Does A Good Father Now Have To Be Rich? Intergenerational Income Mobility in Rural India

A. R. Shariq Mohammed*
University of Arizona

December 2016

Abstract

To date, few income mobility studies for developing counties have utilized panel datasets linking parents with their children. Institutional features and the absence of reliable panel datasets can explain most of this paucity. Using a unique panel dataset this study contributes to the literature in three ways: 1) it presents intergenerational income elasticity (IGE) estimates for rural India, 2) it finds within and between groups mobility for various socio-economic groups, and 3) it develops strategies that can be used for similar research in other developing countries. Depending on the choice of income scaling parameters, the OLS IGE estimate lies between 0.28 and 0.37, which is relatively lower than available evidence from other developing countries. As the OLS estimator is known to suffer from attenuation bias, I also present the 2SLS IGE estimate as an upper bound on each estimate. Group-wise analysis finds that within-group mobility for disadvantaged groups is relatively high and convergence speed between groups is remarkably low. The between-groups finding, which is corroborated by repeated cross-sections from the National Sample Survey, indicates that India's progress towards cross-caste equality is disappointingly low.

JEL Classification: D31 . I20 . I30 . J62 .

---

*Department of Economics, McClelland Hall 401, PO Box 210108, Tucson, AZ 85721-0108. Email: shariqazeez@email.arizona.edu. I would like to thank my adviser Gary Solon for patient advising and support. I am grateful to Orley Ashenfelter, Christopher Walters, Leslie Dupont, Swapnil Motghare, Sanguk Nam and Arundhati Tillu for valuable comments. I thank seminar participants at the University of Arizona and the IHDS-NCAER users conference. All errors are my own.

1 Introduction

"*If I were giving a young man advice as to how he might succeed in life, I would say to him, pick a good father and mother, and begin life in Ohio.*" [Wilbur Wright, 1910]

In reality, someone who aspires to success never has the option of choosing his or her parents or the environment he or she is born into; however, a positive environment — both at home and in the larger community — may be crucial for a person's lifelong success. At present, a century after Wilbur Wright's advice, how true is it that being born to good parents will significantly improve the odds of success? Or is it now the case that good parents must also be rich and, in the context of India, belong to a higher caste? Using a unique panel dataset, I address such questions for rural India.

Speaking on the eve of India's independence from Great Britain, Jawaharlal Nehru, the first prime minister of India, proclaimed that "we end today a period of ill fortunes and India discovers herself again ... The service of India means ... ending of poverty and ignorance and disease and inequality of opportunity." One can find a number of studies that appraise India's performance when it comes to ending poverty, lack of education, and diseases. But then, how much do we know about India's performance in ending inequality of opportunity? Unfortunately, due to insufficient data and institutional features, India's progress in reducing inequality of opportunity across generations has received very little attention. This absence of results is lamentable because many developing countries, including India, are actively seeking policies that can fast track them on a high growth rate trajectory. The desirability of such policies, however, may turn out to be questionable in the light of what they entail for the equality of opportunities for everyone. Further, given the long history of the discriminatory caste system in India, mobility research is particularly crucial.

One popular concept of equitable access to opportunity is the intergenerational income elasticity (IGE) — a measure that estimates persistence of economic status across generations. A primary requirement for any intergenerational analysis is a reliable panel dataset that links offspring with their parents. Availability of such datasets in developing countries is rare. Even when one finds a panel dataset, it may be fraught with sampling and other selection issues. This situation has led many researchers to work with cross-sectional datasets to make some conjectures about the intergenerational mobility. While such studies offer some light in a sea of darkness, they fail to match the quality of estimates derived from a reliable panel dataset. In this study, I use a unique panel dataset of rural India that — to the current author's knowledge — is the only available panel dataset for rural India. Households in this panel were first surveyed by the National Council of Applied Economics Research (NCAER) in the Human Development Profile of India (HDPI) dated 1993-1994. These households were re-interviewed by the Indian Human Development Survey (IHDS) — a joint collaboration between the University of Maryland and the NCAER — in 2005 and 2012.

In a simple model of intergenerational mobility, a researcher adopts the most parsimonious choice of regressing proxies for a son's economic status on his father's economic status. Using income to proxy for economic status in this model — as it is popularly done in the literature — one can estimate the IGE. However, having access to a panel dataset is not sufficient for estimating the IGE. Another key requirement is to have reliable measures of income. In developing countries such as India, the majority of people are self-employed in agriculture or businesses. Estimating income for such occupations poses serious accounting difficulties. Previous studies have either ignored a significant proportion of population that is self-employed, or they have used self-reported income that may not have been collected with due care. Unlike most other datasets, the HDPI-IHDS panel dataset is among the first set of developing country surveys that contain detailed information on net earnings of the household from farm and non-farm production. Other than collecting information on household consumption expenditures or household assets, which are commonly collected by most other developing country surveys, the HDPI-IHDS queried about 50 different income sources. Though income from all sources is meticulously reported, income from many sources has been reported only at the household level. This feature of the dataset means that working with a comprehensive measure of income that better represents income of a large proportion of the population has a trade-off; rather than working with the usual sons and their father's income, I can only estimate persistence in the income of sons' households with that of their father's household.

Now that we have access to a reliable panel dataset with meticulously reported household income, and that we agree on regressing a son's household income on his father's household income, is that sufficient to find reliable estimates of IGE for rural India? The answer is no. There is at least one more key issue that needs ironing out. A simple glance at the household demographics within and across generations raises a red flag that there exists pervasive heterogeneity in household size. It makes sense not to compare household income for a household with 2 members with that of a household whose size can be as high as 32. This major concern is addressed in this paper by scaling household income. A popular scaling approach, hailed as a best practice in the development literature, is adopted to scale income across households (Deaton and Zaidi, 2002).

With these solutions at hand, we are finally ready to estimate the IGE for rural India. A higher IGE from the benchmark model reflects higher persistence in economic status across generations, and thus, lower intergenerational income mobility. Results from the benchmark model suggest that the OLS IGE lies between 0.28 and 0.37. I also find that the OLS IGE is robust to the choice of a son's age group in 1994 and selection rules. It is possible that the OLS estimates suffer from measurement error and may be downward biased. One possible solution to overcome such bias is to use plausible instruments for a father's income. Though noisy, the 2SLS IGE estimate places an upper bound on each OLS IGE estimate.

In India, comparison of economic performance among various socio-economic groups is often a very crucial policy exercise. Such an interest by policymakers and researchers makes a great deal of sense given the long history of an inequitable and discriminatory caste system. With a resolution to correct for this gross historical injustice, in the mid-twentieth century, India became the first country to experiment with affirmative action for its marginalized groups. Understandably, due to data limitations, comparison of groups is often carried out by studying time series of poverty rates for each group. Useful as they are, reducing poverty rate among certain groups cannot be the end of a government's role in helping its people, especially those who have been discriminated for thousands of years. Thus, a government needs to go beyond poverty rate comparisons and ensure that every group has access to opportunities for improving their relative economic status. With this motivation, I estimate group-wise measures of intergenerational income mobility in rural India. Group-wise analysis finds that within-group mobility for disadvantaged groups is relatively high, which is coupled with a remarkably low speed of convergence between groups. The between-groups finding is corroborated by evidence from the National Sample Survey data. Strikingly, the between-groups rate of convergence in the range 0.7 to 0.8 is comparable to that for most common surnames in the US (Chetty et. al., 2014), and also to Gregory Clark's (2014) finding.

In an attempt to answer the opportunity question for rural India, this paper contributes to the literature in at least three important ways. It provides reliable IGE estimates for rural India for a period spanning 1994-2012. In fact, any study pertaining to rural India is crucial because historically, the majority of the Indian population has lived in rural areas.[1] Second, this paper finds intergenerational income mobility estimates for major socio-economic groups. Results from such analysis will provide useful insights into the debate on affirmative action for disadvantaged groups

[1] Around 85% of India lived in rural areas during 1950, 74% in 1990, and current estimates suggest that around 70% of the Indian population continues to live in rural areas (Census 2010).

in India, which has been in place for more than six decades. Lastly, this paper addresses a few institutional features related to developing countries and adopts strategies to tackle them. Useful lessons can be learned from this study for other intergenerational income mobility studies focusing on developing countries facing similar challenges.

The remaining paper is organized as follows. In section II, I present evidence of intergenerational mobility from other developing countries. Section III describes the data. Section IV presents main findings of the paper. In section V, I present various robustness exercises. Section VI discusses the main findings of this paper.

