4465)	0.3237 (0.4703)	0.0591 (0.2833)
Market access	0.3704** (0.1663)	0.4104* (0.2112)	0.2689* (0.1417)	0.4115** (0.1777)	0.2025* (0.1116)
Government expenditure per GDP	-0.7205 (0.4620)	0.6287 (0.6159)	0.7098 (0.5255)	-0.7440 (0.5840)	-0.4101 (0.3308)
Educational expenditure per GDP	2.0334 (2.0825)	-2.4099 (2.6591)	-2.7618 (2.2030)	0.3208 (2.2873)	1.8425 (1.2666)
Voter turnout last election	0.0010 (0.0021)	0.0033 (0.0025)	-0.0026 (0.0020)	-0.0001 (0.0024)	-0.0017 (0.0014)
Political competition	0.0540 (0.0660)	0.1029 (0.0778)	-0.0523 (0.0755)	0.0708 (0.0816)	-0.0320 (0.0529)
Observations	200	200	200	200	200
Adjusted $R^2$	-0.065	0.102	0.404	0.286	0.463

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years. Instruments for upward mobility: share of Brahman (1931) and share of scheduled tribes (1961) interacted with the survey year.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Castelló-Climent et al. (2017), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and (STICERD and LSE 2005–19).

## 6 Conclusion

Despite a consistently high economic growth rate in the past decades, India stands out as one of the countries with the least social mobility and rising income inequality. The historical segregation of ethnic groups with lower socio-economic status led to large differences in socio-economic outcomes in different regions in India. Since India's independence, affirmative action programmes

were meant to increase opportunities for scheduled castes and tribes and other backward castes, which resulted in varying degrees of success. These conditions and the high quality of historically available data mark India as one of the most interesting cases to study intergenerational social mobility and its effect on economic growth and human capital accumulation at different rungs of the socio-economic ladder.

We study the period between 1983 and 2011. During this period, illiteracy decreased from one in three young adult men (ages 20–25 years) to 10 per cent, and the percentage of high school graduates increased from 23 to 55 per cent. A large part of the gains in educational attainment occurred among the most disadvantaged socio-economic groups with illiterate fathers. Despite these positive developments, we do not find conclusive evidence in the data that socio-economic mobility is improving in India.

This makes it all the more important to investigate the potential benefits of higher socio-economic mobility in India. We hypothesize that higher socio-economic mobility leads to higher economic growth, which at the same time benefits particularly the lower socio-economic groups (inclusive growth). Using upward mobility to measure intergenerational socio-economic mobility as proposed by Asher et al. (2018), we find evidence for this hypothesis. The effect estimated using instrumental variables regression is the strongest in the sample of large states of India (using the expanded set of controls).

We further conjecture that human capital accumulation denotes an important channel via which intergenerational mobility has an impact on growth and on the degree of ‘inclusivity’ of growth. We test this hypothesis by investigating the effect on upward mobility on the years of education of boys in the age group 20–25 years conditional on their father’s education level. Our results indicate that upward mobility particularly increases the educational attainment of boys whose fathers are in the lower ranks of the education ladder (below primary).

These results suggest that policies aimed at improving socio-economic mobility not only improve the socio-economic position of disadvantaged groups, thereby decreasing inequality, but they also have a beneficial effect on the performance of the economy. Further research is needed to elaborate on the channels that arise from these effects.

## References

- Allen, Treb, and David Atkin (2016). ‘Volatility and the Gains from Trade’. NBER Working Paper 22276. Cambridge, MA: National Bureau of Economic Research (NBER). Available at: <https://www.nber.org/papers/w22276.pdf> (accessed 2 January 2019).
- Asher, Sam, Paul Novosad, and Charlie Rafkin (2018). ‘Intergenerational Mobility in India: Estimates from New Methods and Administrative Data’. World Bank Working Paper. Available at: <http://www.dartmouth.edu/~novosad/anr-india-mobility.pdf> (accessed December 2018).
- Azam, Mehtabul, and Vipul Bhatt (2015). ‘Like Father, Like Son? Intergenerational Educational Mobility in India’. *Demography*, 52: 1929–59.
- Besley, Timothy, and Robin Burgess (2002). ‘The Political Economy of Government Responsiveness: Theory and Evidence from India’. *The Quarterly Journal of Economics*, 117: 1415–51.
- Butler, David, Ashok Lahiri, and Prannoy Roy (1995). ‘Vidhan Sabha Elections’. In *India Decides: Elections, 1952–1995*. New Delhi: Books and Things.



