INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD • INDIA

Research and Publications

Is the Past Still Holding Us Back? A Study on Intergenerational Education Mobility in India

P K V Kishan

W. P. No. 2018-01-03
January 2018

INDIAN INSTITUTE OF MANAGEMENT
AHMEDABAD-380 015
INDIA

Is the Past Still Holding Us Back? A Study on Intergenerational Education Mobility in India

P K V Kishan
Economics Area
Indian Institute of Management, Ahmedabad
pkvkishan@iima.ac.in

Abstract

This paper deals with intergenerational education mobility in India. We make use of IHDS-II (2011-12) and prepare a representative dataset that goes beyond co-resident pairs of son-father by utilizing a unique feature of IHDS data. From the resulting sample of 44,532 son-father pairs and appropriate cohort analysis, we find that there is still a high degree of intergenerational persistence in education, although the same is decreasing steadily over time since independence. Through quantile regression, we find that sons are most mobile at the top end of the education distribution. Finally, "Higher Inequality (during fathers' generation) → Lesser Mobility" nexus in education plays out for the Indian scenario and thus corroborates the 'Great Gatsby Curve'.

***Keywords*: Intergenerational education mobility, quantile regression, co-residence, Great Gatsby Curve**

1. Introduction

Inequality in a society can be roughly ascribed two reasons – one, a disparity in efforts of individuals or category of individuals, and two, differences in predetermined circumstances outside the locus of control of individuals. While the former is essential in society to promote merit and provide incentives for individuals to work hard; the latter is unfair as it manifests into inequality of opportunity where the life chances of an individual are determined by the relative socio-economic status of his/her parents or ancestors. This intergenerational persistence or lack of intergenerational mobility is said to be prevalent in a society if the same group(s) of the population continue to remain trapped at low levels of education or maintain low social status or remain in relative poverty as compared to other sections of the population for more than one generation.

Mobility is an essential marker for growth and development of a society. A child whose life chances are determined by the social or economic strata he/she belongs to and not by his/her industry in a rigid society is bound to have no inducement to try and wiggle out of the low-level equilibrium and contribute to nation's progress. This has been articulated to a similar effect by Bourguignon et al. (2007) who suggested that in a society where the poor and the rich (and their respective children) are equally likely to succeed, people have a higher incentive to work hard.

Our objective in this paper is to check whether intergenerational mobility has materialized in India by focusing on the correlation of educational attainments between parents and children (specifically, of fathers and sons) as it has been reasoned in several quarters apropos the transmission of opportunity from parents to children through the channel of education. For example, as per Becker et al. (1990), maternal education can improve the efficiency of human capital production across generations leading to increasing returns. Other direct and indirect channels include – parents' schooling impacting children's schooling as the educated parents value their children's education more (cultural dependency), parental education influencing children's income, educated parents being capable of supporting their children in their studies (teaching practices), and the relation between "parental economic ability" (owing to their educational attainment) and their children's education (Hertz et al., 2007; Bussolo, Checchi & Peragine, 2017).

There are other cogent reasons behind our choice of education as a medium to study intergenerational mobility. Education is less prone to errors than income in terms of measurement, and in case of developing countries data on educational outcomes/attainment is mostly available unlike data on earnings, which is not reported in a typical household survey (Azam & Bhatt, 2015). Also, once an individual reaches mid-twenties, formal education gets fixed. This precludes life cycle biases which occur in case of income measurement as they are volatile and age-dependent (Hertz et al., 2007; Azam & Bhatt, 2015). It is also cited in Hertz et al. (2007) that even though the "*main role of education is to promote social mobility*", however, at the same time "education is also *the main vehicle of social reproduction*" (Ganzeboom, Treiman, & Ultee, p. 284). Thus, the examination of equality or inequality of opportunity through the lenses of education makes for an interesting study and a relevant study.

The rest of the paper is organized as follows. In section two, we review the literature on the subject with a particular focus on studies in the Indian context. Section three contains description on data, choice of variables, and the empirical strategy. The results are presented in section four. And finally, in section five, we summarize our findings and conclude.

2. Review of Literature

A high degree of variation in children's outcomes emerges even before they enter the labour market as averred by Chetty et al. (2014) and hence it becomes crucial to identify the factors that affect the children while they are growing up which drive the differences in mobility across different social groups and wealth classes. For the American population, in terms of spatial mobility, Chetty et al. (2014) identify those factors to be racial shares at the community level, urban sprawl (as characterized by commute times to work), income inequality, school system (as measured by test scores, dropout rates, class sizes), social capital (which are proxies of strength of social networks and community involvement in an area) and family structure (fraction of single parents in the area, parents' marital status).

One of the procedures for measuring intergenerational persistence involves calculating the correlation between the socioeconomic status of parents and their adult children. The sign and magnitude of these correlations can help evaluate a society's success or failure to provide equality of opportunity to children from various family backgrounds based on the rate of transmission of inter-personal equality (Hertz et al., 2007). The authors take up the study of status persistence/mobility to assess international differences and trends, a gap hitherto unfilled as most studies concentrated on these statistics, region-wise. Hertz et al. (2007) estimate 50-year trends in intergeneration persistence by one, regressing children's schooling against that of their parents and two, calculating the correlation of educational attainment between the parents and children, for 42 countries (a mix of developed and developing countries). The regression coefficient as the predictor of persistence of parents' education on the next generation demonstrates that the impact of parents' fortune on children's outcomes has decreased over time. On the other hand, the correlation coefficient doesn't suggest the same. On an average, for the countries in the sample, the intergenerational correlation coefficient has held steady for a century as per the study. While the Latin American countries displayed the highest persistence, the Nordic countries stood out for their relatively higher measure of mobility.

2.1 The Indian Setting

India's economic growth since the 1980s has been concurrent with increasing inequalities in outcomes and consequently raises a concern of whether it reflects inequalities in opportunities in the society. As it is, the Indian society is deeply stratified by caste and beset by poor learning outcomes and very low mobility (Gupta, 2004). And, this lack of mobility as Maitra and Sharma (2009) contend, excludes many parts of our society from reaping rewards of the prolific levels growth the country has experienced during the last two decades.

The empirical literature on intergenerational persistence/mobility in India is scarce (Maitra & Sharma, 2009; Hnatkovska, Lahiri & Paul, 2013; Emran and Shilpi, 2015). The earliest paper in this regard is by Jalan and Murgai (2008) who use two rounds of National Family Health Survey (NFHS) in 1992-93 and 1998-99 to study inequalities in educational attainments and its persistence across generations for different groups of population in India. The results reflect significant and consistent improvements in education mobility and decreasing education gaps between various caste groups once other characteristics are controlled for. The gaps, though, exist between the rich and the poor characterized by the differences in intergenerational education elasticity across various wealth classes as defined by the authors. However, the authors advise against phasing out of positive discrimination policies as they argue that “caste remains a useful proxy for targeting the poor because of its close association with low wealth” (p. 10).

Maitra and Sharma (2009) examine the intergenerational transmission of human capital by analyzing the role of two aspects of parents' education on child's educational attainment – the number of years of schooling and progression across different levels of schooling. Using data from Indian Human Development Survey (IHDS) 2005, the authors account for the potential endogeneity of parents' educational attainment being correlated with some of the unobserved determinants of schooling of their children. They do so by employing instrumental variables estimation. The findings affirm the results obtained by Jalan and Murgai (2008). One, there has been a significant increase in educational attainment over the last few decades and that the influence of parents' education on the education of their children is little, stressing upon the fact that public investments in education matter much more over private investments. Two, the sequential probit analysis of the school progression shows that while mother's education is an important determinant at the start of children's schooling and the middle school, father's education becomes crucial in the decision for the child to continue beyond post-secondary levels. Among other results, one that stands out is for more educated mothers; the average educational attainment of daughters is higher than that of sons underlining the popular argument that education of mother is one of the most important ways to alleviate gender gap in education.

National Sample Survey (NSS) data has also been the choice of database for a few mobility studies (Majumder, 2010; Hnatkovska, Lahiri & Paul, 2013). The latter uses data in the time span of 1983 and 2005 from successive rounds of NSS Employment-Unemployment surveys to compare the intergenerational mobility rates of scheduled castes (SC) and scheduled tribes (ST) against the non-SC/ST households in terms of educational outcomes, occupation choices and wages. The authors posit that the co-resident nature of Indian households is helpful in getting around the shortcoming of lack of panel studies on families and hence useful for inferring intergenerational mobility patterns in other such developing countries as well. The results indicate that convergence has taken place in the intergenerational mobility rates, more so in case of educational attainment and wages, between the SC/STs and the others; although the non-SC/STs are still more likely to work in a different profession than their parents as compared to their SC/ST counterparts and their parents. Another key finding here is that the mobility improvement in education and income has occurred for both low and highly educated/ high-income households among SC/STs. The authors attribute these improvements in mobility rates to the structural changes in India during the period of research.

In a panel data analysis of rural India, Mohammed (2016) examines intergenerational economic mobility by using the income of sons and fathers as a proxy for their respective economic status. The panel is composed of the initial data of the households surveyed by National Council of Applied Economics Research (NCAER) in the Human Development Profile of India (HDPI) 1993-94 along with the data on the same households present in IHDS-I and IHDS-II. NSSO data is also employed to corroborate the findings on between-socioeconomic group differences in economic mobility. The results indicate an intergenerational economic persistence ranging from 0.28 to 0.37, slightly comparable to the education mobility estimates obtained in Jalan and Murgai (2008). Further, the author finds a relatively higher measure of within-group mobility in case of disadvantaged groups while also detecting low rates of convergence between groups. A couple of more studies have made use of IHDS-I data.

An enquiry by Azam and Bhatt (2015) produces results that are different to those obtained in Jalan and Murgai (2007) and Maitra and Sharma (2009). To tell apart the discrepancy in results, Azam and Bhatt (2015) decompose the intergenerational correlation on fathers' educational distribution and find that whilst mobility has improved at the lower end of the fathers' educational distribution, it has declined at the top end of the distribution, thus giving rise to a neutral trend in the overall correlation between fathers and sons education. The authors also stress upon the salience of IHDS data to circumvent the limitation of the co-residence condition in other studies that use data from either NHFS or NSS. Owing to the comparability of IHDS data to the data set used by Hertz et al. (2007), the authors rank the average intergenerational correlation coefficient obtained for India against 42 nations following Hertz et al. (2007). While India fares better than Latin American countries in terms of educational mobility, its intergenerational educational persistence is higher than the World average and expectedly does worse than Western and Eastern Europe. An earlier study by Motiram and Singh (2012) complements these results with its take on intergenerational occupational mobility in India. Contrary to the findings on persistence in educational attainments in the studies discussed so far and the study on occupational mobility by Hnatkovska, Lahiri and Paul (2013), this paper detects considerable occupational persistence, although with differences across occupational categories. Proponents of affirmative action would have fodder for thought as downward mobility and persistence in low-skilled/low-paying occupations for SC/STs is higher as compared to the same for non-SC/STs. Majumder (2010) comes up with similar findings from his study based on two rounds (1993 and 2004) of NSSO data. In this study, Majumder (2010) finds intergenerational stickiness, even in educational attainment along with occupational distribution, although to a lesser extent in educational achievements, for SC/STs; pointing towards possible discrimination in the labour market.

One study that stands out in its methodology, as well as results, is by Emran and Shilpi (2015). While most studies preceding this one have mostly reported considerable improvements in educational mobility, the measures used in this study, namely Sibling Correlation (SC) and Intergenerational Correlation (IGC), for similar age cohorts as other studies, suggest strong intergenerational persistence in education; the only beneficiaries being the urban women, especially the low caste ones. This is the second paper after Jalan and Murgai (2008) to have used NFHS data (1992-93 and 2006 rounds) but owing to the use of SC and IGC as measures, which the authors aver (with support from literature) are more robust and informative than Intergenerational Regression Coefficient (IGRC), report evidence that "paints a more sober picture" (Emran and Shilpi, 2015, p. 362). Even after 15 years of liberalization, the sibling and intergenerational coefficients have declined only marginally and indicate more adverse equality of opportunity as compared to Latin American countries and other Asian countries. When accounted for neighbourhood fixed effects, geographic location emerged as an important factor in the measurement of sibling correlation and intergenerational correlation. For example, the SC is higher for urban men as compared to rural men, indicating a higher inequality in opportunities for the urban men. There is no such difference between rural and urban women, but the gender gaps in both measures for rural areas are substantial with the value of SC being much higher for rural women as compared to rural men, despite moderate improvements in mobility among women.

In a later study, Emran, Greene and Shilpi (2017) illustrate the merit behind the preference of IGC over IGRC, especially in the context of developing countries where data limitations restrict the study of the phenomenon to co-resident samples only. Using a simple model of truncation[1] followed up by an empirical exercise on household surveys of India and Bangladesh, the authors find a significant downward bias in IGRC as compared to IGC when the sample consideration is reduced to co-resident cases. According to the study, the downward truncation bias in IGRC

[1] Truncation in the data is due to children leaving their parental household post marriage.