2 Income Mobility in Developing Countries

To date, few studies on income mobility in developing countries have used panel datasets, while a vast majority of available studies rely on cross-sectional datasets. Following are two examples of panel dataset based studies. Using Panel CASEN survey from Chile, Celhay et. al. (2011) estimate an IGE of 0.59 for men in Chile. The second study is by Lilland and Kilburn (1995), who use a series of wages for a son and his father from the Malaysian Family Life Surveys to find an IGE estimate of 0.26 for Malaysia. The cross-sectional studies, on the other hand, mostly adopt co-resident households or two-sample instrumental variables (TSIV) approaches in their analysis. These approaches, though restrictive, are adopted when panel datasets are insufficient. In a co-resident household, adult children reside in the same house as their parents. Such a sample allows researchers to observe a father's and a son's incomes in a single cross-sectional dataset. On the other hand, in a TSIV approach, researchers predict fathers' earnings from a secondary sample and use the predicted earnings to estimate the IGE. Evidence from cross-sectional studies suggests an IGE in the range of 0.4 to 0.5 for Taiwan (Chu and Lin, 2016), 0.52 to 0.54 for Urban Chile (Nunex and Miranda ,2011), and 0.68 for Brazil (Dunn ,2007). Hnatkovska et. al. (2013) find an IGE estimate of 0.55 for scheduled caste and scheduled tribe (SC/ST) households and 0.61 for non-SC/ST households in India. Evidence from urban China suggests an IGE of 0.74 for father-son, 0.84 for father-daughter, 0.33 for mother-son, and 0.47 for mother-daughter pairs (Gong et. al., 2012). There is also evidence of high intergenerational income elasticity among fast-growing Asian Tigers: Korea (Choi and Hong 2011; Ueda 2013), Singapore (Ng 2007; Ng, Shen, and Ho 2009), and Taiwan (Kan, Li, and Wang 2015; Sun and Ueda 2015). As mentioned above, cross-sectional studies mostly rely on either a co-resident sample, or they predict a parent's economic status based on limited information available during offspring's generation. Findings in the mobility literature suggest that using non-representative samples, observing fathers and sons during different stages of their life cycle, or bias in predicting father's income generally result in a downward bias in the IGE estimates (Solon, 1992; Haider and Solon, 2006; Chu and Lin, 2006).

3 Data

For analyzing intergenerational income mobility, an ideal dataset would be a panel that is representative of the concerned population; and it must contain mid-career income information for fathers and sons. The panel used in this paper comes from the Human Development Profile of India (HDPI) and the Indian Human Development Survey (IHDS). This section will provide basic details about the construction of panel, measurement of income, approach for scaling household income, and descriptive statistics.

3.1 Panel

The HDPI data was collected by the NCAER in 1993-1994. The IHDS, which is a joint collaboration between the University of Maryland and NCAER, collected the follow up data in the years 2004-05 and 2011-12. The original HDPI survey included a sample of 33,230 households that was representative of the rural Indian population. When the IHDS conducted a similar survey in 2004-05, its goal was to select around 50% of the HDPI sample and then to add a refresher sample that will make it nationally representative for both the urban and the rural Indian populations. The households that the IHDS selected from the original HDPI sample for re-interview were chosen in such a way that, starting from the year 1993-94, the panel sample continued to be representative of rural India.[2] This study focuses on rural India because the 1993-94 NCAER survey was limited to this region. Among other things, a nice feature of the IHDS dataset is that it followed every separated household as long as they continued to live in the same village. This feature averts having to work with only coresident households, which are most likely special because the choice of adult/married sons to live with their parents is non-random and is effected by factors such as local customs, social norms, and land ownership (agricultural/non-agricultural).

[2] Later in the robustness analysis, I show that the panel is representative of rural India, as it was originally intended.

3.2 Income

Measurement of income is one of the most common problems that researchers face when studying income mobility in developing countries. In most datasets, earnings are reported for only those individuals that receive wage and salary income for their work. That is to say, information on wage and salaries will be reported as missing for a significant proportion of the population that works on their own farm or businesses. If one were to estimate intergenerational income mobility for only those individuals who report wage and salary earnings, then results will most likely be biased. The 66th round of National Sample Survey reported that 51% of India's total workforce is self-employed, only 15.5% receive wages or salaries, and 33.5% work as casual laborers. The number of people self-employed in rural areas were found to be 54.2%, much higher than the self-employment proportion of 41.4% within urban areas. In light of these facts, it becomes clear that restricting sample to households who report positive earnings will result in biased estimates due to sample selection.

The first challenge in measuring income for individuals that are self-employed in farm or non-farm production is that the detailed information on net earnings from these activities is difficult to collect. Unlike most other datasets, the HDPI-IHDS panel dataset is one of the first developing country surveys that contains detailed information on net-earnings of the household. Other than collecting information on household consumption expenditures or household assets, which are commonly collected by most other developing country surveys, the HDPI-IHDS survey queried about 50 different income sources. For farm income, one of the most difficult income source to estimate, HDPI-IHDS collected data on crop production and prices, use of crop residues, animal ownership and home-produced animal and crop products, expenses for a variety of farm inputs, and agricultural rents paid and received. Working with such comprehensive measure of income, I use state level poverty lines to convert household income in real terms. In my analysis I express all consumption expenditure values in 1993-94 rural Andhra Pradesh prices.[3]

3.3 Scaling of household income

This study uses total household income instead of individual income because HDPI-IHDS reports income from many sources at the household level. Doing so averts the problem of working with only those individuals who report wages and salaries. In dealing with total household income, one needs to address the issue of comparability of income across households. When a developing country like

[3]For 1993-94, I use state level poverty lines derived from the Lakdawala approach. For 2011-12, I estimate the Lakdawala lines by adjusting the 2004-05 Lakdawala lines using the index implicit in the official Tendulkar lines for 2004–2005 and 2011–2012.

India, the institution of joint family is quite prevalent. In a joint family, married adults live with their parents, siblings' families and other relatives. Figure 1 presents the histograms of household size in a fathers' and a sons' generation. Although household size for the most part lies between 2 and 11, the right end of the distribution can get as high as 32. Given such pervasive heterogeneity in household size, a simple comparison of total household income can no longer be used to infer the economic status of a member of the household.

I standardize total household income by the number of adults and children, and the standardized income will henceforth be referred to as scaled income. In the development literature, various scaling approaches are commonly adopted to standardize household aggregates such as consumption expenditure. The following scaling approach, hailed as the best practice in the development literature (Deaton and Zaidi, 2002), is adopted in this paper:

$$\text{Scaled HH Income} = \frac{\text{Household Income}}{\text{Adult Equivalents}}$$

where Adult Equivalents $= (\text{Adults} + \alpha * \text{Children})^{\theta}$. Members of the household older than 15 years are treated as adults, and those below 15 years of age are treated as children. α is the cost of a child relative to that of an adult, and it lies between 0 and 1. When α takes the value of '0', it means that there is no additional cost to having children. On the other extreme, when α takes the value of '1', it means that having a child costs as much as having an adult in the household. The second parameter in the adult equivalents equation is θ, which is a measure of economies of scale when subtracted from '1', i.e., $(1-\theta)$. When θ is '0' (i.e., $(1-\theta)$=1), there are perfect economies of scale. In this case, all consumption goods are public goods, so whether there is one adult or more, they all share the same goods and no cost will exit for having additional adults. On the other extreme, when θ is '1' (i.e., $(1-\theta)$=0), there are no economies of scale. That is to say, all consumption goods are private goods and a household needs to spend the same amount of money for every additional adult.

Findings in the development literature suggest that the cost of raising children in industrialized economies is relatively higher than in poorer agricultural economies. Another set of findings in the literature suggest that the share of budget spent on private goods, for example food and clothing, is very high in poorer countries, indicating that the economies of scale must be very limited in the these countries (Deaton and Zaidi, 2002). From these two findings, it is often recommended in the literature that for poorer economies the researchers will do good by selecting a very low value of α and a very high value of θ. In this paper, I first present results corresponding to all possible combinations of the income scaling parameters α and θ, and afterwards, following the recommendation in the development literature, I present results corresponding to the values α=0.3 and θ=0.9.

3.4 Descriptive Statistics

The 1994-2012 panel consists of 1,776 father-son pairs. The households in the panel were first surveyed in 1994, and they were subsequently resurveyed in 2005 and 2012. For the main sample, I restrict attention to only those households that included a son in 1994 with an age lying between 16-22 years. Eighteen years later, in 2012, the sons were observed again when they were 34-40 years old. I drop observations if difference between 2012 reported age and 1994 reported age differs from 18 by more than 5-years.[4] The observations are also dropped if the son reported zero hours of work during 2011-12.[5]

Summary statistics for the 1994-2012 panel are presented in Table 1. Column 1 in this table pertains to fathers that were observed in 1994. Sons were observed in 2012 and their details are presented in column 2. Sons' average age in the panel is 36.47 years and that for fathers is 50.48 years. From the point of view of capturing long term income, it is comforting that both father and son are observed towards the mid or late-mid stages in their working career. Average household size is slightly lower, and literacy and income are on average higher in sons' generation.

4 Results

This section provides results from ordinary least squares (OLS) and two stage least squares (2SLS) approaches. Given a long history of the caste system, inequitable access to opportunities among various socio-economic groups takes a central stage in wide array of policies such as affirmative action and various other planned programs. As such, I then provide a detailed analysis of within and between groups IGE, including similar results from multiple rounds of National Sample Survey.

[4]A few households who reported negative income were dropped to avoid problems with natural log being undefined over the negative values.

[5]Later I show that main results do not change in any meaningful way when I relax these regularity conditions.

4.1 OLS Results

In line with the literature, I use the following parsimonious specification to estimate the IGE:[6]

$$\log(y_{son,i}) = \alpha + \beta\log(y_{dad,i}) + \text{age}_{dad,i} + \text{age}^2_{dad,i} + \text{age}_{son,i} + \text{age}^2_{son,i} + \varepsilon_i$$

Where $\log(y_{son,i})$ is the natural log of sons' income and $\log(y_{dad,i})$ is the natural log of father's income. As discussed above, by income I refer to the scaled real household income. The control variables are age of father, age of son, and their squares. These age covariates control for any age related variation in each generation that may be affecting the estimate of persistence. The resulting co-efficient of interest that gives us the value of the IGE from this specification is β. This coefficient captures the persistence in the economic status of sons compared to that of their father, within their respective generations. The measure of mobility can be derived by subtracting IGE from one, i.e., $1 - \beta$.