- Cassan, Guilhem (2019). ‘Affirmative Action, Education and Gender: Evidence from India’. *Journal of Development Economics*, 136: 51–70. Available online since October 2018: <https://doi.org/10.1016/j.jdeveco.2018.10.001> (accessed on 20 December 2018).
- Castelló-Climent, Amparo, Latika Chaudhary, and Abhiroop Mukhopadhyay (2017). ‘Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India’. *The Economic Journal*, 128(616): 3039–75.
- Census of India (2001). Data Highlights. Registrar General and Census Commissioner, Government of India. Available at [http://censusindia.gov.in/Data\\_Products/Data\\_Highlights/Data\\_Highlights\\_link/data\\_highlights\\_D1D2D3.pdf](http://censusindia.gov.in/Data_Products/Data_Highlights/Data_Highlights_link/data_highlights_D1D2D3.pdf) (accessed January 2019).
- Census of India (2011). Population and Vital Statistics. Registrar General and Census Commissioner, Government of India. Available at [http://ayush.gov.in/sites/default/files/Population%20and%20Vital%20Statistics%20Table 2015.pdf](http://ayush.gov.in/sites/default/files/Population%20and%20Vital%20Statistics%20Table%202015.pdf) (accessed January 2019).
- Datt, Gaurav, and Martin Ravallion (2011), ‘Has India’s Economic Growth Become More Pro-Poor in the Wake of Economic Reforms?’. *The World Bank Economic Review*, 25: 157–89.
- Deaton, Angus (2003). ‘Adjusted Indian Poverty Estimates for 1999–2000’. *Economic and Political Weekly*, 322–6.
- Deaton, Angus, and Alessandra Tarozzi (1999). ‘Prices and Poverty in India’. Princeton University Mimeo, 13 December. Available at: [http://siteresources.worldbank.org/INTPGI/Resources/13236\\_Prices\\_and\\_poverty\\_in\\_India.pdf](http://siteresources.worldbank.org/INTPGI/Resources/13236_Prices_and_poverty_in_India.pdf) (accessed December 2018).
- Desai, Sonalde, and Veena Kulkarni (2008). ‘Changing Educational Inequalities in India in the Context of Affirmative Action?’. *Demography*, 45: 245–70.
- Ferreira, Francisco H.G., and Jérémie Gignoux (2011). ‘The Measurement of Inequality of Opportunity: Theory and an Application to Latin America’. *Review of Income and Wealth*, 4: 622–57.
- Government of India (2012–15a). Consumer Price Index for Agricultural Labourers in India. Ministry of Labour. Available at: <https://data.gov.in/catalog/consumer-price-index-agricultural-labourers-and-rural-labourers> (accessed December 2018).
- Government of India (2012–15b). Consumer Price Index for Industrial Workers in India. Ministry of Labour. Available at: <https://data.gov.in/catalog/consumer-price-index-industrial-workers> (accessed December 2018).
- ICRISAT (n.d.). Village Dynamics in South Asia (VDSA), Meso Database. Available at <http://vdsa.icrisat.ac.in/vdsa-database.aspx> (accessed January 2019).
- Lanjouw, Peter, and Rinku Murgai (2009). ‘Poverty Decline, Agricultural Wages, and Non-Farm Employment in India: 1983–2004’. *Agricultural Economics*, 40: 243–63.
- Narayan, Ambar, Roy Van der Weide, Alexandru Cojocaru, Christoph Lakner, Silvia Redaelli, Daniel Gerszon Mahler, Rakesh Gupta N. Ramasubbaiah, and Stefan Thewissen (2018). *Fair Progress?: Economic Mobility Across Generations around the World*. Washington, DC: World Bank Group.
- Nickell, Stephen (1981). ‘Biases in Dynamic Models with Fixed Effects’. *Econometrica: Journal of the Econometric Society*, 49(6): 1417–26.
- NSSO (1983). ‘National Sample Survey of India, 38th round’. New Delhi: National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.