(when compared to the bias in IGC) is inversely proportional to the extent of co-residency rates observed in the data. From this, it can be inferred that IGRC remains a sizably robust measure of interpreting intergenerational mobility in either of the following two situations. One, the co-residency rates are high in the population. Two, parent-child pairs bearing the requisite information (on education, income, or occupation) can be created irrespective of whether they are co-resident in a household.

From the literature discussed so far, it is clear that the studies on intergenerational educational mobility in India have differed over the choice of measures, choice of age cohorts, and selection of data sources. Although a consensus doesn't emerge, some studies agree upon improvements in educational mobility in India and attribute various reasons to the process ranging from structural changes because of liberalization to the success of positive discrimination policies. We take this investigation further as a part of this study.

3. Data, Variables, and Empirical Framework

Broadly, intergenerational mobility can be analyzed by one of the following ways – one, by assessing variables across a repeated cross-section of the population; two, by measuring the variables over the life courses of sampled individuals; and three, by measuring the variables across age-cohorts (Bussolo, Checchi & Peragine, 2017). Since longitudinal datasets are conspicuous by their absence in India, we are left with options one and three. In such cases, the following model (which we too shall take up) is usually estimated to analyze intergenerational mobility in the specific case of education –

$$S_i = \beta_0 + \beta_1 F_i + (\text{Controls}) + \epsilon_i$$

where, S_i denotes the number of years of schooling of the i^{th} son, F_i is the primary circumstance variable in this study that stands for father's education in terms of his completed years of schooling, and ϵ_i encapsulates the unobserved elements such as ability or/and effort of the individual. β_1 is the main variable of interest and is termed Intergenerational Regression Coefficient (IGRC). β_1 essentially captures the sensitivity of the expected educational outcome of the sons to unit changes in the educational attainment of the fathers. It conveys how strongly past circumstances affect the educational attainment of the son and in turn, his life chances. A value of zero denotes perfect mobility, and a value of one means perfect rigidity.

The cross-sectional database based studies (one and three) are further divided in their adoption of either co-resident household approach or two-sample instrumental variables approach (Mohammed, 2016). The three major sample surveys in India – NSSO, NFHS, and IHDS – amply facilitate co-resident household approach as numerous studies based on each of them in the context of intergenerational mobility in India have shown. However, consideration of only co-resident son-father pairs might generate attenuation bias as cohabitation might be systematically linked to decisions regarding human capital investments in a household. Moreover, as pointed out by Motiram and Singh (2012), we would be missing out on single-member households, two-member households consisting of husband and wife, and nuclear families (husband, wife, and children), which would by itself lead to a substantial loss in observations. Further, the nuclear family structure, which characterizes a significant proportion of the urban-middle-class households in the contemporary age and time, would be grossly underrepresented in a 'co-resident only' sample.

3.1 Choice of Database and the Basic Model

As evident by the review of the literature, there has been a sizeable number of studies in the Indian context to have studied intergenerational mobility in education, income, and occupation. Most of these have quantified the extent of stickiness (in the three variables) between two generations. Through this study, we check as well as update the numbers (for the trends in intergenerational education mobility) with the help of the latest data. Additionally, we apply quantile regression method to assess the difference in effects of fathers' education on the distribution of sons' educational attainments. We thus intend to explore the possible non-linearity in the relationship between educational outcomes of successive generations. Our choices of method and database were influenced by the considerations on data constraints and limitations stated in the earlier two paragraphs. Thus, to examine the evolution of and trends in intergenerational education mobility, we resort to comparing β_1 across age-cohorts for – (i) Overall sample, (ii) Social classes, and (iii) Religions. Next, we shall use the data from the second round of the Indian Human Development Survey (IHDS-II) conducted in 2011-12. IHDS is a collaborative project between National Council of Applied Economic Research (NCAER) and the University of Maryland. IHDS-II is nationally representative and covers 42,152 households in 1420 villages and 1042 urban neighbourhoods across India. The survey includes household information on education, health, employment, economic status, social capital, fertility, etc. It has a panel component as well whereby 83% of the households interviewed as part of IHDS-II were interviewed for the first time in 2004-05 as part of IHDS-I.

With regards to the data requirements to address questions on intergenerational education mobility where information on education for parents, as well as their adult children, is needed, IHDS offers distinct advantages over the rounds of NSSO and NFHS. One, both NSSO and NFHS report information on levels of schooling completed and thus entails imputation of the number of years of schooling from such information resulting in discontinuities. On the other hand, IHDS reports data on the actual number of completed years of schooling as per its questionnaire. Two, more importantly, IHDS contains additional questions[2] which expand the scope of the sample to those beyond co-resident pairs of father-son and hence precludes biases due to consideration of co-resident pairs alone.

We start by preparing a dataset in alignment with the one by Azam and Bhatt (2015) which they created for IHDS-I. The dataset is unique in the sense that in addition to matching father-son data based on "Relationship to head of household" field in the household questionnaire which only ends up linking the co-resident pairs, we also use the additional question pertaining to the educational attainment of the head of the household. The co-resident pairs constitute pertinent information for only 34.58% of the male respondents in the age group of 20 to 65. In our final dataset, however, 96.73% of the males in the said age group are matched with their respective father’s schooling data giving rise to a more representative sample. For our analysis, we finally consider adult males in the age group 25-64. We drop those in the age group 20-24 as 27.01% of them are still enrolled in schools. Our final sample consists of 44, 532 observations with matching data for fathers’ education attainment. We exclude females in this analysis due to following reasons – One, households with women as head are very few (2.95% of all cases). Even for such households, education data is provided for their husbands. Hence, the unique feature (refer to footnote 22) of the IHDS data cannot be utilized for daughter-father, daughter-mother, or son-mother pairing to create a representative sample of such pairings. Two, given the ubiquitous family structure in India, adult females reside in either nuclear households or joint

[2] The question important to our cause is 1.18c on page 3 of the Income and Social Capital Questionnaire. It enquires the educational attainment (in years of schooling completed) of the father/husband of the head of the household.

families along with their respective husbands and kin belonging to the husband's side. Hence, the requisite pairing information is not available for a purported representative sample even if we just wish to consider just the co-residency condition. The downward bias due to such truncation is explained well through a simple model in Emran, Greene, and Shilpi (2017).

Next, using a simple OLS framework, we estimate several variants of the following base model –

$$S_i^c = \beta_0^c + \beta_1^c F_i^c + \gamma^c(\text{Social Class}) + \delta^c(\text{Religion}) + \sigma^c(\text{State}) + \epsilon_i^c$$

where, S_i^c and F_i^c have been defined earlier but appear with a superscript 'c' here that denotes the age cohort. We divide the sample into eight age cohorts – 25-29, 30-34, 35-39, . . ., 60-64 as of the year 2011. Dummies for social classes are assigned in accordance with caste divisions as per IHDS – Brahmin, Forward/General (excluding Brahmin), OBC, SC, ST, and Others. Similarly, religion dummies are assigned for Hindus, Muslims, Christians, Sikhs, Others (Buddhists, Jains, etc.). The above equation is first estimated without the controls at the all India level, and for different social classes/religions/age cohorts. The value of IGRC is then compared across these groupings. The controls are then introduced, and the patterns of IGRC are again analyzed. The age cohorts can also be interpreted as birth cohorts spanning 1947-1951 to 1982-1986. Thus, this analysis also facilitates understanding of the evolution of education mobility in India for adults born since our independence. In all the exercises, IGRC is preferred over IGC as we are more interested in understanding the trends and evolution of intergenerational educational mobility, albeit unconditional on the dispersion of educational outcomes for each generation across various cohorts and groupings. Moreover, since we have precluded the attenuation bias arising from the consideration of co-resident pairings alone, our approach can be considered robust.

3.2 Quantile Regression

Till now, we have enlisted the plan to survey the effect of father's education attainment on son's education outcomes using a simple OLS framework. However, we are also interested in gauging the difference in effects of fathers' education on sons with low educational achievements and those with relatively higher achievements. Essentially, we wish to study the differences in effects along the distribution of the sons' educational attainments. For this, we employ the quantile regression framework.

Alexander, Harding and Lamarche (2011) state, "Quantile regression offers a robust, and therefore an efficient alternative to least squares estimation" (p. 47). The quantiles describe the distribution of the dependent variable. A quantile regression models the relationship between the explanatory variables and the conditional quantiles of the dependent variable rather than just the conditional mean of the regressand. By this, a quantile regression gives a more comprehensive and a deeper insight into the effect of the independent variables on the dependent variable. The quantile coefficients show different effects and provide different interpretations along the distribution of the dependent variable. Thus, in this case, we estimate variants of the following specification –

$$Q_\theta(S_i/X) = \beta_0 + \beta_\theta F_i + \gamma(\text{Social Class}) + \delta(\text{Religion}) + \sigma(\text{State}) + \epsilon_i$$

where, $Q_\theta(S_i/X)$ represents θth quantile of the distribution of the number of years of schooling of the son conditional on the covariates X, in our case – father's years of schooling and other dummy variables. The quantile regression approach is an essential component of this study. If IGRCs across all quantiles are equal, then the two approaches are equivalent, and the relationship between son's educational attainment and father's educational attainment is linear. However, if the β_θs are different, then the mean estimator (β_1) ceases to be the best estimator, and quantile regressions convey important results.

3.3 The Education Transition Matrix

Another way to assess movements along the education distribution is through transition matrices. The use of transition matrices to view intergenerational mobility in education, occupation, and income is quite prevalent in the literature (Motiram & Singh, 2012; Hnatkovska, Lahiri & Paul, 2013; Azam & Bhatt, 2015; Iversen, Krishna & Sen, 2017). The education transition matrices map education attainment levels of sons with the educational attainment levels of fathers. The rows contain the education outcome levels of the fathers, and the columns bear the education attainment levels of the sons. The cells along the diagonal represent persistence at the same level of education, the cells in the upper triangle of the matrix represent upward mobility, and the cells in the lower triangle of the matrix depict downward mobility in educational attainment. In this exercise, we have resolved the number of years of schooling (which has a range of 0 to 16 as per the data) into seven education levels – level 0 (illiterates with zero years of schooling), level 1 (literate but below primary – one to three years of schooling), level 2 (primary – four to six years of schooling), level 3 (middle – seven or eight years of schooling), level 4 (secondary – nine and ten years of schooling), level 5 (higher secondary/diploma/certificate course – 11 to 14 years of schooling), and level 6 (15 or 16 years of schooling). The education transition matrices are prepared for the overall sample, for the urban and rural sample, and for social classes and religions.

3.4 Endogeneity and the IV Strategy

The models we have listed so far ensure that one source of endogeneity (w.r.t. the main explanatory variable of this study), i.e. simultaneity, is taken care of (as in most cases, father's time at school precedes son's schooling). However, endogeneity can also arise due to omitted variable bias and measurement error. As established earlier, we can also rule out measurement error as we deal with the education aspect and not income nor occupation. The potential endogeneity could be due to the correlation between father's education attainment with unobserved or unmeasurable antecedents of son's schooling such as genetics and cultural factors. Thus, to account for such endogeneities and ensure robustness of our obtained results, we make use of instrumental variable estimation and see if the findings still hold.

In choosing instruments for father's education, we follow Maitra and Sharma (2009). Their strategy was based on the correlation between public expenditure on education and schooling outcomes of parents. Greater public investment in education at a given place would translate into better educational facilities in that region. Hence, growing up at that place would have a positive impact on a given person's schooling but would be inconsequential for his son's education. Hence, an ideal instrument for father's schooling outcomes would be a variable that indicates the place where he grew up. Unfortunately, IHDS data does not bear such information. Hence, Maitra and Sharma (2009) use the birth year of parents and birth year of parents interacted with their original location (urban/rural) as instruments for parents' educational attainment. In this paper, we utilize different combinations of the birth year of the father and the interaction term between his year of birth and his original location as instruments.

3.5 The Great Gatsby Curve

Finally, we plot the relationship between education inequality (of fathers) and intergenerational persistence in education with the states/union territories of India being the unit of analysis. This exercise is to check if the plot follows "The Great Gatsby Curve". The Great Gatsby Curve (GGC) displays a positive relationship between economic inequality in one generation and intergenerational income immobility in the next for countries across the world. The phrase (was coined) and the curve was discussed for the first time by Krueger (2012) in a speech on the rise and consequences of inequality in the United States. The curve implies that the persistence in the circumstances handed over by parents to their children greatly depends on the economic inequality prevalent in the said region during parents' time. The implication of the curve was deftly put by Noah (2012) – "it's harder to climb a ladder when the rungs are farther apart". We attempt to see if that indeed is true in the case of education in India.

4. Results

To start with, we present the data summary of our sample. Table 1 and 2 contain the same for the overall sample and various segregations. As mentioned earlier, the overall sample contains information concerning father's educational attainment and other variables for 44, 532 males in the age group 25-64 (both numbers included). The rationale behind age bounds is borrowed from Behrman et al. (2001). The lower age floor ensures inclusion of those individuals who have completed their schooling. On the other extreme, the age ceiling helps in preventing selection bias due to different survival rates of individuals hailing from different family backgrounds.