The OLS IGE estimate lies between 0.28 and 0.37, depending on the income scaling parameters. Panel (A) of Figure 2 depicts the IGE for all possible combinations of α and θ. The first column of Table 2 presents coordinates at the boundary points of this figure. The OLS IGE takes the lowest value of 0.28 when there are zero costs for raising children and there are no economies of scale. Clearly, both these assumptions are not supported by evidence. On the other extreme, the OLS IGE reaches the maximum of 0.37 when there are perfect economies of scale. Going forward with the the a specific profile of parameters $\alpha = 0.3$ and $\theta = 0.9$, Table 3 presents results from the OLS specification for various combinations of sample selection rules. The OLS IGE is 0.30 when a son is dropped if his reported age in 1993-94 and 2011-12 differs from 18 by more than 5 years and if he has not worked during 2011-12. Results from different combinations of these two sample selection rules are almost same. The OLS IGE for an age error tolerance of 10 years and non-zero work hours is 0.30, that for an age error tolerance of 5 years alone is 0.31, and that for an age error tolerance of 10 years alone is 0.31. These estimates suggest that the OLS IGE estimation is not sensitive to any particular combination of sample selection rules. Thus in the following analysis I adopt sample selection rules from the first column.

Are these OLS IGE estimates sensitive to the choice of a son's age group in 1993-94? Table 4 presents findings from the OLS specification for various age groups. The OLS IGE estimate is 0.30 for a random sample of sons who were between the ages 16 and 22 in 1993-94. Results are very

[6] Here and in all subsequent analysis, I use sample weights from the 1993-94 HDPI Survey. Results are very similar when I use IHDS 2012 weights instead.

similar when lower or upper bound is changed by 2 years. The OLS IGE for sons belonging to the age group 14 and 22 is 0.29, that for age group 16 and 20 is 0.33, and that for age group 14 and 20 is 0.31. These estimates suggest that the OLS IGE estimation is not sensitive to the choice of son's age group. Unless otherwise specified, in the following analysis, I work with sons who were between the ages 16 and 22 in 1993-94. In summary, the OLS IGE results suggest that estimates are robust to the choice of selection rules and age group choice, and the IGE lies between 0.28 and 0.37 depending on the choice of income scaling parameters.

4.2 2SLS Results

Reported measures of income are often known to suffer from errors-in-variables problem. This may happen because the respondent recalls his annual income with some error, or perhaps due to transient income shocks his income from the preceding year may not reflect his long-term income. Assuming that the errors-in-variables problem in reported incomes is of the classical measurement error type, I can focus on finding ways that can reduce this problem. Two widely used techniques in the mobility literature to address such a problem are averaging a father's income from multiple years and using instruments for the father's income. Since I do not have multiple observations for a father's household income in 1994-2012 panel, I therefore adopt the second approach.[7]

The instrumental variables that I propose to use for a father's household income, all of which are observed during the year 1993-94, are father's years of education, whether a household owns a radio or transistor, and whether a household owns a bicycle. Evidence suggests that a father with higher income will in general have higher years of schooling, own a radio or a transistor, and own a bicycle. Years of schooling has previously been used as an instrument for father's income (e.g., Solon, 1992). During the 1990s, when television and cars were not common in rural India, radio was the main source of household entertainment and bicycle was the main mode of local transportation. An advantage of radio and bicycle as instruments is that they are short term assets. Eighteen years later, these assets would be physically depreciated and perhaps out of date also, thus not playing any crucial role in a son's household income.

It is hard to contend that these instruments satisfy exclusion restriction. Consider father A and father B with similar incomes. Suppose that the years of schooling for father A is greater than that

[7] Although household information is available for the year 2004-05, I do not use father's income from this year because of two reasons: first, income information may not be available for all the fathers; and second, sons may start contributing to the household income by 2005, which will result in the problem of regressing son's income against some mixture of son's income.

of father B, and only father A owns both a radio and a bicycle. Now it makes sense to argue that the son of father A is more likely to do better than son of father B. More education, and holding of assets such as radio and bicycle by father A may improve the performance of his son in school, provide his son with better information about the world through radio, and improve his son's access to local opportunities by providing him a bicycle. Nevertheless, as shown by Solon (1992), under plausible assumption that instruments are positively correlated with the error in the benchmark specification, the inconsistency in the 2SLS estimates will be in the upward direction.[8] Thus, the true IGE will be bracketed by the OLS and 2SLS estimates.

Panel(B) of Figure 2 and second column of Table 2 show how the IGE 2SLS estimate varies depending on the choice of income scaling parameters. As for the OLS IGE, the 2SLS IGE reaches its minimum of 0.66 when there are zero costs of raising children and there are no economies of scale. The 2SLS IGE reaches its maximum of 0.80 at all of the remaining three boundary points. Table 5 presents a 2SLS estimate of 0.73 for a specific profile of scaling parameters: $\alpha = 0.3$ and $\theta = 0.9$. Results from the first stage are shown in the bottom section of this table. A glance at the first column, that uses all three instruments, reveals that all three instruments are significantly positive predictors of logarithm of father's scaled real household income in 1994. The F-statistic from the first stage is 23.18. With a 2SLS estimate being so different from the OLS estimator, the Hausman test rejects the null that the logarithm of father's scaled real household income is exogenous. It is also not surprising that the Sargan and Bassman over-identification test rejects the H_0 that the instruments are valid. Column 2, column 3, and column 4 of table present results when only one of the three instruments is used in the first stage. All three instruments, on their own, are significantly positive predictors of father's income. Except for ownership of bicycle, other two instruments provide a significantly positive estimate of the 2SLS IGE. The 2SLS IGE measure from education as an instrument is 0.95, and that from the ownership of radio is 0.590. In summary, for the income scaling parameters $\alpha = 0.3$ and $\theta = 0.9$, the IGE is bracketed between 0.30 and 0.73. Working in the same way, upper bound on the OLS IGE can be obtained for all other combinations of α and θ.

[8] And in case the assumption does not hold, even the 2SLS estimator will underestimate the true IGE.

4.3 Group Analysis

Historians have traced the prevalence of discrimination due to India's caste system to as far back as 1500 BC (Daniels 2010). In order to correct for injustices done to disadvantaged groups, in 1950 the Indian constitution established a plan of affirmative action in employment, education, and politics. These measures were originally intended to last for only 10 years, but they were regularly extended given the perception that the condition of these disadvantaged groups remained to be subpar compared to the rest of the population. Even after 65 years, there is almost a complete consensus among political circles to continue affirmative actions, at least for now.

Table 7 presents sample shares for the six major socio-economic groups in India, of which 23% are Brahmins and forward castes . For thousands of years, these higher castes have been among the most privileged groups. The average log incomes of fathers' and sons' households from this group is towards the high end of the income distribution. The next group is the Other Backward Classes (OBC) that constitutes about one third of the sample. This group is now recognized to be economically backward and is entitled to 'targeted' affirmative action — affirmative action that is available only to those households whose annual income falls below certain level as set by the state government. Next two groups are Scheduled Castes (SC) and Scheduled Tribes (ST), which together constitute about 30% of the sample. These groups are considered to be the most disadvantaged groups and they were amongst the first set of groups that were provided with the affirmative action in 1950. Unlike the OBCs, affirmative action is available to all members of these two groups irrespective of their income and other characteristics. Next group is Muslims, which constitutes about 8% of the sample. Specially in the aftermath of the Government of India's Sachar Committee report (2006), Muslims are widely recognized as being economically backward, as their performance in many socio-economic measures is worse than the SCs and the STs. Finally, the other category includes all other minorities.

Hertz (2005) has shown that the IGE estimator can be decomposed into a weighted sum of within and between groups estimators, implying that higher within-group and higher between-groups estimators result in a higher overall IGE. Within-group mobility informs us about the extent to which children are mobile with respect to their parents within a certain group, and it is measured by running the benchmark IGE model only for the members belonging to a group. Between-groups estimator, on the other hand, captures the degree to which people belonging to a group fall above or below the sample mean.[9] The decomposition of overall IGE from Hertz (2005) for a two-groups case is as follows:

$$\hat{\beta} = \hat{\beta}_1\Big(\hat{\pi}_1 \frac{\hat{\sigma}^2_{x1}}{\hat{\sigma}^2_x}\Big) + \hat{\beta}_2\Big(\hat{\pi}_2 \frac{\hat{\sigma}^2_{x2}}{\hat{\sigma}^2_x}\Big) + \delta\Big(\frac{\hat{\pi}_1\Big(\bar{x}_1-\bar{x}\Big)^2+\hat{\pi}_2\Big(\bar{x}_2-\bar{x}\Big)^2}{\hat{\sigma}^2_x}\Big)$$

Where $\hat{\beta}_g$ is the within-group estimator, δ is between-groups estimator, $\overline{x}$ represents the mean of log scaled real household income for a father, $\hat{\sigma}^2_x$ is variance of log scaled real household income for a father, and $\hat{\pi}_g$ is the group sample share. It is immediate from the decomposition that the overall IGE varies positively with changes in within and between groups estimators. Interestingly, when comparing two groups, between-groups estimator becomes the Wald estimator: $\delta = Wald\ Estimator = \dfrac{\overline{y}_{g1} - \overline{y}_{g2}}{\overline{x}_{g1} - \overline{x}_{g2}}$, where $\overline{y}$ represents the mean of log scaled real household income for a son, and $\overline{x}$ represents the mean of log scaled real household income for a father. It is easy to see that a Wald estimator traces overtime convergence in the log real income between groups. Wald being greater than 1 (less than 1) demonstrates overtime expansion (contraction) in the real wage gap between groups.