- NSSO (1988). 'National Sample Survey of India, 43th round'. New Delhi: National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.
- NSSO (1994). 'National Sample Survey of India, 50th round'. New Delhi: National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.
- NSSO (2000). 'National Sample Survey of India, 55th round'. New Delhi: National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.
- NSSO (2005). 'National Sample Survey of India, 61st round'. New Delhi: National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.
- NSSO (2012). 'National Sample Survey of India, 68th round'. New Delhi: National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.
- RBI (1998). 'Banking Statistics, 1972–1995'. Available at <https://rbidocs.rbi.org.in/rdocs/Publications/PDFs/bs72-95.pdf>, Accessed on 11 January 2019.
- RBI (2000). 'Statistical Tables Relating to Banks in India, 1999–2000'. Available on <https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Statistical+Tables+Relating+to+Banks+in+India> (accessed January 2019).
- RBI (2005). 'Statistical Tables Relating to Banks in India, 2004–2005'. Available on <https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Statistical+Tables+Relating+to+Banks+in+India> (accessed January 2019).
- RBI (2012). 'Statistical Tables Relating to Banks in India, 2011–2012'. Available on <https://rbi.org.in/Scripts/AnnualPublications.aspx?head=Statistical+Tables+Relating+to+Banks+in+India> (accessed January 2019).
- Roemer, John E. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- STICERD and LSE (2005–19). EOPP Indian States Database. London School of Economics. Available at [http://sticerd.lse.ac.uk/eopp/\\_new/data/indian\\_data/default.asp](http://sticerd.lse.ac.uk/eopp/_new/data/indian_data/default.asp) (accessed January 2019).
- Van der Weide, Roy, and Branko Milanovic (2018). 'Inequality Is Bad for Growth of the Poor (But Not for That of the Rich)'. *The World Bank Economic Review*, 32: 507–30.
- World Bank (2018). *India Systematic Country Diagnostic: Realizing the Promise of Prosperity*. Washington, DC: World Bank Group. Available at: <http://documents.worldbank.org/curated/en/629571528745663168/India-Systematic-country-diagnostic-realizing-the-promise-of-prosperity> (accessed December 2018).

## Appendix A: Data construction

### A1 State-region observation units

We constructed the state-region identifiers following Lanjouw and Murgai (2009), where the definitions used in the 55th NSS round (1999–2000) are used to define state-regions. In order to harmonize this with earlier NSS rounds (1983, 1987–88, and 1993–94), we created state identifiers for the newly created states of Jharkhand, Chhattisgarh, and Uttaranchal, using the appropriate regions in the states of Bihar, Madhya Pradesh, and Uttar Pradesh, respectively. Due to the creation of the new states, the region identifiers of Bihar, Madhya Pradesh, and Uttar Pradesh have to be adjusted in the NSS rounds of 2004–05 and 2011–12. In addition, the state identifiers were also adjusted to those used in the 1999–2000 round for all the other survey rounds. For all rounds, Goa and Daman and Diu remained pooled together because it is not possible to separate them for the survey rounds before 1999–2000.

As a result of the above-mentioned harmonization, 77 state-region units were created. However, some of these units contained only a few hundred households, which is deemed too few given our interest in the different percentiles of the distribution of economic growth. Therefore, we merged the following smaller state-region units keeping in mind geographic considerations:

- The two Inland Southern regions (66th round region name) of Andhra Pradesh
- The states of Arunachal Pradesh and Nagaland
- The Eastern Plains and Cachar Plain regions (66th round region name) of Assam
- The states of Goa, Daman and Diu, Dadra and Nagar Haveli, and the Coastal region of Maharashtra state
- The two South Eastern regions (66th round region name) of Gujarat
- The Saurashtra and Dry Areas regions of Gujarat
- All two regions of Haryana
- All regions of Jammu and Kashmir (because of inconsistencies in the state-region areas covered by the surveys)<sup>8</sup>
- The Coastal and Ghats, Inland Eastern and Inland Southern regions of Karnataka
- The Central and Northern regions of Madhya Pradesh
- The South and South Western regions (66th round region name) of Madhya Pradesh
- The Inland Eastern and Eastern regions of Maharashtra
- All two regions of Manipur
- All two regions of Orissa
- The state of Chandigarh and the Southern region of Punjab
- The Southern and South Eastern regions of Rajasthan
- The state of Sikkim and the Himalayan region of West Bengal
- The Central and Southern regions of Uttar Pradesh
- The state of Pondicherry and the Coastal Northern region of Tamil Nadu

In addition, we removed observations for the island state of Lakshadweep due to too few observations. After these adjustments, we were left with 54 state-region units.