Table 1

Summary Statistics for the overall sample and other groupings

Variable	Observations	Mean	Std. Dev.	Min	Max
		Overall			
yrssch	44,532	7.331	4.942	0	16
fatheryrssch	44,532	3.422	4.362	0	16
		Rural India			
yrssch	28,138	6.306	4.791	0	16
fatheryrssch	28,138	2.487	3.723	0	16
		Urban India			
yrssch	16,394	9.092	4.695	0	16
fatheryrssch	16,394	5.027	4.880	0	16
		Brahmins and Other Upper Castes			
yrssch	13,124	9.117	4.755	0	16
fatheryrssch	13,124	5.071	4.891	0	16
		Other Backward Castes (OBCs)			
yrssch	17,981	7.084	4.743	0	16

fatheryrssch	17,981	3.150	4.081	0	16
			SCs and STs		
yrssch	12,702	5.835	4.842	0	16
fatheryrssch	12,702	2.094	3.551	0	16
			Hindu		
yrssch	36,369	7.474	4.930	0	16
fatheryrssch	36,369	3.464	4.383	0	16
			Muslim		
yrssch	5,264	5.910	4.900	0	16
fatheryrssch	5,264	2.787	4.023	0	16
			Others (Christians, Sikhs, Jains, etc.)		
yrssch	2,899	8.124	4.709	0	16
fatheryrssch	2,899	4.046	4.553	0	16

Notes: yrssch – Years of schooling of the individuals; fatheryrssch – Years of schooling of an individual's father; Rural/Urban classification is as per 2011 census.

In the overall sample and all other groupings, sons unequivocally have a higher mean educational attainment than fathers. We observe that the level of educational attainment is higher in urban India as compared to rural India for both fathers and sons. As for caste groups, Brahmins and other upper castes are significantly more educated than their counterparts from lower castes. And, as far as religious groups are concerned, while Muslims are the least educated, the average educational outcomes for the rest of the population that includes Christians, Sikhs and Jains, are better than those of Hindus. Overall, it can be construed that in terms of absolute numbers, it has been a clear case of educational mobility for a given generation over the previous generation. The difference in years of schooling between two generations is more than three for all groupings.

Table 2

Summary Statistics by age cohorts

Son's Age Cohort	Sample Size	Percent	Average Years of Schooling	
			Son	Father
25-29	7,827	17.58	8.842	4.658
30-34	6,702	15.05	8.213	4.214
35-39	6,524	14.65	7.865	3.627
40-44	5,943	13.35	7.151	3.175
45-49	5,738	12.89	6.499	2.734

50-54	4,660	10.46	6.332	2.776
55-59	3,882	8.72	5.989	2.493
60-64	3,256	7.31	5.645	2.109
Total	44,532	100		

Notes: The ages are as of the year 2011. Thus, the age cohorts could also be understood as the following respective birth cohorts – 1982-1986, 1977-1981, 1972-1976, 1967-1971, 1962-1966, 1957-1961, 1952-1956, and 1947-1951.

In table 2, we report the sample means of education attainment by age/birth cohorts. As can be observed from the table, all the cohorts are well represented in terms of their respective sample sizes. This data too showcases that there has been a clear and steady growth in educational attainment over the years and across generations. Sons have consistently exceeded their fathers w.r.t. the time they have spent in school since India's independence in 1947.

Next, we look at the rate increases in schooling attainments per annum. To obtain the same, we regress the number of years of schooling on the birth year of individuals using simple OLS framework. The coefficient estimates listed in table 3 mark the respective rates of increases of schooling for the overall sample and various categorizations.

Table 3

Average growth rate in educational attainment (Dependent Variable – 'year of birth')

	Growth per Year	N
Overall	0.0910***	44532
	(0.00198)	
Rural	0.113***	28138
	(0.00234)	
Urban	0.0529***	16394
	(0.00321)	
Brahmins & Other UCs	0.0666***	13124
	(0.00359)	
OBCs	0.0953***	17981
	(0.00295)	
SCs and STs	0.126***	12702
	(0.00352)	
Hindu	0.0928***	36369
	(0.00218)	
Muslim	0.0933***	5264
	(0.00557)	

Others	0.0920***	2899
	(0.00725)	

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

From regression 1 of table 3, the yearly increase in educational attainment of individuals in the sample is seen to be 0.09 years. This translates to an increase of approximately 0.9 years of schooling per decade. Convergence in educational attainment can also be observed between rural and urban India as the growth in schooling outcomes in rural areas is almost twice as seen in the urban regions. Similar convergence can also be discerned with respect to the rates between upper castes and the rest. However, the respective increases in schooling outcomes have moved at a similar rate for all three religion groupings, which in turn is almost the same as the overall average growth rate. In figure 1, we fit a LOWESS curve to represent the underlying trend in schooling improvements since independence pictorially. A LOWESS (locally weighted scatterplot smoothing) plot is a smooth curve resulting from non-parametric regression methods.

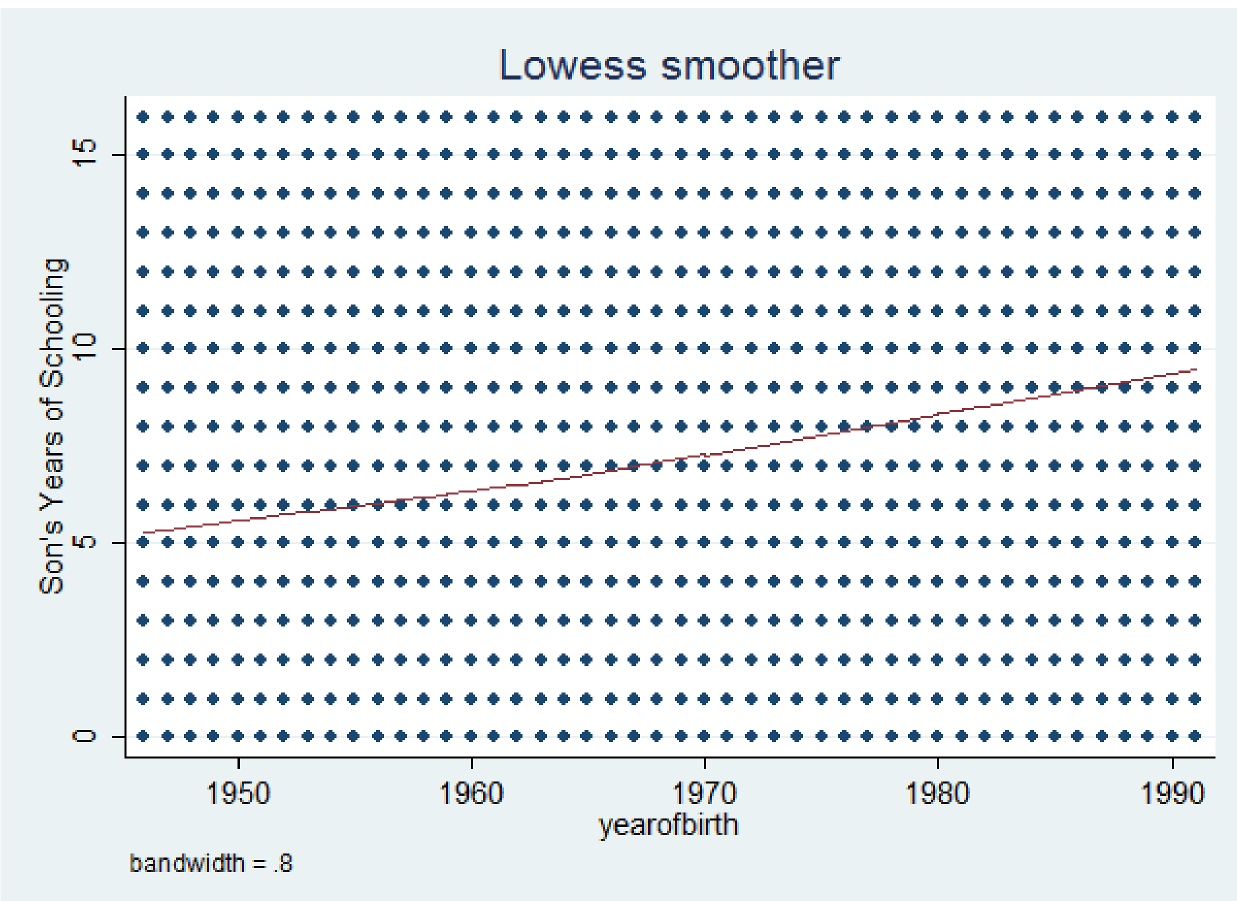

Figure 1. *Lowess plot of sons' educational attainment by year of birth*

Having sifted through the broad figures, a couple of pivotal questions now arise – through all the years since independence, has there been a level playing field? If not, are we at least moving in the right direction? From here on, we empirically attempt to determine the influence circumstances (proxied by father's education in this paper) have on the socioeconomic outcomes (proxied by educational attainment) of the individual and see if it has lessened over time, and further understand its nature across various categories of individuals.

4.1 Preliminary Empirical Results

In table 4, we lay out the OLS estimation results for the overall sample. We start with the base model at first and then add controls for caste, state, and religion, sequentially. For the specification with no controls, the estimated intergenerational education coefficient comes to be 0.588. As it is also statistically significant, this underscores a sizable high degree dependency of an individual's life chances on his father's status. Next, we apply controls to account for factors that could have a bearing on schooling achievements of individuals. There is evidence available in literature as to how caste plays a role in school participation of individuals (E.g. in Hickey & Stratton, 2007), how there exists a link between religion and education (E.g. in Booroah & Iyer, 2005), and how educational opportunity differs across states in India (Asadullah & Yalonetzky, 2012). Hence, we control for the three factors by employing respective dummy variables. The last row evidently shows that the fit of the model improves with the addition of each control. With the addition of controls, the degree of persistence decreases, in turn underlining the importance of caste, state, and religion in inequality of opportunity debate. Caste seems to have a higher bearing on IGRC as compared to the other two factors. The statistical significance of the results remains robust to the addition of controls.

Table 4

Intergenerational Regression Coefficients (All India) (Dependent Variable – 'yrssch')

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	yrssch	yrssch	yrssch	yrssch	yrssch	yrssch	yrssch
fatheryrssch	0.588***	0.544***	0.534***	0.582***	0.529***	0.570***	0.522***
	(0.00405)	(0.00426)	(0.00430)	(0.00406)	(0.00428)	(0.00413)	(0.00433)
cons	5.318***	7.131***	7.325***	5.456***	7.233***	6.684***	8.140***
	(0.0271)	(0.0827)	(0.162)	(0.0290)	(0.0824)	(0.145)	(0.165)
Caste Controls	No	Yes	Yes	No	Yes	No	Yes

State Controls	No	No	Yes	No	No	Yes	Yes
Religion Controls	No	No	No	Yes	Yes	Yes	Yes
N	44532	44411	44411	44532	44411	44532	44411
adj. R-sq.	0.270	0.287	0.306	0.276	0.299	0.296	0.316

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

Overall, the above results agree with the mobility estimates obtained in Azam, and Bhatt[3] (2015) and Emran and Shilpi[4] (2015) but are a departure from the estimates in Jalan and Murgai[5] (2007) and Maitra and Sharma[6] (2009).

The lowess curve in figure 2 supports these findings. First, the positive relation between son's and father's educational attainment is unmistakable. Second, as the slope of the curve is less than 45 degrees, on an average, an additional year of education for father translates to less than a year of schooling for his son. However, till the curve is above the 45-degree line, on an average, the educational attainment of sons is greater than that of fathers. The intersection of the curve with the 45-degree line happens at about 12.5 years of schooling. This is comparable to the figure obtained by Maitra and Sharma (2009) for a similar exercise. However, it must also be mentioned that the authors were dealing with a much smaller sample (only co-resident pairs) drawn from IHDS-I.

[3] Azam and Bhatt (2015) obtained an IGRC of 0.634 from a similar exercise performed on IHDS-I data.

[4] Emran and Shilpi (2015) estimated sibling correlation (SC) – another measure of intergenerational persistence – to be ranging from 0.614 to 0.624 for brothers over two rounds of NFHS data (1993 and 2006). They also calculate IGRC for the males which comes out to be 0.541 and 0.523 respectively for the two rounds.

[5] Jalan and Murgai (2007) worked with the 1998-99 round of NFHS and obtained IGRC estimates ranging from 0.236 for the 1969-1973 birth cohort of males to 0.153 for the males in the 1979-1983 birth cohort.

[6] The IGRC estimates for urban males and rural males in Maitra and Sharma (2009) stand at 0.3332 and 0.3831 respectively. Their study is based on IHDS-I data.