Results from within-group and between-groups estimators are presented in Table 6. Group sample shares are presented in the first column. Log of father's scaled real household income and log of son's scaled real household income are presented in column 2 and column 3. Column 4 presents the within-group mobility, and the last column presents the between-groups estimator. Panel(A) of this table compares SC/STs with everyone else. Log of scaled real household income is relatively lower for SC/STs when compared with that for Non-SC/ST groups. Within-group IGE for SC/STs is lower and statistically different from that for Non-SC/ST groups, indicating that mobility is relatively high within the SC/STs. Between-groups estimator, which is a simple Wald estimator in a two groups comparison, is found to be 0.74. As discussed in the foregoing analysis, Wald being close to one suggests that convergence in log real incomes between groups is very slow. Panel(B) presents results when Muslims, another widely acknowledged disadvantaged group, are included with the SC/STs. Results from the SC/ST/Muslims vs the rest are similar to those from panel(A). In panel(C), I include the Other Backward Classes (OBC) with the SC/STs and Muslims, which results in a relatively higher within group IGE and a lower Wald estimator as compared to those from panel(A) and panel(B). These changes are in the expected direction as every household among the OBCs is not considered to be disadvantaged, and this is reflected in the fact that affirmative action for the OBCs is targeted. Since I do not have a clean measure of the OBC households

[9] By scaling the between-groups effect with appropriate variances, Hertz (2005) further decomposes overall between-groups effect across groups.

who qualify for targeted benefits, it is not possible to separate the disadvantaged OBC households from others. For this reason, preferred estimates of within and between-groups estimators in rural India are from panel(A) and panel(B). In summary, results suggest that within-group mobility among disadvantaged groups is relatively high, which is coupled with a remarkably low speed of convergence between groups.[10]

Strikingly, the between-groups degree of convergence is comparable to that for most common surnames in the US (Chetty et. al., 2014), and also to Gregory Clark's (2014) finding. A rather strong hypothesis from Gregory Clark's (2014) work is that the persistence rate in the social law of mobility is in the region of 0.7-0.8, which means that it will take many hundreds of years for families who are initially far below or far above the mean to regress towards the mean of the society. A testable prediction from Clark's hypothesis is that intergenerational correlation based on group-average estimation should be in the range of 0.7 to 0.8. Presenting evidence from group-average based intergenerational studies, Solon (2015) does not find support for the high persistence rates as hypothesized by Clark. Interestingly, Chetty et. al. (2014) obtain an estimate of 0.81, similar to Clark's hypothesis and also to the finding in this paper, from their US income tax data with millions of observations when they estimate group-average based regression on the seven most common surnames. This led them to conjecture that (p.1575) "(Clark's) focus on distinctive surnames partly identifies the degree of convergence in income between racial and ethnic groups (Borjas 1992) rather than across individuals ...". The low rate of convergence for rural India is in agreement with other evidence on growing inequality. In its recent regional economic outlook of Asia and Pacific, International Monetary Fund warned about rampant and growing inequality in India. The report found that India's Gini coefficient rose to 51 by 2013, from 45 in 1990.

4.4 Between-Groups Convergence: Evidence from the NSS data

A striking finding from the previous section is that the rate of convergence in log real household income between groups has been very slow. Can we trust this finding, or is it an artifact of some data problems? In this section, I use multiple rounds of an alternative cross-sectional dataset to analyze between-groups convergence. If these estimates turn out to be similar to previous results, it will corroborate the between-groups finding obtained from the HDPI-IHDS panel dataset.

I use the National Sample Survey (NSS) to verify between-groups finding from the HDPI-IHDS

[10]Table 7 presents results from six-groups comparison. Though noisy and that within-group estimates are in most cases statistically not different, the between-groups estimator continues to suggest a slow rate of convergence among groups.

panel dataset. The NSS is the largest cross-sectional socio-economic dataset available for India. I use three employment and unemployment survey rounds — Round 38 (1983), Round 61 (2004-05) and Round 68 (2011-12) — in my analysis. Each round includes information on 100,000 to 120,000 households and 460,000 to 600,000 individuals. However, the only available income information from the NSS dataset is the wage earnings from the previous week. For my analysis, this is a highly inadequate measure of income for at least two reasons: 1) only wage earnings from external sources are reported, which ignores a majority of households who derive their income from work on their own farms or businesses; 2) wage earnings are reported only from the previous week, which is a poor measure of potential long-term earnings. Given such limitations, I work with consumption expenditure and years of education instead of income.[11] Evidence in the development literature suggests that consumption is not affected by short-term and idiosyncratic shocks to a household income, suggesting that consumption can be a reasonable replacement in the absence of reliable income measures.[12] Moreover, years of education are commonly known to be among the important determinants of income.

Although NSS is a cross-sectional dataset, I can still estimate between-groups estimator by utilizing information on group average consumption. However, an interpretation of between-groups estimator holding across generations will be warranted if the gap between two survey rounds is long enough to encompass a generation. This explains my choice of 1983 and 2011-12 as the two preferred ends of the analysis, where an interim period of thirty years will most likely cover two generations. I also check for the robustness of results by using an alternative pair of survey rounds with a relatively smaller interim gap: Round 38 (1983) and Round 61 (2004-05). I observe SC, ST and Muslim identity of households in every round, but the OBC identity is not available for the Round 38 (1983). For this reason, the following analysis will compare some combination of SC, ST and Muslims with the rest. In order to be comparable with the between-groups estimate from the HDPI-IHDS panel, I restrict sample to rural India, male head of the household between the ages 16 and 65 who is not currently enrolled and reports his educational information. Table 8 presents summary statistics for the resulting sample. For each group in each round, average age is between

[11] I use state level poverty lines to convert consumption expenditure in real terms. In my analysis I express all consumption expenditure values in 1983 rural Maharashtra prices. For 1983 and 2004-05, I use state level poverty lines derived from the Lakdawala approach. For 2011-12, I estimate the Lakdawala lines by adjusting the 2004-05 Lakdawala lines using the index implicit in the official Tendulkar lines for 2004–2005 and 2011–2012.

[12] Using 84 months of panel data from Townsend Thai project, Chiappori et. al., (2014) find evidence of consumption smoothing in the face of idiosyncratic household shocks, suggesting that consumption can be a reasonable measure of long term income. Chiappori et. al., (2013) find similar results using a less-demanding portfolio choice model.

40 and 44 years, and the proportion of married observations is between 0.91 and 0.95. SC/STs tend to have a relatively smaller household size, and there is a clear over time decline in household size for both the categories.

Figure 3 presents density and cumulative distribution of log monthly per capita (real) consumption expenditure (LMPCE) for each group by survey rounds. A feature common to all the densities is that in each round the density of the Non-SC/ST households is relatively more skewed to the left. And a common feature to all the CDFs is that the Non-SC/ST households have a relatively higher mass towards the higher quantiles in the LMPCE distribution. I next calculate the between-groups estimator for each of the two pairs of survey rounds. Table 9 presents the between-groups estimators for Survey rounds 38(1983)-68(2011-12) and Survey rounds 38(1983)-61(2004-05). From panel(A), which compares SC/STs with the rest, the between-groups estimator from the survey rounds 38 and 68 is found to be 0.75. Interestingly, this estimator is almost similar to that of 0.74 obtained from the HDPI-IHDS panel. Not surprisingly, I fail to reject the null that these two estimates are significantly different. Although I prefer a comparison from the survey rounds 38 and 68 which have an interim gap of 30 year, I present estimates from survey rounds 38 and 61 as well to check whether it gives completely different estimates. It is reassuring that between-groups estimator from these two rounds is found to be 0.81, which is not far off from the earlier estimate. Panel(B) presents results when Muslims are included with the SC/STs. Interestingly, Wald from each pair of rounds is slightly higher but very similar to the results found in panel(A).

Note that the foregoing results were based on a sample of male household heads in the age group of 16 to 65. Are these results sensitive to the choice of age group? Table 11 reveals that between-groups estimator varies between 0.73 and 0.77 when age criteria is changed from 16-65 to 25-6, 35-65, or 45-65. Thus, the findings of the low rate of convergence is robust to the choice of age group. Another appealing outcome of convergence is years of education. Table 12 presents between-groups estimators measured in terms of years of education. These estimators, for various combinations of survey rounds and groups, are slightly above one. Although the time scale used for education is completely different from the monetary scale applied to income and consumption, it nevertheless suggests that there is either slow or no convergence between groups in terms of years of education .