---

<sup>8</sup> Notably, there are no observations for Jhelum Valley in Jammu and Kashmir in the 1993–94 round.

## A2 Consumer price index

The consumer price index (CPI) is used to correct for differences in prices between urban and rural areas, different states within the survey rounds, and inflation between the survey rounds. We obtained information on the CPI from different data sources for the major states, which we combined to calculate the CPI with a reference period of 2004–05 for rural all India.

The main data sources are the time series of the CPI for agricultural labourers (CPI-AL) obtained for the period of 1982–83 to 2012–13, and the time series of the CPI for industrial workers (CPI-IW) obtained for the period of 1994–95 (or 2001–02) to 2011–12. These series are available on a yearly basis for the major states. However, they can only be compared across time and not across states and urban/rural areas.

To create spatial comparability of the CPI series, we use the poverty lines for 2004–05 and calculate the relative prices compared with the urban all-India poverty line. The obtained ratios are used to normalize the CPI-AL and CPI-IW series. The obtained (comparable) CPI values are used for rural areas in the 1983, 1987–88, 1993–94, 2004–05, 2009–10, and 2011–12 survey rounds, and urban areas of the 2004–05, 2009–10, and 2011–12 survey rounds.

In order to obtain CPI values for urban areas for the 1983, 1987–88, and 1993–94 survey rounds, we consulted two other data sources. The Planning Commission provided the average monthly per capita consumption expenditure (MPCE) at current and constant prices (base 1993–94) for urban and rural areas of major states for periods including 1983, 1993–94, and 2004–05. The 2004–05 series provides a good match with the CPI-AL and CPI-IW series, and these values are used to calculate the CPI for urban areas for the 1983 and 1993–94 survey rounds (in the major states). In order to obtain urban CPI for the 1987–88 round, the study by Deaton and Tarozzi (1999) was consulted, which provide the price factors between the 1987–88 and 1993–94 survey rounds (see Deaton and Tarozzi 1999: 41).

These data sources make it possible to adjust for inflation between the survey rounds. In order to correct for inflation within the survey rounds by quarter of survey, the CPI values for the year following the survey rounds are approximated using log-linear approximation where necessary (urban areas). The quarterly CPI values are then approximated using log-linear approximation for all survey rounds.

These procedures provide us with the CPI for the major states. We use the CPI values of the neighbouring major states to approximate the CPI of the smaller states (where historical CPI data is not available) for all rounds except the 2011–12 round. The following rules are followed:

- Jammu and Kashmir: values of Himachal Pradesh are used (urban for urban, rural for rural areas)
- Chandigarh: urban values of Punjab are used for both urban and rural areas
- Uttaranchal: values of Himachal Pradesh are used (urban for urban, rural for rural areas) formerly part of Uttar Pradesh, but Himanchal Pradesh is also more mountainous just like Uttaranchal
- Delhi rural: values of urban Delhi are used
- Sikkim: values of urban West Bengal are used for both urban and rural areas
- Arunachal Pradesh, Nagaland, Manipur, Tripura and Meghalaya: values of Assam are used (urban for urban, rural for rural areas)
- Jharkhand: values of Bihar are used (urban for urban, rural for rural); formerly part of Bihar

- Chhattisgarh: values of Madhya Pradesh are used (urban for urban, rural for rural areas); formerly part of Madhya Pradesh
- Goa, Daman and Diu, and Dadra and Nagar Haveli: values of Maharashtra are used (urban for urban, rural for rural areas)
- Pondicherry and Andaman and Nicobar: values of Tamil Nadu are used (urban for urban, rural for rural areas).