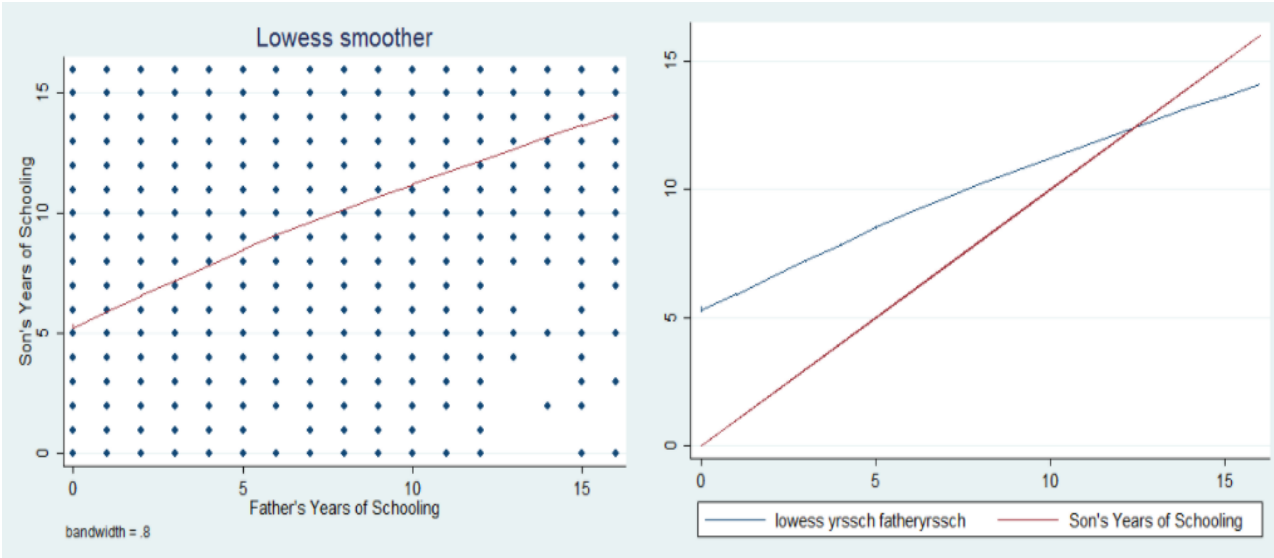

Figure 2. *Lowess plot of sons' educational attainment by fathers' educational attainment*

Note: Both plots are same. The only difference is that in the right plot, a reference 45-degree line has been shown.

4.1.1 Sub-Sample Analysis – Intergenerational Mobility in Co-resident Households

Co-resident household father-son pairs make up for only 26.63% of the potential pairings which could have been possible if information on father's education was expressly available for each male individual in the age-group of 25 – 64 from the survey. On the other hand, by making use of the unique feature of IHDS-II (question 1.18c on page 3 of the Income and Social Capital Questionnaire), we account for 96.73% of the potential pairings. In this section, by estimating IGRCs for the co-resident father-son pairs, we highlight the attenuation bias that creeps in due to the neglect of non-co-resident pairs. Table 5 lists the results.

Table 5

Intergenerational Regression Coefficients (All India) for co-resident pairs (Dependent Variable – 'yrssch')

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	yrssch	yrssch	yrssch	yrssch	yrssch	yrssch	yrssch
fatheryrssch	0.487***	0.464***	0.456***	0.482***	0.453***	0.472***	0.446***
	(0.0069)	(0.0072)	(0.0072)	(0.0069)	(0.0072)	(0.0070)	(0.0073)
cons	6.711***	7.728***	8.237***	6.849***	7.822***	8.113***	9.005***

	(0.06)	(0.15)	(0.26)	(0.06)	(0.15)	(0.24)	(0.27)
Truncation Bias[7]	20.75%	17.24%	17.11%	20.75%	16.78%	20.76%	17.04%
Caste Controls	No	Yes	Yes	No	Yes	No	Yes
State Controls	No	No	Yes	No	No	Yes	Yes
Religion Controls	No	No	No	Yes	Yes	Yes	Yes
N	12227	12173	12173	12227	12173	12227	12173
adj. R-sq.	0.271	0.28	0.299	0.276	0.286	0.297	0.306

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

It can be observed in table 5 that the number of observations goes down massively and that there is a downward truncation bias ranging from about 13% to 21% in the coefficients estimated from the co-resident pairs. The figures are in line with the estimates obtained from a similar comparative analysis in Azam and Bhatt (2015). This underscores the point made earlier about there being a severe issue with sample selection if 'co-resident only' father-son pairs are chosen.

4.2 Intergenerational Education Mobility Across Cohorts

In this section, we look at how educational mobility has evolved since independence. For this purpose, we perform a cohort trend analysis by estimating IGRCs in respective OLS regression models for each of the five-year age/birth cohorts. While table 6 presents the estimates for the base model, table 7 contains the estimates after application of controls.

Table 6

Age cohort trend in Intergenerational Regression Coefficients (All India) (Dependent Variable – 'yrssch')

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
fatheryrssch	0.481***	0.512***	0.507***	0.583***	0.600***	0.673***	0.701***	0.750***
	(0.0089)	(0.0096)	(0.0101)	(0.0115)	(0.0128)	(0.0137)	(0.0152)	(0.0188)
cons	6.600***	6.055***	6.028***	5.301***	4.858***	4.463***	4.240***	4.065***
	(0.0685)	(0.0723)	(0.0713)	(0.0734)	(0.0728)	(0.0785)	(0.0828)	(0.0885)
N	7827	6702	6524	5943	5738	4660	3882	3256
adj. R-sq.	0.248	0.259	0.229	0.252	0.228	0.285	0.288	0.275

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

[7] $Truncation\ Bias = \left(\frac{IGRC_{Table\ 3.4} - IGRC_{Table\ 3.5}}{IGRC_{Table\ 3.5}}\right) * 100$

Table 7

Age cohort trend in Intergenerational Regression Coefficients (All India) (in presence of controls) (Dependent Variable – 'yrssch')

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
fatheryrssch	0.430***	0.449***	0.444***	0.503***	0.498***	0.565***	0.609***	0.637***
	(0.0098)	(0.0103)	(0.0109)	(0.0128)	(0.0140)	(0.0158)	(0.0169)	(0.0209)
cons	9.528***	9.357***	8.456***	7.210***	8.575***	7.333***	7.216***	7.315***
	(0.361)	(0.447)	(0.451)	(0.501)	(0.516)	(0.475)	(0.469)	(0.650)
Caste Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Religion Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7799	6681	6505	5933	5726	4651	3872	3244
adj. R-sq.	0.300	0.316	0.285	0.304	0.302	0.356	0.356	0.330

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

From tables 6 and 7, it can be observed that there is a clear reduction in intergenerational persistence in educational attainments as the years have progressed, although the decrease in non-monotonic (For example, in table 7, IGRC at age cohort of 40-44 is marginally higher than IGRC at 45-49). Father's education has a consistently positive and statistically significant effect (at a significance level of one percent) on the son's education across all cohorts and even after accounting for caste, state, and religion differences. A decrease in IGRC from 0.750 (highly persistent relationship) for the oldest birth cohort (1947-1951) to 0.481 (moderately persistent) for the youngest birth cohorts (1982-1986) augurs well in our society's path towards greater mobility, and in turn towards facilitating an economic environment of greater equality of opportunity.

Considering the role caste and religion plays in determining socio-economic outcomes and status in India, we determine the IGRC estimates for Brahmins and other Upper Castes, Other Backward Castes (OBCs), Scheduled Castes and Scheduled Tribes (SCs and STs), Hindus, and Muslims by age cohorts (the youngest 10-year age cohort – 25 to 34, and the oldest 10-year age cohort – 55 to 64) to understand its evolution in the sub-samples and differences between them. These estimates are displayed in tables 8 and 9.

Table 8

Cohort trends in Intergenerational Regression Coefficient by Caste (Dependent Variable – 'yrssch')

	(1)	(2)	(3)	(4)	(5)	(6)
	Brahmins and Other UCs		OBCs		SCs and STs	
	25-34	55-64	25-34	55-64	25-34	55-64
fatheryrssch	0.422***	0.596***	0.449***	0.601***	0.447***	0.670***
	(0.0126)	(0.0188)	(0.0119)	(0.0221)	(0.0149)	(0.0376)
cons	9.190***	6.870***	8.511***	4.372***	7.549***	5.134***
	(0.357)	(0.450)	(0.612)	(0.691)	(0.431)	(1.146)
State Controls	Yes	Yes	Yes	Yes	Yes	Yes
Religion Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	4042	2217	5919	2931	4313	1882
adj. R-sq.	0.337	0.359	0.271	0.233	0.228	0.237

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

Table 9

Cohort trends in Intergenerational Regression Coefficient by Religion (Dependent Variable – 'yrssch')

	(7)	(8)	(9)	(10)
	Hindu		Muslim	
	25-34	55-64	25-34	55-64
fatheryrssch	0.426***	0.599***	0.533***	0.650***
	(0.00813)	(0.0146)	(0.0237)	(0.0452)
cons	8.595***	7.592***	7.159***	3.338***
	(0.326)	(0.492)	(0.404)	(0.498)
Caste Controls	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
N	11700	5858	1899	771
adj. R-sq.	0.292	0.351	0.335	0.277

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

Within all categorizations in tables 8 and 9, there is a conspicuous improvement in education mobility across back-to-back generations since independence. The pace of this progress is different for different groups, though. Thanks to affirmative action policies (in education, public sector jobs, and state legislatures) by the government, especially in favour of SCs and STs, the improvement in their education mobility has happened at a faster rate as compared to the upper castes and the OBCs. IGRC for SCs and STs has fallen by 33.28% over 30 years as compared to 25.29% for OBCs and 29.19 percent for the upper castes. Although this result is not generalizable as we are dealing with a limited sample and do not have other robustness checks in place, our arguments still fit the general narrative (Jalan & Murgai, 2007; Hnatkovska, Lahiri & Paul, 2013). In fact, in addition to the reason mentioned earlier, Hnatkovska, Lahiri and Paul (2013) attribute breaking down of the caste-based shackles towards mobility on structural changes in the Indian economy in the last two decades as well.

Apropos grouping by religion, Hindus have always held precedence over Muslims in case of educational mobility. Moreover, the percentage decrease in intergenerational persistence across cohorts separated by 30 years for Hindus (at 28.88%) is more than 1.5 times than for Muslims (18%). The IGRC for Muslims in the age cohort of 25-34 is also much higher than the IGRC for the overall sample in that age cohort. The status of Muslims continues to be majorly encumbered by their previous generations.

4.3 Intergenerational Education Mobility Across Education Distribution

Where in the educational attainment distribution of India's population are the individuals most mobile with respect to their predecessors? Do those sons whose parents are illiterate or sparsely literate manage to better their life chances by spending a greater number of years at school? To what extent are they helped by factors external to their circumstances? How have sons in the rest of the distribution fared w.r.t. their mobility rates? We cast our attention to these questions in this section. We rely on quantile regression framework to discern between IGRCs along various points of distribution of sons' schooling outcomes. The estimates are presented in tables 10 through 14.

Table 10

Intergenerational Regression Coefficients along the distribution of 'years of schooling' (Overall Sample)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	Quantile (.10)	Quantile (.20)	Quantile (Median)	Quantile (.60)	Quantile (.75)	Quantile (.90)	Quantile (.95)
fatheryrs sch	0.667***	0.900***	0.667***	0.500***	0.438***	0.333***	0.200***
	(0.0166)	(0.0176)	(0.00619)	(0.00718)	(0.00985)	(0.00844)	(0.0230)
cons	0	0	5***	7***	9***	12***	14***

	(0.10)	(0.11)	(0.03)	(0.04)	(0.06)	(0.08)	(0.13)
N	44532	44532	44532	44532	44532	44532	44532

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; In Breusch-Pagan / Cook-Weisberg test for heteroscedasticity – chi2(1) = 2394.11 and Prob > chi2 = 0.0000. Hence, we reject the Null. Use of Quantile Regression is justified.

Table 11

Intergenerational Regression Coefficients along the distribution of 'years of schooling' (Rural Sample)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	Quantile (.10)	Quantile (.20)	Quantile (Median)	Quantile (.60)	Quantile (.75)	Quantile (.90)	Quantile (.95)
fatheryrs sch	0.600***	0.833***	0.583***	0.500***	0.455***	0.500***	0.400***
	(0.0590)	(0.0381)	(0.00747)	(0.00807)	(0.00486)	(0.0101)	(0.0127)
cons	0	0	5***	6.500***	8.182***	10***	12***
	(0.337)	(0.217)	(0.0537)	(0.0613)	(0.0622)	(0.0877)	(0.0738)
N	28138	28138	28138	28138	28138	28138	28138

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; In Breusch-Pagan / Cook-Weisberg test for heteroscedasticity – chi2(1) = 529.25 and Prob > chi2 = 0.0000. Hence, we reject the Null. Use of Quantile Regression is justified.

Table 12

Intergenerational Regression Coefficients along the distribution of 'years of schooling' (Urban Sample)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	Quantile (.10)	Quantile (.20)	Quantile (Median)	Quantile (.60)	Quantile (.75)	Quantile (.90)	Quantile (.95)
fatheryrs sch	0.750***	0.857***	0.500***	0.467***	0.400***	0.250***	0.0909***
	(0.0322)	(0.0191)	(0.00662)	(0.00577)	(0.00398)	(0.0338)	(0.00635)
cons	0	0.714***	7***	8***	10***	13***	15***
	(0.217)	(0.182)	(0.0518)	(0.0493)	(0.0269)	(0.274)	(0.0429)
N	16394	16394	16394	16394	16394	16394	16394

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; In Breusch-Pagan / Cook-Weisberg test for heteroscedasticity – chi2(1) = 915.14 and Prob > chi2 = 0.0000. Hence, we reject the Null. Use of Quantile Regression is justified.