Although 70% of India's population continues to live in rural areas, it will be interesting to learn about between-groups estimators from a sample that covers both rural and urban areas. The HDPI-IHDS panel data does not allow such a comparison, but information from the NSS samples can be used to make some conjecture for the entire population. Table 10 presents between-groups estimators for both rural and urban samples from Survey rounds 38(1983)-68(2011-12) and Survey rounds 38(1983)-61(2004-05). Surprisingly, between-groups estimator is relative high for each pair of rounds and for each combination of groups. Does it mean that convergence between groups for rural and urban sample is slower than that for only the rural sample? Given the absence corresponding estimates from the HDPI-IHDS panel, nothing decisive can be said yet, even though cross-sectional evidence is in affirmation. For the rural India, however, evidence is clear from the aforementioned findings that between-groups convergence is very slow.

In a recent study, Hnatkovska, Lahiri and Paul (2012) (henceforth HLP) investigate patterns in group-wise mobility using multiple rounds of the NSS. Besides numerous similarities in sample selection criteria between current study and HLP, sample used in current study varies from working sample in HLP by not requiring 3-digit occupational code and full time work status. Furthermore, current study incorporates evidence from the NSS round 68 (2011-12), allowing 30 years long interim gap between 1983 and 2011-12, which was not available to HLP. I obtain similar consumption results as those in HLP. Figure 4 reproduces Figure 7(B) in HLP. In each panel of Figure 4, the two dotted lines plot the group ratio in log-real-consumption for various NSS rounds. By comparing round 61 with round 38 for both rural and urban sample, Panel(C) comes closest to the underlying sample in figure 7(B) from HLP. It is comforting that dotted lines in panel(C) are almost exactly the same as those in figure 7(B) of HLP, indicating that sample in current study is not significantly different from that in HLP. A common feature in all the panels is that the two dotted lines show similar pattern. What is most remarkable about this figure is the solid line overlaying the two dotted lines — this overlaying solid line is similar in spirit to the Wald estimator: the ratio of group differences in log-real-consumption between two rounds. Since this solid line lies above 0.7 at almost all the percentiles in each panel, it suggests that the Wald from underlying data is also above 0.7. Strictly speaking, the solid line is not quite the Wald, but findings from Table 9 and Table 10 provide the definitive evidence that Wald from HLP is indeed above 0.7.

5 Robustness Analysis

In this section, I present results that address a few valid concerns. I check whether the HDPI-IHDS panel is representative of rural India, whether out-migration of sons causes significant bias, and whether there is any evidence of nonlinearity.

When the IHDS was first conceived in the early 2000s, a proportion of rural households from the HDPI survey (1993-1994) were included to form a panel. A refresher rural sample and an urban sample was then added to make the IHDS representative at the all-India level. I check whether the HDPI-IHDS panel is representative of rural India, as it was originally intended to be, by comparing panel households that were re-interviewed in 2005 (Panel HH) with two other samples: 1) rural households in 1994 that were subsequently not re-interviewed (Non-Panel HH [1994]) and 2) refresher rural households that were interviewed for the first time in 2005 (Refresher HH [2005]). Results from both these comparisons are presented in Table 13. Comparing column 1 and column 2 reveals that the panel HH and the non-Panel HH in 1994 were similar in size, household income, composition, land ownership, and cultivation practices. A comparison between panel households (column 3) and the refresher rural households (column 4) in 2005 indicates that they were similar in sizes, numbers of adults and married couples, poverty rates, and household incomes. In summary, comparison suggests that the selection of panel households went as intended, and we can trust the panel sample to be representative of rural India.

As is true for any other panel dataset, it is possible to lose track of certain individuals or even the entire household between two time periods. Such missing observations may give rise to the concern that results are biased. An advantage of the IHDS dataset is that it provides some relevant information, such as educational attainment, about the non-resident members of the household. Information on education in the absence of income is especially relevant given the evidence that income and educational attainment are closely related (Hertz et. al., 2007). In order to understand the direction of bias, I present both relative and absolute measures of intergenerational educational elasticity (IGEE).[13] Results pertaining to both these measures of IGEE are presented in Table 14. The relative IGEE for the sample of 1,774 father-son pairs used in the main analysis is 0.465, and the absolute IGEE is 0.389. And the relative IGEE when 143 non-resident sons are added to the original sample is 0.461, and the absolute IGEE is 0.390. Interestingly, no discernible difference is evident between these two sets of results, suggesting no systematic bias in the IGEE estimates from missing information on non-resident sons. Given the evidence on positive correlation between educational status and income, it seems that no systematic bias is caused in the IGE due to missing non-resident sons.[14]

[13] Relative measure regresses sons' years of schooling against fathers' years of schooling. On the other hand, the absolute measure of IGEE regresses sons' years of schooling against fathers' years of schooling after scaling each by their corresponding standard deviations.

[14] Similar results are obtained when Table 14 is reproduced for the time period 1994-2005.

Finally, it is worth asking if the relationship between fathers' and sons' status is at all linear. If the relationship were not to be linear, then it makes sense to use a flexible non-linear specifications and check how results are effected. Figure 5 plots the mean of sons' incomes for each percentile of the parents' income. An easy way to account for non-linearity is to estimate a flexible specification that includes quadratic or even cubic terms. Results from such flexible estimation are presented in Table 15. Unlike coefficient on cubic term, coefficient on quadratic term is significant in both the specifications, which hints towards prevalence of non-linearity. Table 16 estimates the IGE at mean and four quintiles for both specifications. It is clear that IGE varies at different quintiles. At the first quintile, for example, the value is 0.087, but it increases to 0.365 at the fourth quintile.

6 Summary and Discussion

Reliable estimates of income mobility using panel data are hard to find for developing countries. This paper uses a unique panel dataset to estimate the IGE for rural India during 1994-2012. The contribution of this paper is threefold: 1) it provides reliable estimates of IGE for rural India; 2) it finds mobility estimates various different socio-economic groups; and 3) it adopts strategies that can be used for similar research in other developing countries.

Depending on the choice of income scaling parameters, the OLS IGE estimates lie between 0.28 and 0.37. What do these estimates really mean? Following Solon (1992), under the assumption that economic status of father and sons is normally distributed, the IGE estimates can be interpreted as the probability of mobility across quintiles. For the IGE of 0.4 and a father with 5th percentile in his generation, the probability of his son lying in the bottom quintile is 0.42, moving above median is 0.24 and moving to the top quintile is 0.05. It will also be instructive to compare the IGE for rural India with evidence from other countries. The OLS IGE estimates seem to be a little lower than the estimate of 0.59 for Chile (Celhay et. al., 2011), but they are certainly higher than the estimate of 0.26 that was found for Malaysia about 20 years ago (Lilland and Kilburn, 1995). Jantti et. al. (2006) find the IGE estimates of 0.306 and 0.517 for the UK and the US respectively. Among developed countries, the OLS IGE estimates for rural India seem closer to those from the UK and the US, which are considered to be among the least mobile developed countries.

This study also finds group wise measures of mobility. In the past, many researchers have compared poverty rates between various socio-economic groups (Panagariya and More, 2013).[15]

[15] A popular measure of poverty rate comparisons is the Head Count Ratio- a measure of the proportion of population that lives below

However, there are not many studies that investigate intergenerational income mobility patterns for these groups. In comparison to the poverty rate analysis, mobility measures inform us about the access to opportunities among various groups, and whether any particular group's aspirations are as limitless as that of any other group. There is no disputing of the fact that poverty rate analysis is invaluable in the face of widespread deprivation, but this in no way can obviate the need for policies that go beyond poverty eradication and ensure equitable access to opportunities for everyone. Moreover, pulling disadvantaged groups out of the poverty trap may have been the short-term objective of the constitutional writers, but their choices of education, employment, and politics as spheres for affirmative action clearly imply that the writers were primarily interested in improving their accumulation of human capital. A study of intergenerational income mobility among the groups can be one way to ascertain whether and to what extent these coveted objectives have been achieved. An analysis of the IGE among groups will also serve well by informing the current debate about affirmative action in India. Given that affirmative action takes away as much as 50% of the total available spots in public sector jobs and seats in educational institutions, access to opportunities by all the groups is tied to the future of affirmative action.[16]

Group-wise analysis suggests that the SC/STs have relatively higher within-group mobility. Improvement in within-group mobility for SCs and STs is a positive outcome, and it suggests that for income distribution within these groups, children from lower income families have a better chance of moving up the economic ladder than they did before. A possible story behind this improvement could be that implementation of affirmative policies has improved, and benefits from such policies are no longer limited to the affluent households. Results from the between-groups comparisons are rather dismal. The between-groups estimator during 1994-2012 is estimated as 0.74, which is corroborated by the evidence from the NSS data. Though income in absolute terms and economic well-being may have increased for most groups, convergence across groups in terms of log real income (or log real consumption) has been very slow. How slow is the slow convergence when the between-groups estimator is 0.74? Suppose that, ceteris paribus, the speed of convergence stays the same, then it will take about seven generations to reduce the gap between groups to 10% of the gap in 1994. Within-group and between-groups results together suggest that even though within-group mobility for certain disadvantaged groups may have improved, improvement in their relative access

the 'poverty line' as set by the government.

[16] In central-government funded higher education institutions, 22.5% of available seats are reserved for Scheduled Caste (SC) and Scheduled Tribe (ST) students (7.5% for STs, 15% for SCs). This reservation percentage has been raised to 49.5% by including an additional 27% reservation for OBCs.

to opportunities during previous generation has been very slow.