### **A3 Industry and occupational codes**

Three-digit industry and occupation codes are provided following the National Industrial Classification and National Classification of Occupations. In the last three decades several changes have occurred to these registers, which had to be harmonized among the survey rounds. Because we are only interested in broad levels of classifications, we only harmonized the codes at the two-digit level, and grouped them into the following categories for industry:

- Primary sector (agriculture)
  - Agriculture
- Secondary sector (industry)
  - Mining and quarrying
  - Manufacturing
  - Public works
  - Construction
- Tertiary sector (services)
  - Trade and repair
  - Restaurants and hotels
  - Transport and storage
  - Information technology, communications and post
  - Financial, legal, and related services
  - Public administration and defence
  - Education, science, and research
  - Health and social work
  - Other services

and the following categories for occupations:

- Administrators, managers
- Professionals, associated professionals
- Clerks
- Sales and service workers
- Skilled agriculture and fisheries workers
- Craftsmen, machine operators
- Labourers, unskilled workers

## Appendix B: Additional results

### B1 Ordinary least square (OLS) regression results

#### *Intergenerational mobility and inclusive growth*

Table B1: Determinants of real per capita consumption expenditure growth at different percentiles (OLS regression: Model 1)

	10%	25%	50%	75%	90%	95%
AR(1)	−0.0760*** (0.0095)	−0.0653*** (0.0091)	−0.0563*** (0.0089)	−0.0496*** (0.0092)	−0.0468*** (0.0112)	−0.0502*** (0.0126)
Upward mobility	0.0007 (0.0005)	0.0004 (0.0004)	0.0003 (0.0004)	0.0003 (0.0004)	0.0001 (0.0004)	0.0001 (0.0005)
Inequality in RMPCE	−0.0258 (0.0479)	0.0005 (0.0417)	0.0105 (0.0363)	−0.0259 (0.0329)	−0.0721* (0.0372)	−0.0957** (0.0442)
Share of age 16–64 years	0.0941* (0.0510)	0.0712 (0.0457)	0.0743* (0.0435)	0.1406*** (0.0477)	0.1882*** (0.0520)	0.2102*** (0.0533)
Years of education	−0.0011 (0.0016)	−0.0006 (0.0014)	−0.0005 (0.0014)	0.0001 (0.0014)	0.0004 (0.0016)	0.0005 (0.0016)
Services industry	0.0904*** (0.0277)	0.0816*** (0.0258)	0.0848*** (0.0266)	0.0580** (0.0254)	0.0547* (0.0286)	0.0570* (0.0298)
Manufacturing industry	0.0668*** (0.0212)	0.0595*** (0.0195)	0.0520*** (0.0197)	0.0492** (0.0207)	0.0613** (0.0240)	0.0739*** (0.0249)
Access to electricity	−0.0124 (0.0083)	−0.0117 (0.0077)	−0.0106 (0.0069)	−0.0118 (0.0074)	−0.0144* (0.0076)	−0.0147* (0.0088)
Log bank per capita	0.0036 (0.0032)	0.0034 (0.0025)	0.0033 (0.0020)	0.0039* (0.0021)	0.0028 (0.0020)	0.0033* (0.0020)
Share of urban	−0.0401*** (0.0145)	−0.0351*** (0.0130)	−0.0320*** (0.0122)	−0.0210* (0.0126)	−0.0147 (0.0150)	−0.0068 (0.0148)
Share of Christians	0.0405*** (0.0122)	0.0355*** (0.0103)	0.0286*** (0.0101)	0.0231** (0.0100)	0.0221** (0.0106)	0.0234** (0.0109)
Observations	268	268	269	269	269	269
Adjusted $R^2$	0.417	0.437	0.483	0.529	0.574	0.594

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region fixed effects (FEs) (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011) and Government of India (2012–15a, 2012–15b).

Table B2: Determinants of real per capita consumption expenditure growth at different percentiles (OLS regression: Model 2)