Table 13

Intergenerational Regression Coefficients along the distribution of 'years of schooling' (Age Cohort: 25-34)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Quantile (.10)	Quantile (.20)	Quantile (Median)	Quantile (.60)	Quantile (.75)	Quantile (.90)	Quantile (.95)
fatheryrs sch	0.714***	0.800***	0.455***	0.438***	0.467***	0.333***	0.111***
	(0.0287)	(0.0207)	(0.0112)	(0.00646)	(0.00974)	(0.0144)	(0.00658)
cons	0	1***	7***	8***	9***	12***	14.78***
	(0.187)	(0.183)	(0.0720)	(0.0235)	(0.0842)	(0.0935)	(0.0855)
N	14529	14529	14529	14529	14529	14529	14529

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; In Breusch-Pagan / Cook-Weisberg test for heteroscedasticity – chi2(1) = 563.28 and Prob > chi2 = 0.0000. Hence, we reject the Null. Use of Quantile Regression is justified.

Table 14

Intergenerational Regression Coefficients along the distribution of 'years of schooling' (Age Cohort: 55-64)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Quantile (.10)	Quantile (.20)	Quantile (Median)	Quantile (.60)	Quantile (.75)	Quantile (.90)	Quantile (.95)
fatheryrs sch	0.667***	0.938***	0.733***	0.700***	0.533***	0.600***	0.500***
	(0.105)	(0.0728)	(0.0227)	(0.0175)	(0.0239)	(0.0142)	(0.0240)
cons	0	0	4***	5***	8***	10***	12***
	(0.597)	(0.413)	(0.148)	(0.0802)	(0.0879)	(0.0249)	(0.135)
N	7138	7138	7138	7138	7138	7138	7138

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; In Breusch-Pagan / Cook-Weisberg test for heteroscedasticity – chi2(1) = 76.29 and Prob > chi2 = 0.0000. Hence, we reject the Null. Use of Quantile Regression is justified.

It is clear from tables 10 to 14 that the effect of fathers' education on sons' education is not linear along percentiles of sons' schooling attainment distribution as IGRCs estimated at different quantiles of the distribution are not equal. With reference to regressions in each table, we observe a similar general trend – intergenerational mobility in education (1 – IGRC) decreases till about the 20th percentile of sons' education distribution and then increases along the rest of the distribution; although in some case, the increase in non-monotonic. For the overall sample, the mobility starts at a value of 0.334 at the 10th percentile of the distribution, decreases and reaches a value of 0.1 at the 20th percentile, and then maintains an upward monotonic trend along the rest of the distribution to reach an (almost) peak value of 0.8 at the 95th percentile. This means that the individuals at the highest point of educational attainment are the ones who are least affected by their circumstances. Even for the rest of the sub-samples (rural, urban, youngest

10-year age cohort, oldest 10-year age cohort), this holds true, albeit to different extents. Although it is expected, by intuition, that urban regions of a country promote greater mobility as compared to the rural areas, from table 11 and 12, it is not apparent at all points of educational distribution. At lower levels of schooling (till about 20th percentile of the respective distributions), rural areas have facilitated greater educational mobility. Conversely, at the other end of the spectrum, educational mobility is much more prominent in urban areas as compared to the rural regions. From tables 13 and 14, we can safely contend that there has been a marked improvement in educational mobility over time at almost all points of the distribution. Next, we graphically present quantile regression coefficients of fathers' years of schooling (our measure of intergenerational mobility/persistence) through figures A.1 to A.5 in Appendix A. The quantiles of son's years of schooling are on the x-axis and the coefficient magnitudes are on the y-axis.

Each regression, for the overall sample and all other sub-samples, yields a statistically significant relationship between father's schooling and son's schooling. However, we still check for the robustness of these findings (for the overall sample only) by implementing controls and see if the results still hold. Table 15 contains the estimates.

Table 15

Intergenerational Regression Coefficients along the distribution of 'years of schooling' in presence of controls (Overall Sample)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	Quantile (.10)	Quantile (.20)	Quantile (.35)	Quantile (Median)	Quantile (.60)	Quantile (.75)	Quantile (.90)	Quantile (.95)
fatheryrssch	0.667***	0.800***	0.667***	0.500***	0.471***	0.426***	0.367***	0.200***
	(0.0135)	(0.0127)	(0.0058)	(0.00465)	(0.0037)	(0.0051)	(0.0046)	(0.0186)
cons	1.333***	2.000***	5.667***	9.000***	9.765***	11.13***	13.83***	15.20***
	(0.181)	(0.538)	(0.228)	(0.216)	(0.134)	(0.125)	(0.185)	(0.374)
Caste Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Religion Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	44411	44411	44411	44411	44411	44411	44411	44411

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; In Breusch-Pagan / Cook-Weisberg test for heteroscedasticity – chi2(1) = 1771.30 and Prob > chi2 = 0.0000. Hence, we reject the Null. Use of Quantile Regression is justified.

The results in Table 15 are robust to the addition of controls. The statistical significance of the associations remains intact without a major change in the coefficient magnitudes. Having said that, findings from our study are in dissonance with those obtained through similar exercises performed in Tassinari (2017) and Gaentzsch and Roman (2017) in the contexts of Italy and Latin American Countries (Chile and Peru) respectively. While our findings suggest intergenerational persistence to be at its lowest at the upper end of the educational distribution, Gaentzsch and Roman (2017) find evidence of high persistence at both upper and lower ends of the distribution and attributes the same to the ceiling and floor effects. In terms of income mobility, Tassinari (2017) finds proof of the highest amount of dependence of a given generation on the previous generation at the top end of the distribution. He finds a greater deal of mobility at the middle and lower ends of the distribution. Tassinari (2017) attributes this stickiness at the top to the parental tendency of rich parents to pass on their respective advantages and social network effects. The Indian case is evidently different. At the top end of the education distribution, an Indian son has managed to break away from his circumstances (to a great extent), while at the lower and middle parts of the distribution, a son is still encumbered by his background, and even his own effort coupled with external factors such as government policies (in education and elsewhere) has done little to diminish his dependency on his father. The situation is improving, though. Sons born in the 1980s are doing better than those born in 1950s in breaking away from their respective circumstances.

4.4 The Education Transition Matrices

Education transition matrices provide non-parametric means of studying intergenerational mobility. The interpretation of each cell of an education transition matrix is this - Given father's education level of F_L in the L^{th} row out of n rows representing n education levels, each cell in that row represents the probability of his son reaching one of the n education levels marked in n columns. So, a cell with an address (L, L) marks the probability $P(S_L|F_L)$ of a son attaining L^{th} level of education given his father's education attainment of L. Hence, the elements in a row always add up to 100. As mentioned earlier, the dimension of education transition matrices in this study is 7×7 amounting to seven levels of education. We list the transition matrix for the overall sample in table 16 and abridged summaries of respective transitions matrices for other groupings in tables 17 through 20.

Table 16

Education Transition Matrix for the Overall Sample

Levels		Son's Education						
		0	1	2	3	4	5	6
Father's Education	0	33.65	6.35	17.11	14.5	17.52	6.86	3.99
	1	10.3	10.18	18.66	15.85	26.72	10.57	7.73
	2	7.49	3.41	19.71	17.14	28.39	13.48	10.38
	3	3.62	1.84	8.19	18.94	30.45	19.18	17.79
	4	2.35	0.8	4.18	8.12	34.5	23.07	26.99
	5	1.2	0.53	2.59	4.99	20.74	30.12	39.83
	6	0.62	0.21	0.96	2.94	11.83	17.72	65.73

Table 17

Comparison of Educational Transitions between Urban Population and Rural Population

Urban India				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	23.14	37.17	6.71
	Primary		59	14.16
	Secondary			32.52
	Higher Secondary			48.2
	At least Graduation			70.62
Rural India				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	37.39	25.26	3.03
	Primary		47.23	7.57
	Secondary			19.85
	Higher Secondary			26.75
	At least Graduation			51.09

Table 18

Comparison of Educational Transitions across various Caste Groups

Brahmins + Other Upper Castes				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's	Illiterate	25.17	37.61	5.86

Education	Primary		61.05	14.49
	Secondary			33.65
	Higher Secondary			49.05
	At least Graduation			70.27

Other Backward Castes				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	32.31	27.85	3.72
	Primary		50.23	9.04
	Secondary			22
	Higher Secondary			30.8
	At least Graduation			59.52

Schedules Castes and Scheduled Tribes				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	40.05	23.57	3.19
	Primary		44.11	7.25
	Secondary			18.91
	Higher Secondary			28.36
	At least Graduation			57.31

Table 19

Comparison of Educational Transitions across various Religions

Hindus				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	32.49	29.55	4.16
	Primary		53.6	11.2
	Secondary			27.64
	Higher Secondary			39.56
	At least Graduation			67.15

Muslims				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	42.94	19.07	2.76
	Primary		39.53	6.29
	Secondary			23.74
	Higher Secondary			40.46

	At least Graduation			54.63
Christians, Sikhs, and Others				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	28.19	33.7	4.51
	Primary		55.02	6.95
	Secondary			24.55
	Higher Secondary			42.42
	At least Graduation			61.02

Table 20

Comparison of Educational Transitions between the Youngest and the Oldest 10-year Age Cohort

Age Cohort 25-34				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	23.56	35.27	5.1
	Primary		55.35	9.73
	Secondary			28.27
	Higher Secondary			41.37
	At least Graduation			62.78
Age Cohort 55-64				
Educational Levels		Son's Education		
		Illiterate	At least Secondary	At least Graduation
Father's Education	Illiterate	43.41	19.87	2.75
	Primary		47.91	10.91
	Secondary			29.76
	Higher Secondary			35
	At least Graduation			73.28

In table 16, the diagonal elements indicate intergenerational persistence. The off-diagonal elements lying to the right of the diagonal reflect upward mobility and the rest of the elements reflect downward mobility. In the matrix (Table 16), the upper-triangle non-diagonal elements dominate the lower-triangle non-diagonal cells. Hence, right at the outset, we can attest to higher upward mobility than downward mobility in education for the overall sample. In most of the unabridged matrices[8], the top-left and the bottom-right numbers are substantially higher than the ones in the rest of the cells. This indicates high persistence at the topmost and the bottommost levels of education. A son of an illiterate will 'usually' remain illiterate and a son of someone with the highest level of education will mostly go on to complete his education. Nonetheless, this persistence is decreasing across cohorts as can be observed in table 20. 43.41 percent of the sons born in the 1950s to illiterate fathers would remain illiterate. However, this figure reduces to almost half its value at 23.56% for the sons born in the eighties. Although this is a welcome improvement, the same can't be said about the other end of the spectrum. Out of all sons whose fathers completed the highest level of education, 73.28% of them went on to attain the highest level of schooling as well for the oldest age cohort in our study. That number reduced to 62.78%. This is certainly a digression. Nonetheless, it must be stressed that upward mobility has improved for sons of fathers with low educational achievements (illiterate, below primary, and primary) in case of younger cohorts.

The matrices divided along regional lines (urban vs rural) (Table 17) point to higher educational mobility for the urban population as compared to the rural folk. In the urban sample, 37.17% of the sons with illiterate fathers, 59% of those whose fathers just completed primary schooling went on to complete secondary level of education. The same numbers languish at 25.26% and 47.23% respectively in case of the rural population. Even among sons with relatively better-educated fathers (e.g. fathers who completed secondary education or higher secondary education), the upward mobility is more pronounced in urban areas than rural areas.

The disadvantage of being born as a scheduled caste or in a scheduled tribe shows vividly among the education transition matrices categorized by castes (Table 18). For SCs and STs, the persistence is highest at the lowest levels of education (illiterates) and lowest at the highest level of schooling (graduates and above) when compared to the upper castes and the OBCs. Upward mobility conditional on lower levels of fathers' education (below primary and primary) is also least for SCs and STs. Even the OBCs only do marginally better. This trend of "Upper castes better off than OBCs better off than SCs and STs" continues even in cases where sons of fathers with secondary and higher secondary education go on to complete the highest levels of education.

[8] Unabridged Transition Matrices for all sub-sample/groupings shall be provided on request.

Coming to categorizations done by religions (Table 19), the education transition matrices show a clear position of disadvantage for Muslim offspring, whereas the advantage is divided between respective sons of Hindus and other religions (Christians, Sikhs, Jains, etc.). To cite an instance, educational persistence is highest at the lowest level of education and lowest at the highest level of education for Muslims when compared with all others. Between Hindus and other religions, while persistence is highest at the highest level of education for Hindus (67.15% vs 61.02% for other religions), it is the lowest at the lowest level of education for Christians, Sikhs, Jains, etc. (28.19% vs 32.49% for Hindus). The propensity for sons of fathers at low educational levels (below primary and primary) to reach the highest education levels (secondary and above) is the least for Muslims. The same numbers are much better for Hindus and other religions with Hindus at a slightly worse off position than members of other religions.