An alarmingly low rate of convergence between groups signals a need to critically appraise existing policies. The findings emphasize the need for expanding the usual discussion about poverty rates to a broader discourse on the access to opportunities and mobility among various groups. Useful as they are, poverty rates should not become the sole focus of developmental policies. In other words, bringing these disadvantaged sections barely above poverty line should not be the end goal of these policies. Instead, the focus must be on creating an equitable society — a society where no one is restricted from improving their economic status just because they were born in a particular caste or a religious group.

References

[1] Azam, Mehtabul. (2013). "Intergenerational Occupational Mobility in India." IZA Discussion Papers 7608, Institute for the Study of Labor (IZA).

[2] Azam, Mehtabul. (2015). "Intergenerational Educational Persistence among Daughters: Evidence from India." IZA Discussion Papers 9518, Institute for the Study of Labor (IZA).

[3] Borjas, George J. (1992). "Ethnic Capital and Intergenerational Mobility." Quarterly Journal of Economics 107 , pp. 123-150.

[4] Celhay, Pablo, Claudia Sanhueza, Jose R. Zubizarreta. (2011). "Intergenerational Mobility of Income and Schooling: Chile 1996-2006." Revista de Análisis Económico – Economic Analysis Review, [S.l.], v. 25, n. 2, p. 43-64.

[5] Chand, Ramesh and Raju.S. S. (2009). "Instability in Indian Agriculture during Different Phases of Technology and Policy." Indian Journal of Agricultural Economics. 64(2): 187-207.

[6] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. (2014). "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." Quarterly Journal of Economics 129, pp. 2633-2679.

[7] Chiappori, Pierre-Andre, Krislert Samphantharak, Sam Schulhofer-Wohl, and Robert Townsend. (2013). Portfolio choices and risk preferences in village economies. Federal Reserve Bank of Minneapolis Working Papers No. 706.

[8] Chiappori, Pierre-Andre, Krislert Samphantharak, Sam Schulhofer-Wohl, and Robert M. Townsend. (2014). Heterogeneity and Risk Sharing in Village Economies. Quantitative Economics 5 (1):1-27.

[9] Choi, Ji Eun, and Ki Seok Hong. (2011). "An Analysis on Intergenerational Income Mobility in Korea: Father-Son Correlations in Labor Earnings." Korean Social Security Studies no. 27 (3):143-163.

[10] Chu, Yu-Wei Luke and Ming-Jen Lin. (2016). "Economic Development and Intergenerational Earnings Mobility: Evidence from Taiwan." Last accessed on 11/13/2016 from SSRN: https://ssrn.com/abstract=2835881

[11] Clark, Gregory. (2014). “The Son Also Rises: Surnames and the History of Social Mobility.” Princeton: Princeton University Press.

[12] Daniel, Aharon. (2010). “Caste System in Modern India.” Adaniel’s Info Site. Web.

[13] Deaton, Angus and Salman Zaidi. (2002). “Guidelines for Constructing Consumption Aggregates for Welfare Analysis.” Living Standards Measurement Study Working Paper 135. Washington DC: World Bank.

[14] Desai, Sonalde, Reeve Vanneman, and National Council of Applied Economic Research, New Delhi. (2005). “India Human Development Survey (IHDS).” ICPSR22626-v8. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-06-29. http://doi.org/10.3886/ICPSR22626.v8

[15] Desai, Sonalde, Reeve Vanneman and National Council of Applied Economic Research, New Delhi. (2011-12). “India Human Development Survey-II (IHDS-II).” ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-07-31. http://doi.org/10.3886/ICPSR36151.v2

[16] Desai, Sonalde, Amaresh Dubey, B.L. Joshi, Mitali Sen, Shariff, Abusaleh, and Reeve Vanneman. (2007). “India Human Development Survey (IHDS): User’s Guide Release 3.” ICPSR22626-v2. University of Maryland and National Council of Applied Economic Research, New Delhi [producers]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-05-01.

[17] Dunn, C. (2007). “The intergenerational transmission of lifetime earnings: evidence from Brazil.” The B.E. Journal of Economic Analysis and Policy 7(2), doi:10.2202/1935-1682.1782.

[18] Gong, C. H., A. Leigh, and X. Meng. (2012). “Intergenerational Income Mobility in Urban China.” Review ofIncome and Wealth, 58, 481–503.

[19] Haider, Steven, and Gary R. Solon. (2006). “Life-cycle variation in the Association Between Current and Lifetime Earnings.” The American Economic Review, 96 , 1308-1320.

[20] Hnatkovska, V., Lahiri, A. and Paul S.B. (2013). “Breaking the Caste Barrier: Intergenerational Mobility in India.” Journal of Human Resources, 48(2), 435-73.

[21] Hnatkovska, Viktoria, Amartya Lahiri, Sourabh Paul. (2012). “Castes and Labor Mobility.” American Economic Journal: Applied Economics 4 (2): 274–307.

[22] Hertz, Tom. (2005). “A group-specific measure of intergenerational persistence.” Economics Letters, Elsevier, vol. 100(3), pages 415-417, September.

[23] International Monetary Fund. (2016). “Regional Economic Outlook. Asia and Pacific. Building on Asia’s Strength during Turbulent Times.” https://www.imf.org/external/pubs/ft/reo/2016/apd/eng/pdf/areo0516.pdf (accessed on 09/25/2016).

[24] Jalan, J., and Murgai, R. (2008). “Intergenerational mobility in education in India.” India: World Bank.

[25] Jäntti, Markus, Bernt Bratsberg, Knut Røed, Oddbjørn Raaum, Robin Naylor, Österbacka, Eva, Björklund Anders and Tor Eriksson. (2006). “American exceptionalism in a new Light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States.” Discussion paper no. 1938 (Institute for the Study of Labor (IZA), Bonn).

[26] Kan, Kamhon, I Li, and Ruei-Hua Wang. (2015). “Intergenerational Income Mobility in Taiwan: Evidence from TS2SLS and Structural Quantile Regression.” The BE Journal of Economic Analysis and Policy no. 15 (1):257-284.

[27] Lillard, Lee A. and M. Rebecca Kilburn. (1995). "Intergenerational Earnings Links: Sons and Daughters." Unpublished.

[28] Maitra, P., and Sharma, A. (2010). “Parents and children: Education across generations in India.” CAM, Working paper. Monash University.

[29] National Council on Applied Economic Research. (1994). “Human Development Profile of India (HDPI)”. New Delhi.

[30] National Research Council. (1995). “Measuring poverty: a new approach.” Washington, DC. National Academy Press.

[31] National Sample Survey Office. (2011). “Key Indicators of Employment and Unemployment in India 2009-10, 66th Round.”

[32] Ng, Irene. (2007). “Intergenerational income mobility in Singapore.” The BE Journal of Economic Analysis and Policy no. 7(2).

[33] Ng, Irene YH, Xiaoyi Shen, and Kong Weng Ho. (2009). “Intergenerational earnings mobility in Singapore and the United States.” Journal of Asian Economics no. 20 (2):110-119.

[34] Nuñez, J and L. Miranda. (2011). “Intergenerational income and educational mobility in urban Chile.” Estudios de Economía, 38 (1) (2011), pp. 195–221

[35] Panagariya, Arvind, and Vishal More. (2013). “Poverty by Social, Religious and Economic Groups in India and Its Largest States 1993-94 to 2011-12.” Program on Indian Economic Policies, Columbia University, New York City.

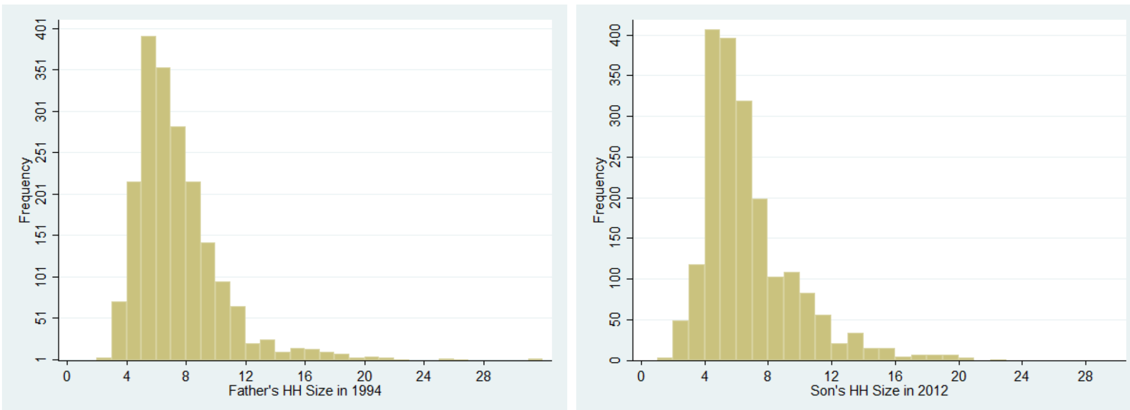
[36] Solon, Gary R. (1992). “Intergenerational Income Mobility in the United States.” American Economic Review, June 1992, 82(3), pp. 393-408.