	10%	25%	50%	75%	90%	95%
AR(1)	-0.0860*** (0.0115)	-0.0686*** (0.0104)	-0.0654*** (0.0110)	-0.0556*** (0.0106)	-0.0482*** (0.0106)	-0.0601*** (0.0126)
Upward mobility	0.0009 (0.0006)	0.0006 (0.0004)	0.0004 (0.0004)	0.0004 (0.0004)	0.0002 (0.0005)	0.0004 (0.0006)
Inequality in RMPCE	-0.0821 (0.0502)	-0.0560 (0.0366)	-0.0536 (0.0343)	-0.0510 (0.0315)	-0.0936*** (0.0346)	-0.1232*** (0.0431)
Share of age 16–64 years	0.1385** (0.0637)	0.1140* (0.0594)	0.1483** (0.0584)	0.1978*** (0.0597)	0.2411*** (0.0680)	0.2682*** (0.0708)
Years of education	-0.0019 (0.0021)	-0.0011 (0.0018)	-0.0010 (0.0020)	-0.0007 (0.0018)	0.0005 (0.0020)	0.0016 (0.0022)
Services industry	0.1511*** (0.0494)	0.1291*** (0.0462)	0.1388*** (0.0499)	0.1179*** (0.0438)	0.0951* (0.0502)	0.0687 (0.0584)
Manufacturing industry	0.0550** (0.0250)	0.0441* (0.0225)	0.0380 (0.0239)	0.0236 (0.0244)	0.0237 (0.0273)	0.0454 (0.0299)
Access to electricity	-0.0194 (0.0129)	-0.0192 (0.0118)	-0.0149 (0.0107)	-0.0181 (0.0113)	-0.0110 (0.0111)	-0.0089 (0.0125)
Log bank per capita	0.0062 (0.0088)	0.0067 (0.0073)	0.0073 (0.0078)	0.0086 (0.0079)	0.0020 (0.0029)	0.0016 (0.0030)
Share of urban	-0.0283 (0.0241)	-0.0236 (0.0204)	-0.0306 (0.0214)	-0.0138 (0.0209)	-0.0096 (0.0223)	0.0147 (0.0254)
Share of Christians	0.0001 (0.0399)	-0.0009 (0.0367)	-0.0028 (0.0386)	0.0090 (0.0336)	0.0149 (0.0351)	0.0298 (0.0370)
Market access	-0.0071 (0.0141)	-0.0037 (0.0116)	0.0029 (0.0124)	0.0059 (0.0129)	0.0155 (0.0140)	0.0176 (0.0162)
Government expenditure per GDP	0.0266 (0.0329)	0.0322 (0.0294)	0.0434 (0.0313)	0.0279 (0.0300)	0.0045 (0.0314)	-0.0046 (0.0394)
Voter turnout last election	0.0001 (0.0002)	0.0000 (0.0001)	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0003* (0.0002)	-0.0001 (0.0002)
Political competition	0.0028 (0.0072)	0.0044 (0.0064)	0.0009 (0.0065)	0.0006 (0.0063)	-0.0033 (0.0068)	-0.0046 (0.0073)
Observations	198	198	198	198	200	200
Adjusted $R^2$	0.488	0.518	0.547	0.572	0.616	0.614

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and (STICERD and LSE 2005–19).

*Intergenerational mobility and human capital*

Table B3: Determinants of change in boys' years of education (20–25 years) by father's education (OLS regression: Model 1)

	Illiterate	Below primary	Primary	Middle	High and above
AR(1)	−0.0791*** (0.0171)	−0.1348*** (0.0133)	−0.1441*** (0.0120)	−0.1503*** (0.0130)	−0.1747*** (0.0113)
Upward mobility	0.0021 (0.0054)	0.0131** (0.0053)	0.0096** (0.0039)	0.0057 (0.0039)	−0.0004 (0.0031)
Inequality in RMPCE	−0.1283 (0.3070)	−0.0680 (0.3840)	−0.1110 (0.3793)	−0.1544 (0.3406)	−0.5686** (0.2884)
Share of age 16–64 years	1.0895** (0.4378)	−0.0080 (0.5958)	−0.2673 (0.5031)	0.9950** (0.4703)	0.3373 (0.3783)
Years of education	0.0440* (0.0247)	0.0575*** (0.0211)	0.0721*** (0.0182)	0.0170 (0.0153)	0.0025 (0.0117)
Services industry	−0.1394 (0.3091)	−0.0119 (0.3710)	−0.4938** (0.2504)	0.1924 (0.2863)	0.0694 (0.2126)
Manufacturing industry	−0.1803 (0.2161)	−0.3137 (0.2598)	−0.2268 (0.2529)	−0.1448 (0.2727)	−0.1302 (0.1734)
Access to electricity	0.0644 (0.0715)	0.0591 (0.0957)	0.0808 (0.0654)	−0.0104 (0.0882)	0.0452 (0.0669)
Log bank per capita	−0.0424*** (0.0156)	−0.0207 (0.0131)	0.0037 (0.0144)	−0.0527* (0.0297)	−0.0091 (0.0200)
Share of urban	−0.2800* (0.1644)	−0.0865 (0.1915)	0.1647 (0.1410)	−0.1372 (0.1369)	0.2432** (0.1150)
Share of Christians	−0.1353 (0.0951)	−0.2300** (0.1105)	−0.2082** (0.0941)	−0.0429 (0.0802)	−0.1808** (0.0840)
Observations	270	270	269	270	270
Adjusted $R^2$	0.217	0.364	0.459	0.432	0.631

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), and Government of India (2012–15a, 2012–15b).