Our quantile regressions had hinted that sons at the top end of the education distribution were relatively unencumbered by their circumstances. This needs to be interpreted in the following way – a low IGRC, say 0.2, obtained for the 90th percentile of the distribution means that if one were to pick an individual with high educational qualifications, his achievements are explained by only 20% of his circumstances (father's education). The rest of the explanation is provided by his own industry and other unobserved factors. The results analyzed from transition matrices in this section complement our earlier findings. Here, the level of educational mobility across cohorts and other groupings is analyzed along the educational attainment distribution of the fathers. We find that the persistence at the top end of the education distribution is relatively stronger than in the rest of the distribution. Even at other levels of education, persistence is sizably strong; although, it has lessened over the years. Most of the mobility is driven by sons of fathers at the lower end of the education distribution attaining higher education levels and by sons of fathers at a relatively higher level reaching even higher. Such an advantage in mobility is best enjoyed by a Brahmin/high caste son living in a city and practicing any religion other than Islam.

4.5 Robustness Check – The IV Strategy

Studies assessing the impact of circumstances on one's life chances by linking parent's education outcomes to his/her schooling achievements are beset with endogeneity problems. The issues of simultaneity and measurement errors of variables can be assumed to be insignificant owing to model set-up (father's schooling precedes son's schooling) and selection of variable (measurement of education attainment defined by number of years of schooling is straightforward, and moreover, in case of IHDS, there is no imputation involved). However, omitted variable bias can eminently creep in as unobservable factors such as genetics and other cultural factors (which are correlated with both father's and son's education) are not accounted for. Hence, to get around this issue, we make use of instrumental variable estimation and see if the results still hold.

Prior studies have used local information (from the place of birth of an individual) as a proxy for his/her current characteristics (like health status, education achievements, etc.) for first stage regressions in IV estimations (Schultz, 2002). For example, a person's early education might be influenced by the facilities, infrastructure, and circumstances existent, with regards to education, at the place of his/her growing up. Accordingly, we follow Maitra and Sharma (2009) and use a combination of the following variables – birth year of the father and birth year of the father interacted with his original location (rural/urban). In doing so, it is assumed that father's education is influenced by the public investments made in education during his years and place of schooling, which in turn is captured through a combination of his year of birth and whether he hails from a rural background or an urban background. The latter is considered as the place of father's birth is not provided in IHDS data. These instruments have no way of affecting the son's education directly other than through an indirect effect via his father's education. The validity of the instruments is checked by employing the standard Sargan test. The results are presented in table 21.

Table 21

Instrumental Variable Estimation of Intergenerational Regression Coefficients

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	OLS	IVREG	IVREG	IVREG	IVREG	IVREG
fatheryrssch	0.522***	0.641***	0.762***	0.720***	0.693***	0.711***
	(0.00433)	(0.0532)	(0.0377)	(0.0309)	(0.0323)	(0.0292)
Caste Controls	Yes	Yes	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes	Yes	No
Religion Controls	Yes	Yes	Yes	Yes	No	Yes

Instrumented			fatheryrssch			
Instruments		fatheryob	father_yob_ol	Both	Both	Both
N	44411	9563	9563	9563	9563	9563
adj. R-sq.	0.316	0.249	0.192	0.215	0.224	0.198
Sargan Statistic[a]				0.0629	0.1006	0.1261

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001; fatheryob – Father's year of birth; father_yob_ol – Interaction between father's year of birth and his original location (rural/urban dummy variable); a – Chi-sq. (1) P-Value for over-identification test of all instruments.

Table 22

Instrumental Variable Estimation of Intergenerational Education Coefficients – First-Stage Regression Results

Variable	R-sq.	Adjusted R-sq.	Test 1[a]	Test 2[b]	F (2,9519)	Prob > F
fatheryrssch (4)	0.2122	0.2087	850.941	464.879	467.625	0
fatheryrssch (5)	0.195	0.1917	772.685	418.545	416.192	0
fatheryrssch (6)	0.1863	0.1853	941.009	521.201	529.075	0

Notes: a – Under-identification Test (Anderson Canonical Correlation LM Statistic); b – Weak identification Test (Cragg-Donald Wald F Statistic)

The OLS estimates in regression one are provided for comparison. In IV regressions 2 and 3, the instruments are found to be weak. However, in IV regressions 4, 5 and 6, the effect of fathers' schooling on sons' schooling comes out to be positive and statistically significant. This result comes through even as a major number of observations were lost as pertinent information needed to build instruments was available for limited sample only. Thus, our results can be claimed to be robust.

4.6 The Great Gatsby Curve

The Great Gatsby Curve (GGC) shows a positive association between economic inequality (for the previous generation) and the ensuing intergenerational economic persistence across countries. We plot the relationship between education inequality among the fathers (quantified by education Gini) and intergenerational education coefficient across the Indian states and check if their association resembles GGC. The Gini Coefficients of education attainment by fathers and IGRCs for respective states are presented in table 23, and the resulting plot is shown in figure 3.

Table 23

State-wise Gini Coefficients of Education and IGRCs

State	Education Gini	IGRC
Orissa	0.336	0.685
Daman & Diu	0.324	0.714
Arunachal Pradesh	0.322	0.462
Sikkim	0.315	0.44
Gujarat	0.308	0.572
Meghalaya	0.306	0.618
Chhattisgarh	0.305	0.647
Madhya Pradesh	0.297	0.607
Maharashtra	0.294	0.516
Karnataka	0.289	0.577
West Bengal	0.287	0.727
Tripura	0.285	0.467
Dadra & Nagar Haveli	0.285	0.784
Kerala	0.279	0.441
Mizoram	0.277	0.565
Rajasthan	0.275	0.619
Andhra Pradesh	0.273	0.629
Assam	0.270	0.548
Uttar Pradesh	0.269	0.601
Uttarakhand	0.268	0.54
Nagaland	0.261	0.342
Goa	0.258	0.307
Pondicherry	0.257	0.786
Tamil Nadu	0.254	0.571
Bihar	0.254	0.67
Jharkhand	0.244	0.627
Himachal Pradesh	0.243	0.422
Delhi	0.234	0.447
Haryana	0.226	0.499
Punjab	0.214	0.511
Chandigarh	0.212	0.419
Jammu & Kashmir	0.190	0.52
Manipur	0.169	0.383

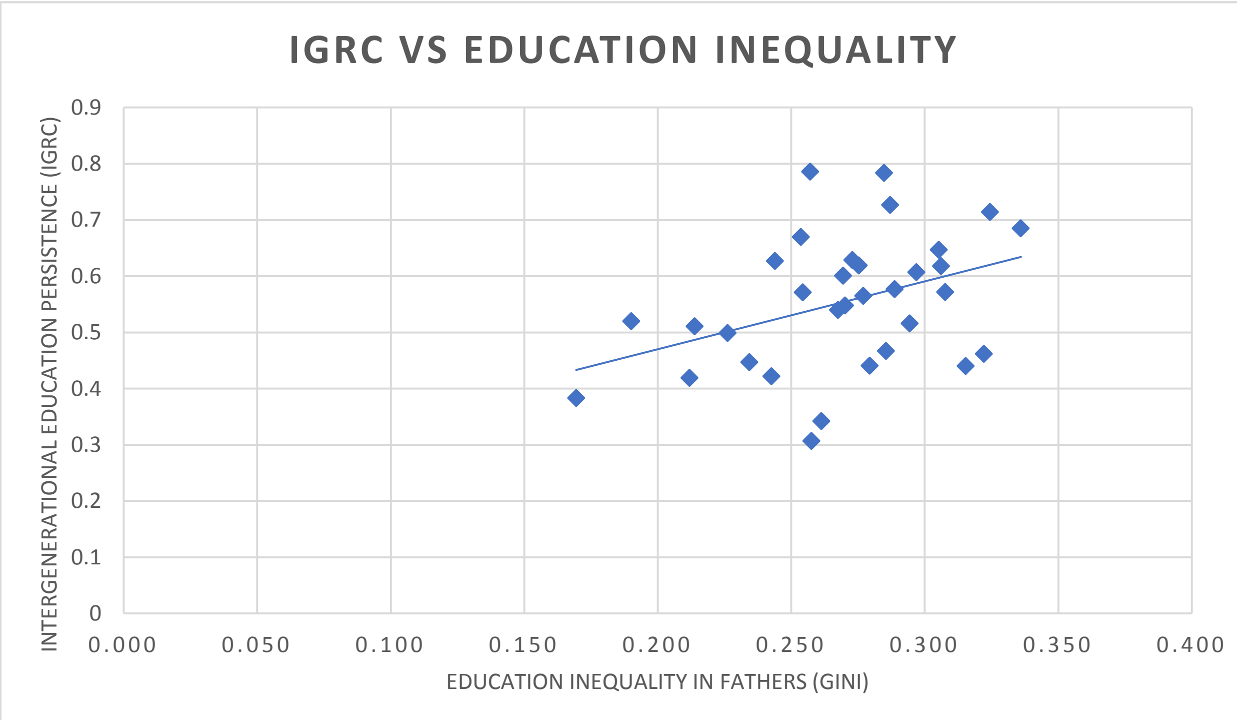

Figure 3. *Intergenerational Regression Coefficient vs Education Gini*

Based on the data, we obtain a Great Gatsby Curve for education. In a state where education inequality is high during father's time, a son's educational attainment and in turn his life chances are mostly determined by his father's education. Hence, in such a state, for a son of a father who is sparsely educated, on an average, it will be difficult to climb the ladder of progress. Such state requires policy attention, and it must duly devise policies that contain education inequality in the first place to ensure that equality of opportunity in education prevails later.

5. Summary and Conclusion

In this paper, we have attempted to discern the role of circumstances in shaping an individual's life chances in India. While an individual's circumstances are proxied by his father's education, his life chances are assumed to depend on his own education outcomes. We have sought to prise out this information by using various methods and for several category and subcategory of individuals. In doing so, we have managed to check and update the numbers on intergenerational education mobility with the aid of the latest data (IHDS-II). More importantly, we have explored the non-linearity in the relationship between educational outcomes of successive generations for various cohorts and regions by employing quantile regressions. Lastly, though not in a causal sense, we find a positive association between education inequality during the previous generation and intergenerational educational persistence within India.

Overall, the sons are more educated than the fathers in the overall sample as well as the sub-groups. Also, there is a clear and steady growth in educational attainment over the years and across generations. Having said that, even after 70 years of independence, a son's life chances are closely tied to his father's relative status in the society. This is reflected in the high values of intergenerational education coefficients that are obtained through various empirical exercises in this paper.

First, for the overall sample, the IGRC varies between 0.522 and 0.588 contingent on inclusion of different controls. Second, across age cohorts, the IGRC displays a decreasing trend, although the fall is not monotonic. The absolute fall in IGRC from 0.75 (for the oldest cohort) to about 0.48 (for the youngest cohort) shows a movement in the right direction towards reducing the inequality of opportunity. Here too, the results are robust to the addition of controls. It must be borne in mind that the figure of 0.48 still points to a high degree of intergenerational education persistence and shows a divergence from much lower values found in Jalan and Murgai (2007) and Maitra and Sharma (2009). Third, a person's caste seems to have a higher effect than his religion or state of residence on his mobility.

In the evolution of education persistence by groups, owing to affirmative action policies by the government since independence, the SCs and STs are closing the gaps on the other castes and now stand less affected by their circumstances as compared to the OBCs, albeit marginally. Muslims, however, continue to languish as their rate of improvement in education mobility is much lesser than other sub-groups.

An important finding of this study is that education mobility is not linear along the distribution of educational attainment of individuals. For the overall sample as well as the sub-groups, sons are most mobile at the top end of the education distribution. Moreover, there has been an improvement in education mobility at almost all points of distribution for the youngest 10-year age cohort as compared to the oldest. A look at the education transition matrices for all groups reveals greater upward mobility than downward mobility in educational achievements for individuals.

Finally, the "Higher Inequality → Lesser Mobility" nexus in education plays out for the Indian scenario and thus corroborates the 'Great Gatsby Curve'. This suggests that a way to ensure equality of opportunity for the future generations is to try and reduce inequality in educational attainment for the present generation. Enhancing public expenditure in education holds the key. This is also to deal with the increased demand for education as sections of the society respond to competitive pressures unleashed by Globalization (Hnatkovska, Lahiri & Paul, 2013). As Iversen, Krishna and Sen (2017) posit, the policy of making education accessible to the masses has worked well in the past in containing the effect of circumstances on individuals' life chances.

References

Alexander, M., Harding, M., & Lamarche, C. (2011). Quantile regression for time-series-cross-section data. *International Journal of Statistics and Management System, 6*(1), 47–72.

Asadullah, M. N., & Yalonetzky, G. (2012). Inequality of educational opportunity in India: Changes over time and across states. *World Development*, *40*(6), 1151-1163.