[37] Solon, Gary R. (1999). “Intergenerational Mobility in the Labor Market.” in Handbook of Labor Economics, Volume 3A. Orley Ashenfelter and David Card, eds. Amsterdam: North-Holland, pp. 1761–800.

[38] Solon, Gary R. (2015). “What Do We Know So Far About Multi-Generational Mobility?” NBER Working Paper No. 21053, Cambridge, MA, 2015.

[39] Sun, Fengye, and Atsuko Ueda. (2015). “Intergenerational earnings mobility in Taiwan.” Economics Bulletin no. 35 (1):187-197.

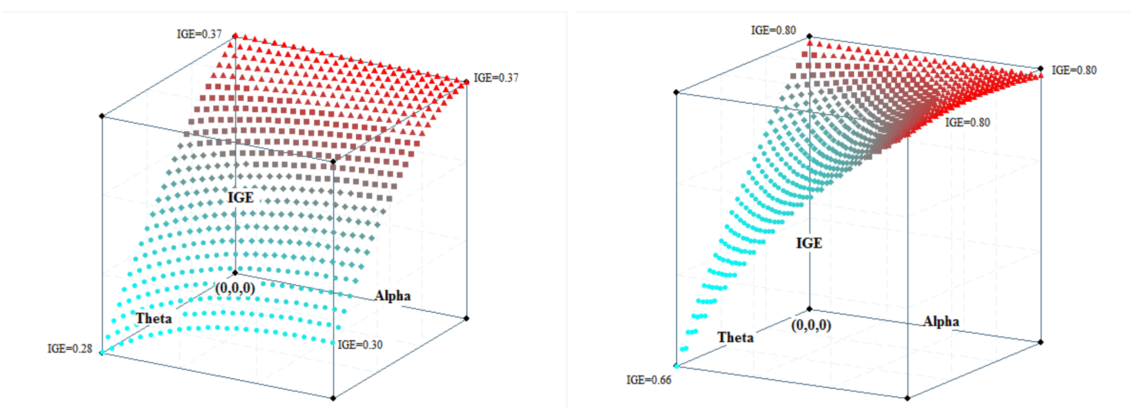
[40] Ueda, Atsuko. (2013). "Intergenerational mobility of earnings in South Korea." Journal of Asian Economics no. 27:33-41.



(a) HH size in father's generation

(b) HH size in son's generation

Figure 1: Density and CDF: Monthly Per Capita Consumption Expenditure

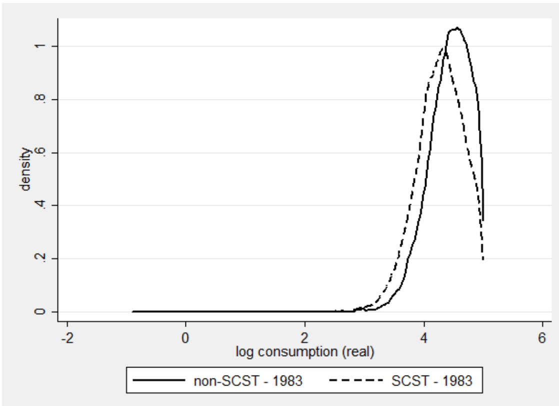


(a) IGE-OLS for all α and θ

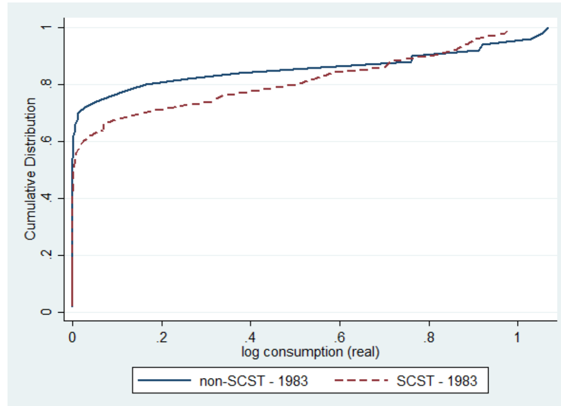
(b) IGE-IV for all α and θ

Figure 2: IGE for all α and θ

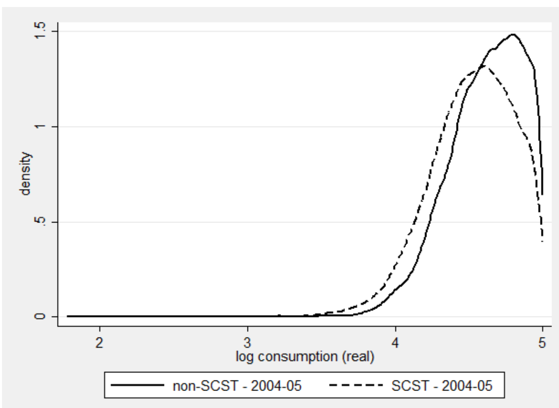
Note: Alpha (between 0 and 1) is the cost of a child relative to that of an adult. Theta is a measure of economies of scale when subtracted from 1.



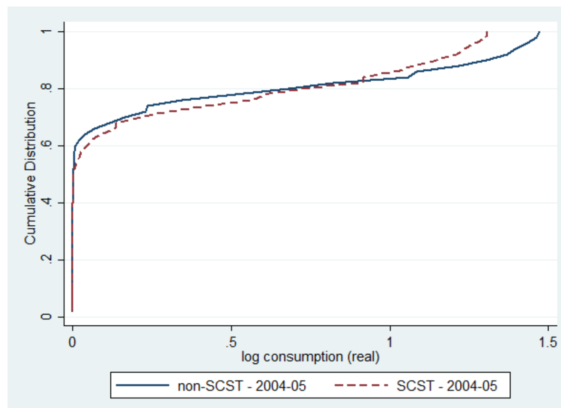
(a) LMPCE density 1983



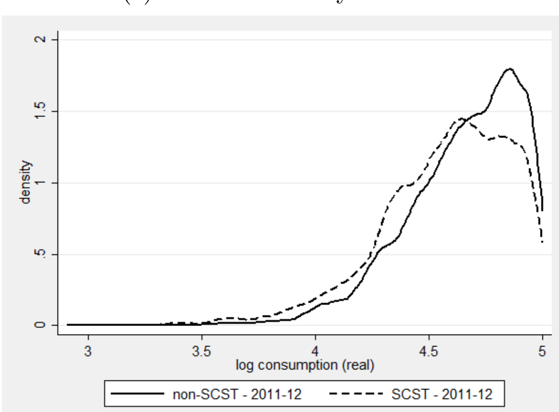
(b) LMPCE cdf 1983



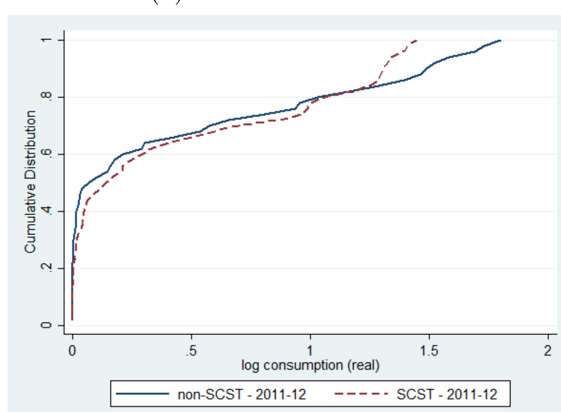
(c) LMPCE density 2004-05



(d) LMPCE cdf 2004-05

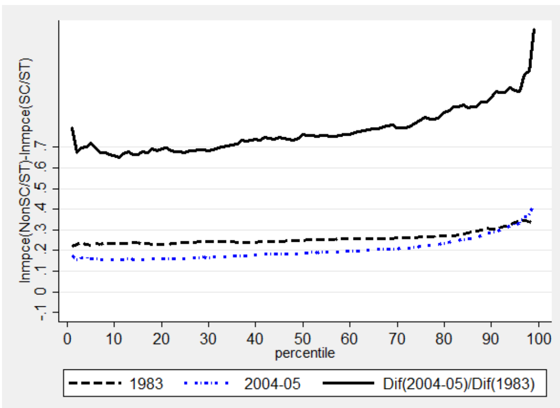


(e) MPCE density 2011-12

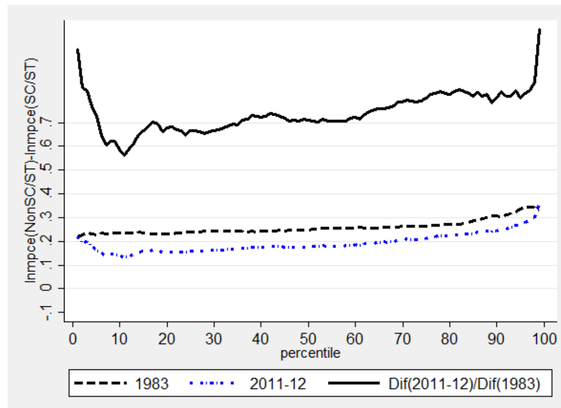


(f) MPCE cdf 2011-12

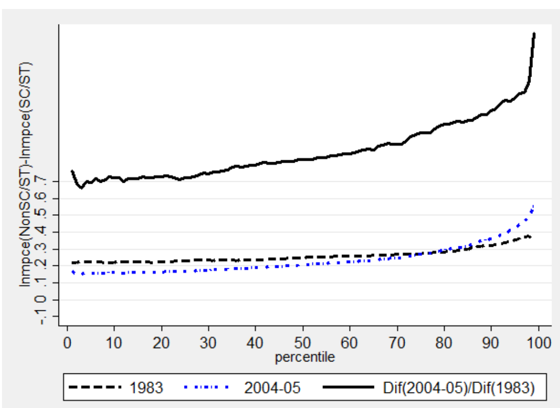
Figure 3: Density and CDF: Monthly Per Capita Consumption Expenditure