Table B4: Determinants of change in boys' years of education (20–25 years) by father's education (OLS regression: Model 2)

	Illiterate	Below primary	Primary	Middle	High and above
AR(1)	−0.0993*** (0.0161)	−0.1368*** (0.0169)	−0.1474*** (0.0158)	−0.1541*** (0.0155)	−0.1629*** (0.0129)
Upward mobility	0.0079 (0.0056)	0.0096 (0.0071)	0.0050 (0.0048)	0.0096* (0.0054)	−0.0035 (0.0040)
Inequality in RMPCE	−0.4103 (0.3566)	−0.7753* (0.4641)	0.3406 (0.4132)	−0.5086 (0.4600)	−0.4761 (0.3342)
Share of age 16–64 years	1.0980** (0.5367)	0.8792 (0.8787)	−0.5958 (0.6711)	1.0540 (0.7040)	0.4748 (0.5169)
Years of education	0.0587** (0.0271)	0.0398 (0.0271)	0.0651** (0.0256)	−0.0053 (0.0234)	−0.0110 (0.0168)
Services industry	−0.3072 (0.4361)	−0.0410 (0.6091)	−0.8289* (0.4532)	0.0269 (0.5208)	0.4213 (0.4058)
Manufacturing industry	−0.0634 (0.2254)	−0.0952 (0.3063)	−0.3569 (0.3303)	0.1719 (0.3256)	0.0319 (0.1975)
Access to electricity	0.1695* (0.0880)	0.0677 (0.1287)	0.2841*** (0.1028)	0.0307 (0.1440)	0.1194 (0.0956)
Log bank per capita	−0.0473* (0.0274)	−0.0533*** (0.0192)	0.0091 (0.0249)	−0.0220 (0.0152)	−0.0463*** (0.0124)
Share of urban	−0.1754 (0.2427)	−0.2321 (0.2850)	0.1627 (0.2661)	−0.0168 (0.2651)	−0.2097 (0.2173)



Share of Christians	-0.2022 (0.3369)	0.0013 (0.3769)	0.1814 (0.3918)	0.4262 (0.3494)	0.0871 (0.2832)
Market access	0.2202* (0.1236)	0.2204 (0.1627)	0.2279 (0.1397)	0.3256** (0.1586)	0.1851 (0.1145)
Government expenditure per GDP	-0.4448 (0.5082)	0.6170 (0.5399)	0.8223 (0.5056)	-0.4838 (0.5245)	-0.3523 (0.3179)
Educational expenditure per GDP	1.8606 (2.1765)	-2.6199 (2.4318)	-3.0691 (2.1881)	-0.4000 (2.1063)	1.6819 (1.2757)
Voter turnout last election	-0.0004 (0.0014)	0.0018 (0.0018)	-0.0036** (0.0017)	-0.0022 (0.0015)	-0.0022* (0.0012)
Political competition	0.0414 (0.0650)	0.1003 (0.0760)	-0.0607 (0.0806)	0.0613 (0.0775)	-0.0344 (0.0562)
Observations	200	200	200	200	200
Adjusted $R^2$	0.226	0.327	0.439	0.433	0.475

Notes: Heteroskedasticity robust standard errors reported (\* $P=0.10$ , \*\* $P=0.05$ , \*\*\* $P=0.01$ ). Regressions control for year and region FEs (not reported). AR(1) stands for lagged dependent variable in levels. Upward mobility for co-residing boys aged 20–25 years.

Source: Authors' compilation based on data from NSSO (1983, 1988, 1994, 2000, 2005, 2012), RBI (1998, 2000, 2005, 2012), Census of India (2001, 2011), Government of India (2012–15a, 2012–15b), ICRISAT (n.d.), and (STICERD and LSE 2005–19).