Azam, M., & Bhatt, V. (2015). Like Father, Like Son? Intergenerational Educational Mobility in India. *Demography, 52*(6), 1929–1959. doi:10.1007/s13524-015-0428-8

Behrman, J. R., Gaviria, A., Székely, M., Birdsall, N., & Galiani, S. (2001). Intergenerational mobility in Latin America [with comments]. *Economia*, *2*(1), 1-44.

Borooah, V. K., & Iyer, S. (2005). Vidya, Veda, and Varna: The influence of religion and caste on education in rural India. *The Journal of Development Studies*, *41*(8), 1369-1404.

Brunori, P., Ferreira, F. H. G., & Peragine, V. (2013). Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons. *IZA Discussion Paper No.7155*, (7155).

Bond, H. C. (2016). Income Inequality and Higher Education.

Bourguignon, F., Ferreira, F. H., & Menéndez, M. (2007). Inequality of opportunity in Brazil. *Review of income and Wealth*, *53*(4), 585-618.

Bussolo, M., Checchi, D., & Peragine, V. (2017). The long-term evolution of inequality of opportunity.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the US. *Executive Summary*.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *NBER Working Paper 19844*, (January), 19844.

Desai, S., Vanneman, R., & National Council of Applied Economic Research, New Delhi. (2005). India Human Development Survey (IHDS), 2005. ICPSR22626-v11. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-02-16.

Desai, S., Vanneman, R. (2011-12). India Human Development Survey-II (IHDS-II), 2011-12. ICPSR36151-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-08-01.

Deshpande, A. (2000). Does caste still define disparity? A look at inequality in Kerala, India. *American Economic Review*, *90*(2), 322–325.

Emran, M. S., & Shilpi, F. (2015). Gender, Geography, and Generations: Intergenerational Educational Mobility in Post-Reform India. *World Development*, *72*(May 2012), 362–380.

Emran, M. S., Greene, W., & Shilpi, F. (2017). When measure matters: coresidency, truncation bias, and intergenerational mobility in developing countries. *Journal of Human Resources*, 0216-7737R1.

Gaentzsch, M., & Roman, G.Z. (2017). More educated, less mobile? Diverging trends in long-term opportunities in Chile and Peru

Ganzeboom, H. B., Treiman, D. J., & Ultee, W. C. (1991). Comparative intergenerational stratification research: Three generations and beyond. *Annual Review of sociology*, *17*(1), 277-302.

Hertz, A., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., & Verashchagina, A. (2007). The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends. *The B.E. Journal of Economic Analysis & Policy*, *7*(2).

Hickey, M. G., & Stratton, M. (2007). Schooling in India: Effects of Gender and Caste. *scholarlypartnershipsedu*, *2*(1), 6.

Hnatkovska, V., Lahiri, A., & Paul, S. B. (2013). Breaking the caste barrier: intergenerational mobility in India. *The Journal of Human Resources*, *48*(2), 435–473. doi:10.1353/jhr.2013.0012

Iversen, V., Krishna, A., & Sen, K. (2017). Rags to riches? Intergenerational occupational mobility in India. *Economic and Political Weekly*, 107-114.

Jalan, J., & Murgai, R. (2007). Intergenerational Mobility in Education in India.

Kearney, M. S., & Levine, P. B. (2014). Income Inequality, Social Mobility, and the Decision To Drop Out of High School. *NBER Working Paper 20195*, 60.

Krueger, A. (2012). The Rise and Consequences of Inequality in the United States, presentation at the Center for American Progress in Washington, DC January 12, 2012.

Maitra, P., & Sharma, A. (2009). Parents and Children: Education Across Generations in India. *Unpublished Manuscript*, (November).

Majumder, R. (2010). Intergenerational Mobility in Educational and Occupational Attainment: A Comparative Study of Social Classes in India. *Margin: The Journal of Applied Economic Research*, *4*(4), 463–494.

Mohammed, A. R. S. (2016). Does a Good Father Now Have to Be Rich? Intergenerational Income Mobility in Rural India.

Motiram, S., & Singh, A. (2012). How Close Does the Apple Fall to the Tree? Some Evidence on Intergenerational Occupational Mobility from India. *WIDER Working Paper No. 2012/101*, (July).

Noah, T. (2012). White House: Here’s why you have to care about inequality. *New Republic. http://www. newrepublic. com/blog/timothy-noah/99651/white-house-heres-why-you-h ave-care-about-inequality*.

Schultz, T. P. (2002). Wage gains associated with height as a form of health human capital. *The American Economic Review*, *92*(2), 349-349.

Tassinari, F. (2017). Intergenerational earnings and income mobility in Italy. New evidence on geographical and temporal differences.

Willen, P., Hendel, I., & Shapiro, J. (2014). Educational Opportunity and Income Inequality. *NBER Working Paper 10879*.

Appendix A

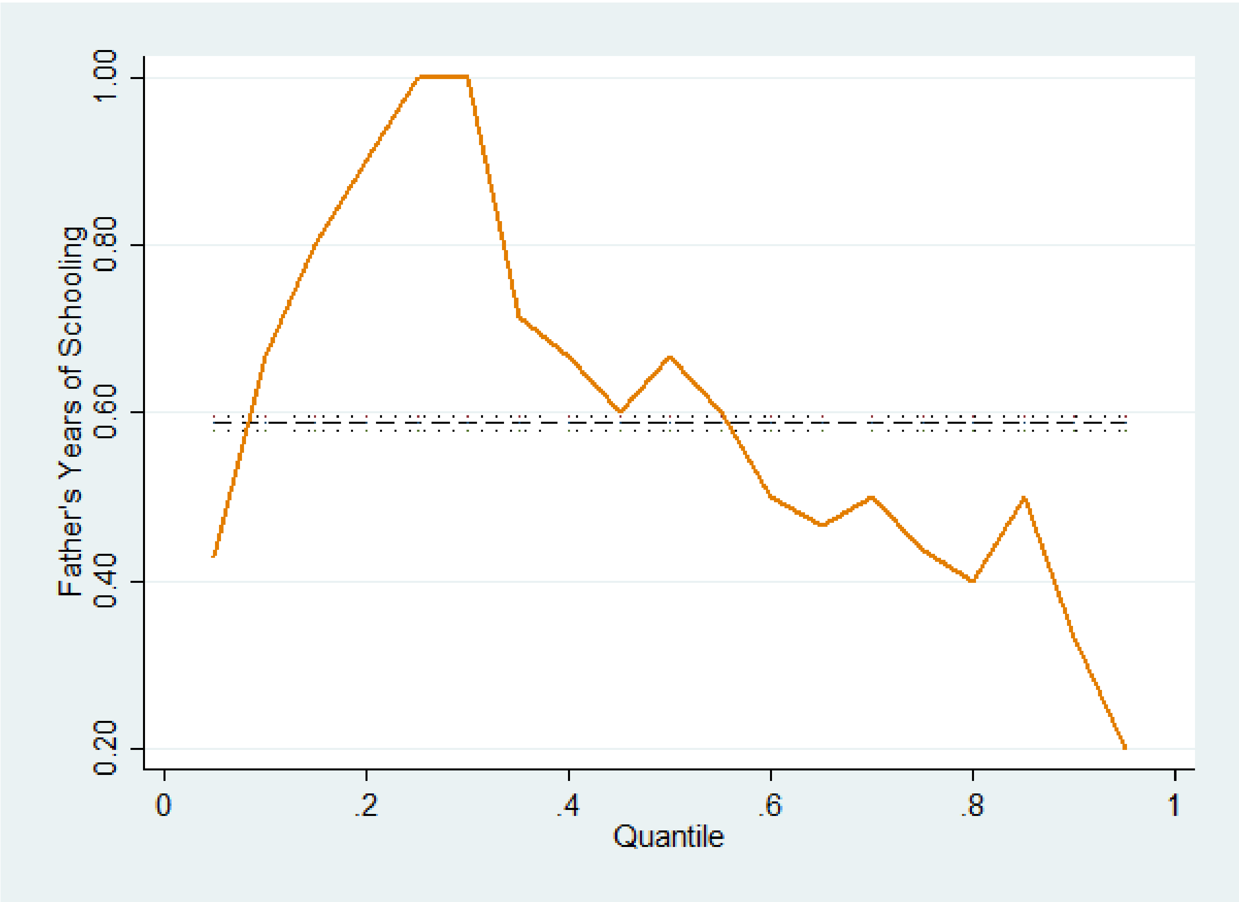

Figure A.1. *Quantile Intergenerational Coefficients of Education (Overall Sample)*

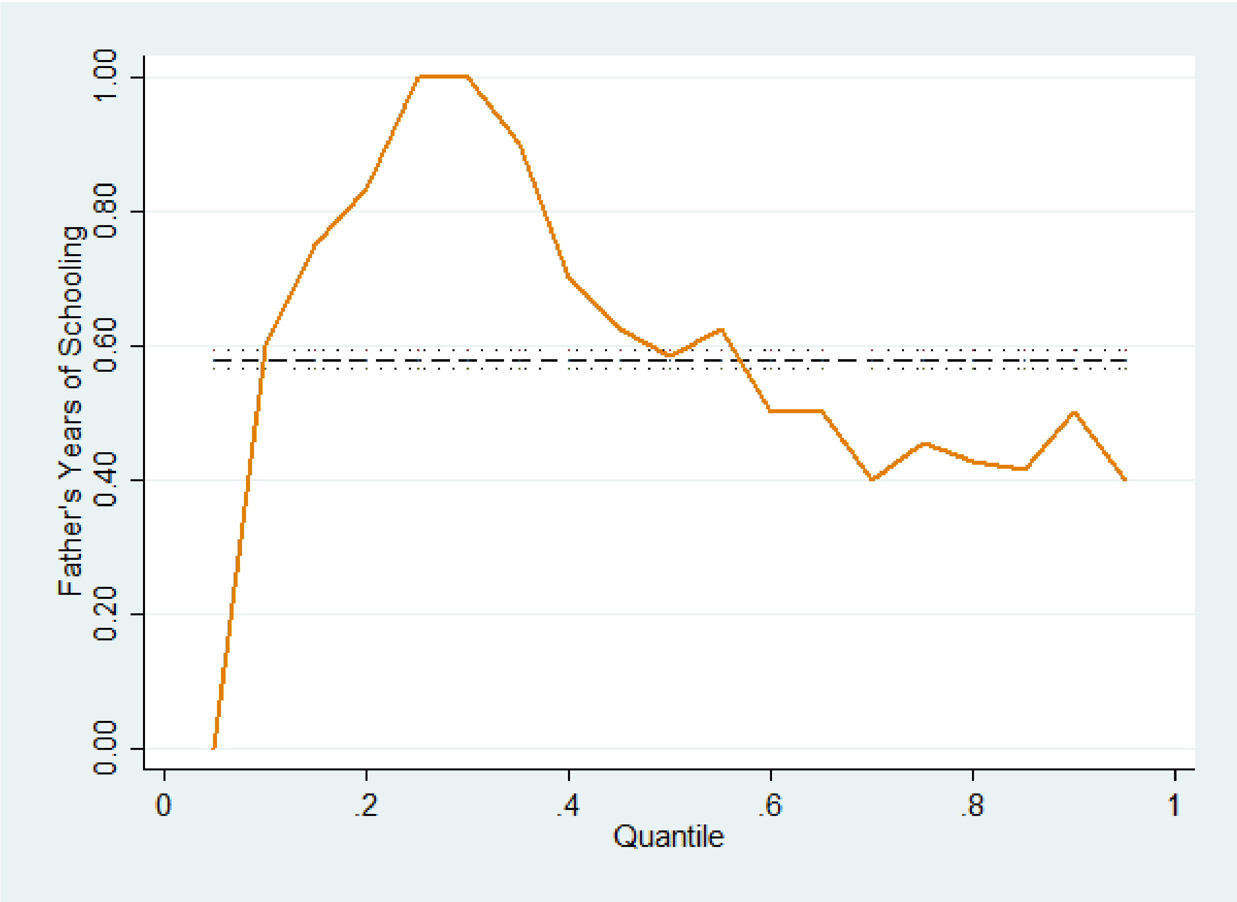

Figure A.2. *Quantile Intergenerational Coefficients of Education (Rural Sample)*

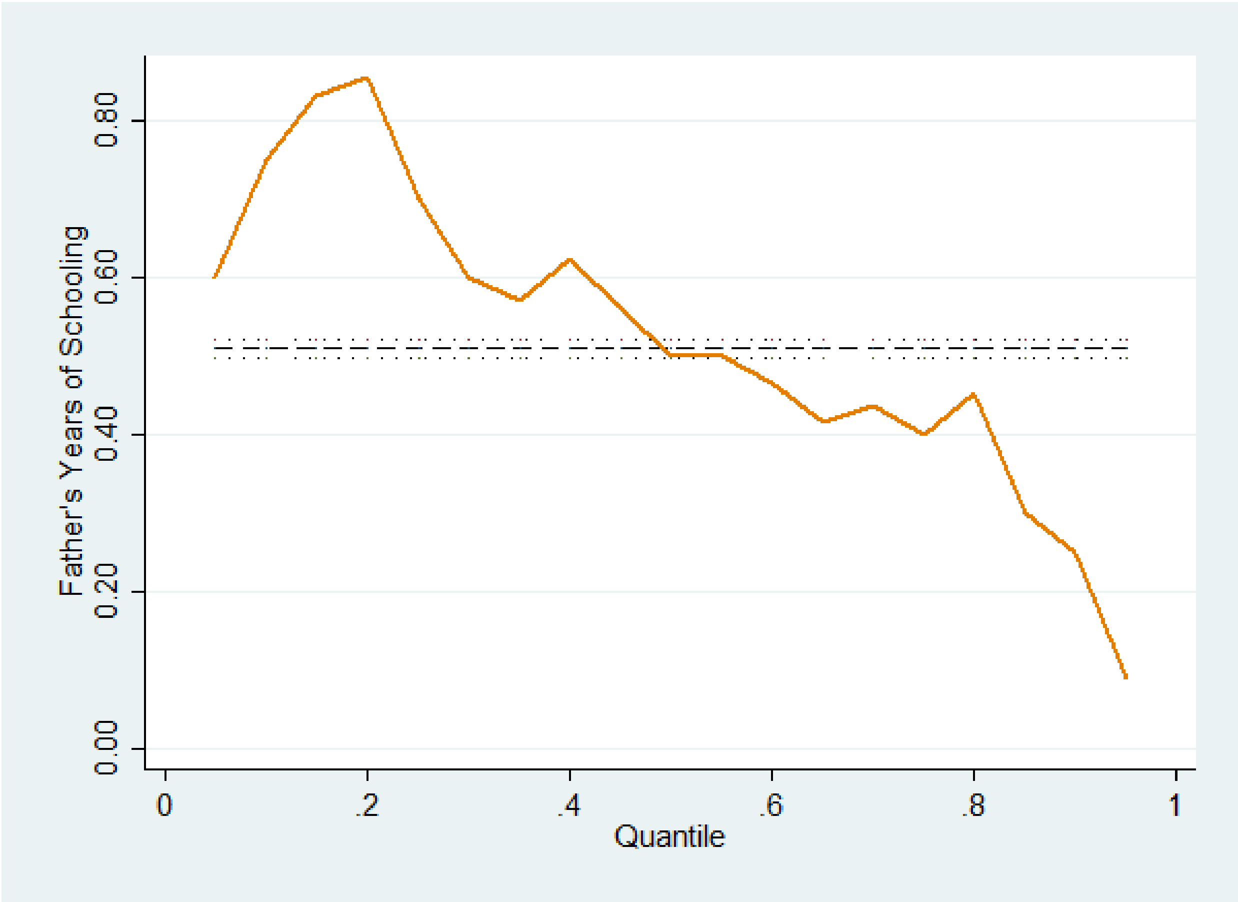

Figure A.3. *Quantile Intergenerational Coefficients of Education (Urban Sample)*

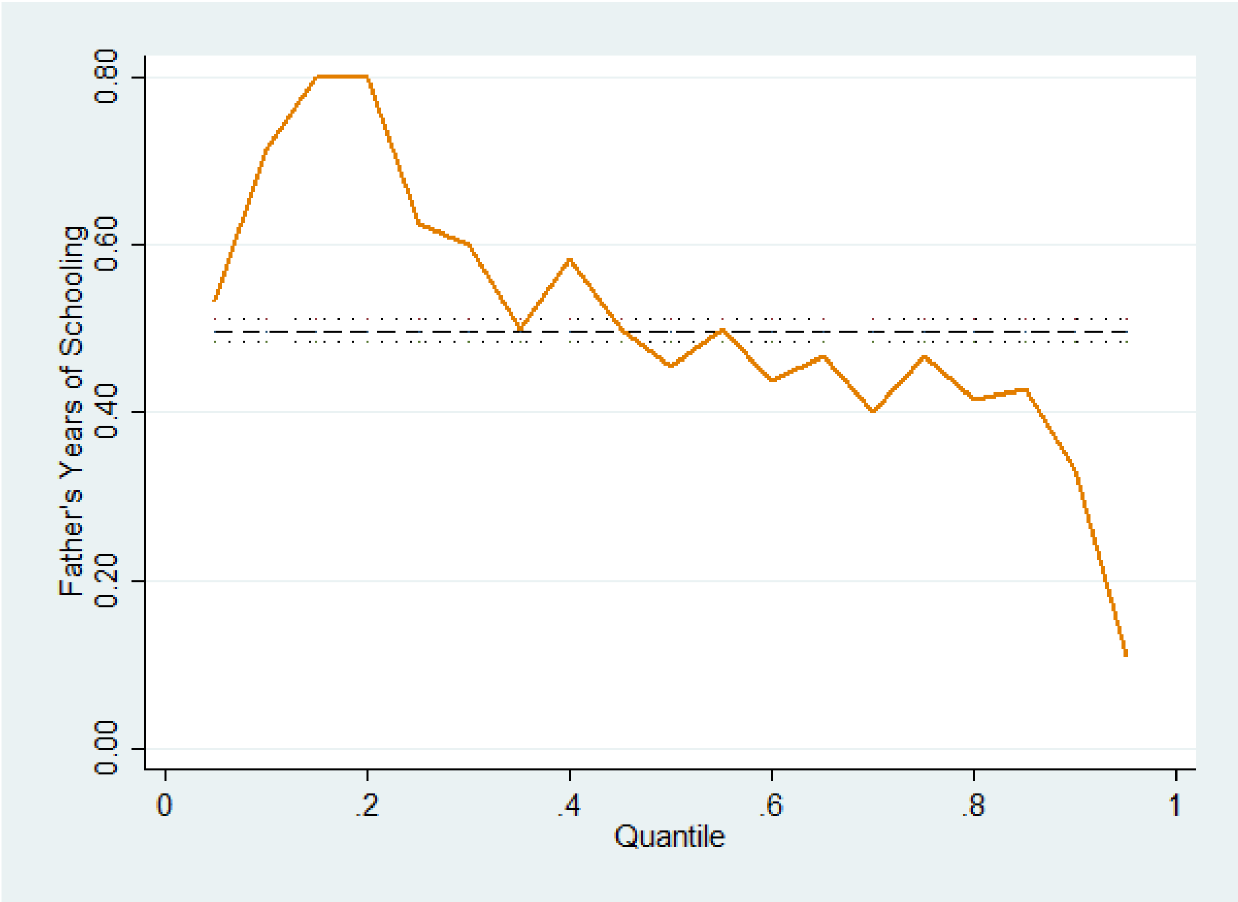

Figure A.4. *Quantile Intergenerational Coefficients of Education (Age Cohort: 25-34)*

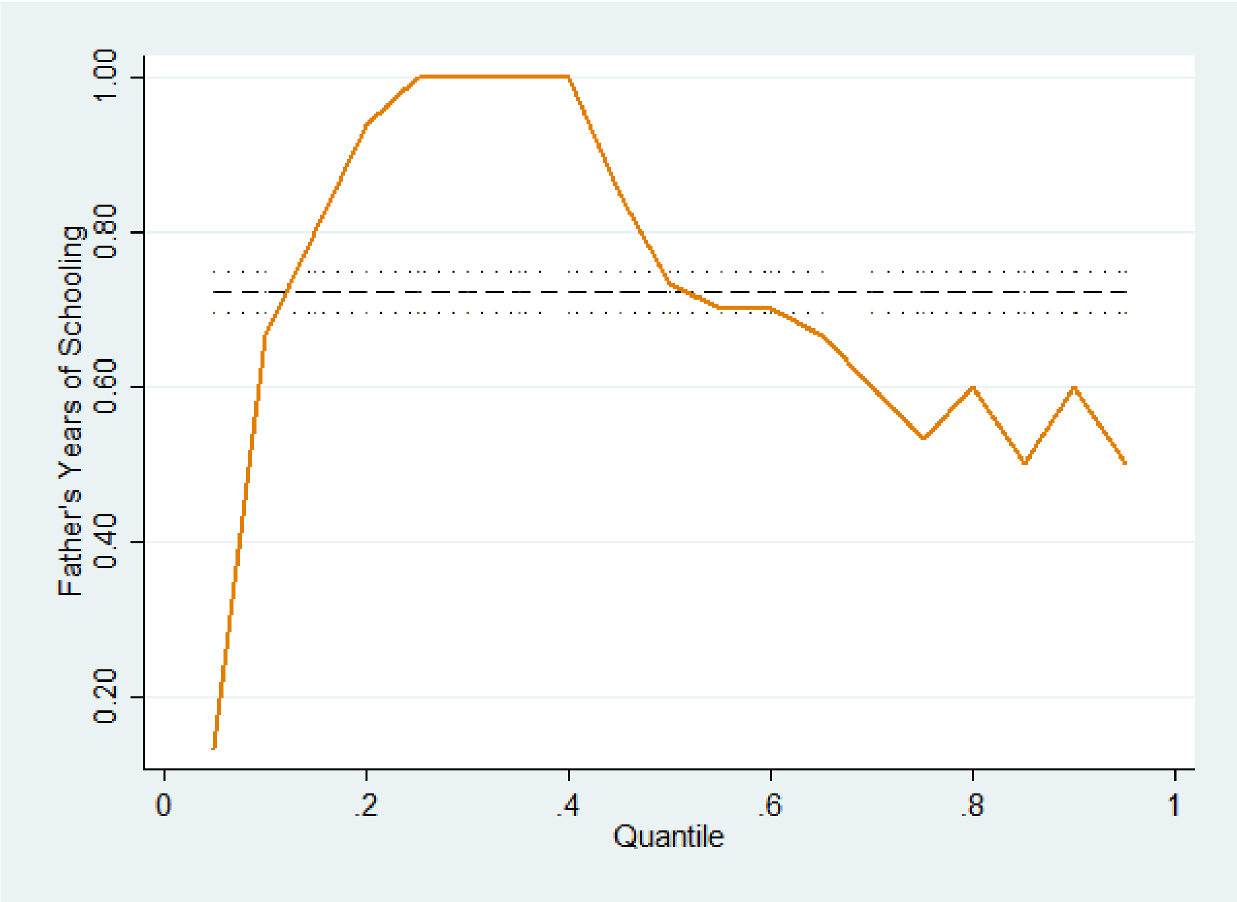

Figure A. 5. *Quantile Intergenerational Coefficients of Education (Age Cohort: 55-64)*

Appendix B

In the overall sample spanning all age groups, fathers are co-resident with their respective sons or daughters in 44.08% of all cases and mothers are co-resident with their respective offspring in 51.17% of the cases. Here, we consider all sons and daughters in the age group 11 - 64 (both ages included) who are no longer enrolled in school and have their respective mothers and fathers residing in the same household. As measures of parental education, we use the number of years of schooling of both the mother and the father of each son/daughter conforming with Jalan and Murgai (2007). Various specifications of the following model are then estimated using an OLS framework.

$$C_i = \beta_0 + \beta_1 F_i + \beta_2 M_i + (\text{Controls}) + \epsilon_i$$

where, C_i is the number of years of schooling of the child (son/daughter), and F_i and M_i are the number of years of schooling of father and mother respectively. The coefficients, β_1 and β_2, are the measures of intergenerational educational persistence emerging from father and mother respectively. Table B.1 contains the regression results.

Table B.1

Intergenerational Regression Coefficients (All India) for co-resident children and their parents (Dependent Variable – 'yrssch')

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	yrssch	yrssch	yrssch	yrssch	yrssch	yrssch	yrssch
fatheryrssch	0.367***	0.355***	0.362***	0.356***	0.338***	0.363***	0.347***
	(0.00631)	(0.00639)	(0.00637)	(0.00631)	(0.00639)	(0.00632)	(0.00637)
motheryrssch	0.275***	0.265***	0.223***	0.273***	0.253***	0.236***	0.216***
	(0.00720)	(0.00729)	(0.00758)	(0.00722)	(0.00733)	(0.00746)	(0.00754)
cons	5.908***	6.893***	7.450***	6.208***	7.100***	7.479***	8.328***
	(0.0376)	(0.120)	(0.206)	(0.0406)	(0.120)	(0.182)	(0.211)
Caste Controls	**No**	**Yes**	**Yes**	**No**	**Yes**	**No**	**Yes**
State Controls	No	No	Yes	No	No	Yes	Yes
Religion Controls	No	No	No	Yes	Yes	Yes	Yes
N	24148	24090	24090	24148	24090	24148	24090
adj. R-sq	0.291	0.295	0.330	0.304	0.312	0.336	0.343

Notes: Standard errors clustered at household level in parentheses; * p<0.05, ** p<0.01, *** p<0.001

In table B.1, the β_1s indicate a much greater degree of intergenerational educational mobility when compared to the IGRCs obtained in the main text of the paper. This sample cannot be considered representative of the population due to the limitation imposed by the co-residency condition. Moreover, the co-resident combinations cannot be argued to be randomly spread in the population. Hence, the coefficients reflect downward truncation bias. The magnitude of truncation bias cannot be calculated as mother's educational attainment cannot be ascertained for the sample, barring the co-resident combinations.