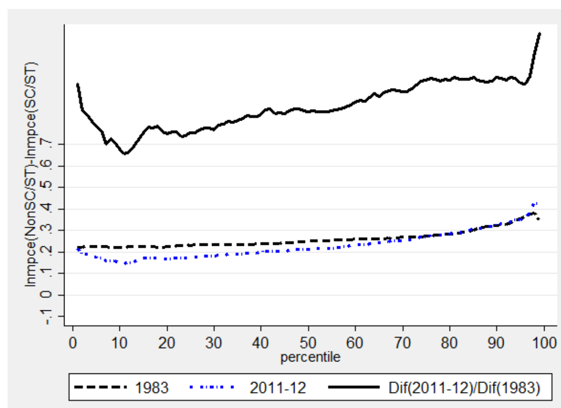
(a) Group Ratio - Rural



(b) Group Ratio - Rural



(c) Group Ratio - Rural+Urban



(d) Group Ratio - Rural+Urban

Figure 4: Group Ratio of Log-Real-Consumption

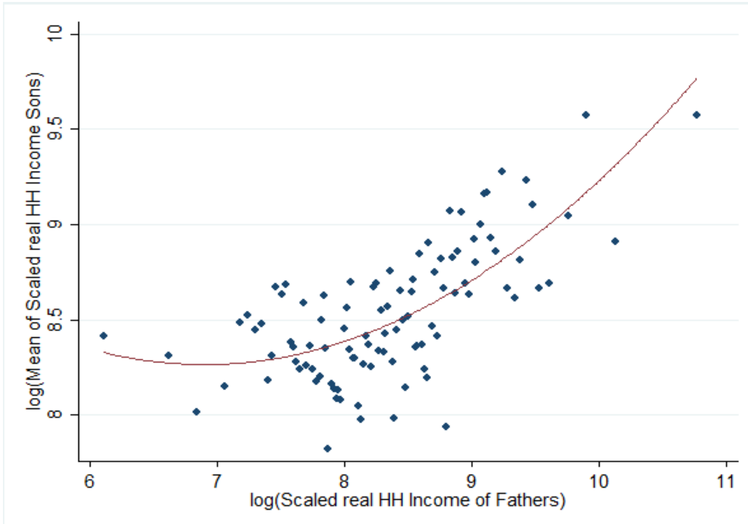


Figure 5: Quadratic Plot

Note: This figure plots average income for sons corresponding to each percentile of father's income. X-Axis plots logarithm of scaled father's HH income in 1994. Y-Axis plots the logarithm of mean of scaled HH income for son in 2012.

	(1)	(2)
	Father (1994)	Son (2012)
Age	50.48	36.47
	(7.97)	(2.77)
Literacy	0.53	0.80
	(0.50)	(0.40)
HH Size	7.00	6.15
	(3.03)	(2.82)
Total HH Income (1,000 Rupees)	34.7	122.8
	(37.0)	(189.8)
log(Total HH Income)	10.13	11.17
	(0.77)	(1.01)
Observations	1,776	1,776

Table 1: Descriptive Statistics

Note: Standard Deviations are reported in parentheses.

	(1)	(2)
Boundary	**OLS**	**IV**
	(α, θ, IGE)	(α, θ, IGE)
Boundary I	(0,1,0.277)	(0,1,0.664)
	(0.033)	(0.103)
Boundary II	(1,1,0.298)	(1,1,0.804)
	(0.033)	(0.108)
Boundary III	(1,0,0.367)	(1,0,0.805)
	(0.036)	(0.102)
Boundary IV	(0,0,0.367)	(0,0,0.805)
	(0.036)	(0.102)
Observations	1,776	1,776

Table 2: Robustness: IGE for All possible values of α and θ.

Note: Robust standard errors for the IGE are reported in parentheses.

	(1)	(2)	(3)	(4)
	$\textbf{Income}_{Son2012}$	$\textbf{Income}_{Son2012}$	$\textbf{Income}_{Son2012}$	$\textbf{Income}_{Son2012}$
	(OLS)	(OLS)	(OLS)	(OLS)
$\textbf{Income}_{Father1994}$	0.30	0.30	0.31	0.31
	(0.03)	(0.03)	(0.03)	(0.03)
Drop Observations if:				
age error over +/-5	Yes	No	Yes	No
age error over +/-10	No	yes	No	Yes
work hours last year=0	Yes	Yes	No	No
Observations	1,776	2,043	1,855	2,137

Table 3: OLS Results: Sample Selection

Note: Robust standard errors are reported in parentheses. Son's income is the logarithm of scaled real household income for son in 2012. Father's income is the logarithm of scaled real household income for father in 1994. The control variables are 'age of father', 'age of son', and their squares. α=0.3 and θ=0.9 are used for scaling HH income.

	(1)	(2)	(3)	(4)
	$\textbf{Income}_{Son2012}$	$\textbf{Income}_{Son2012}$	$\textbf{Income}_{Son2012}$	$\textbf{Income}_{Son2012}$
	(OLS)	(OLS)	(OLS)	(OLS)
$\textbf{Income}_{Father1994}$	0.30	0.29	0.33	0.31
	(0.03)	(0.03)	(0.04)	(0.03)
Age Group:				
16 - 22	Yes	No	No	No
14 - 22	No	Yes	No	No
16 - 20	No	No	Yes	No
14 - 20	No	No	No	Yes
Observations	1,776	2,394	1,332	1,950

Table 4: OLS Results: Age Groups

Note: Robust standard errors are reported in parentheses. Son's income is the logarithm of scaled real household income for son in 2012. Father's income is the logarithm of scaled real household income for father in 1994. The control variables are 'age of father', 'age of son', and their squares. α=0.3 and θ=0.9 are used for scaling HH income.

	(1)	(2)	(3)	(4)
	Income$_{Son2012}$	**Income**$_{Son2012}$	**Income**$_{Son2012}$	**Income**$_{Son2012}$
	(2SLS)	(2SLS)	(2SLS)	(2SLS)
Income$_{Father1994}$	0.733	0.952	0.590	-0.023
	(0.103)	(0.124)	(0.224)	(0.197)
First Stage Results:				
	Income$_{Father1994}$	**Income**$_{Father1994}$	**Income**$_{Father1994}$	**Income**$_{Father1994}$
Father's Education	0.047	0.052		
	(0.005)	(0.005)		
Own Radio	0.128		0.235	
	(0.042)		(0.042)	
Own Bike	0.177			0.275
	(0.047)			(0.046)
Observations	1,776	1,776	1,776	1,776

Table 5: 2SLS Results

Note: Robust standard errors are reported in parentheses. Father's education is measured in years of schooling as reported in 1994; own radio is a dummy variable that takes a value of 1 is the HH owns a radio or transistor in 1994; Own bike is also a dummy variable that takes a value of 1 if the HH owns a bicycle in 1994. F-Statistic from first stage in the first column is 23.18. Father's income is the logarithm of scaled household income for father in 1994. The control variables in both the first stage and the second stage are 'age of father', 'age of son', and their squares. α=0.3 and θ=0.9 are used for scaling HH income.

Group	**Share**	**Father's Income**	**Son's Income**	$\boldsymbol{\beta_g}$	$\boldsymbol{\delta}$
A: SC/ST					
SC/ST	0.30	8.162	8.341	0.128	0.735
		(0.029)	(0.034)	(0.052)	(0.001)
Rest	0.70	8.454	8.556	0.329	
		(0.022)	(0.026)	(0.040)	
B: SC/ST/Muslims					
SC/ST/Muslims	0.37	8.184	8.361	0.181	0.711
		(0.024)	(0.030)	(0.052)	(0.121)
Rest	0.63	8.486	8.577	0.320	
		(0.023)	(0.028)	(0.042)	
C: SC/ST/Muslims/OBC					
SC/ST/Muslims/OBC	0.74	8.270	8.429	0.269	0.654
		(0.020)	(0.023)	(0.040)	(0.142)
Rest	0.26	8.672	8.692	0.304	
		(0.035)	(0.045)	(0.067)	
Pooled	**1.00**	**8.375**	**8.498**	**0.298**	
		(0.012)	(0.0.021)	(0.033)	

Table 6: Results from Two-groups analysis

Note: Standard Errors are reported in parentheses. The between-groups estimator is titled as $\boldsymbol{\delta}$. Standard Errors of the between-groups estimator is calculated from Delta method. SC stands for Scheduled Castes, ST stands for Scheduled Tribes, and OBC stands for Other Backward Classes. α=0.3 and θ=0.9 are used for scaling HH income.

Group	Share	Father's Income	Son's Income	$\boldsymbol{\beta_g}$	$\boldsymbol{\delta}$
Forward Castes	0.23	8.619	8.733	0.316	0.909
		(0.793)	(0.990)	(0.070)	
OBCs	0.36	8.345	8.508	0.320	
		(0.793)	(0.860)	(0.056)	
SCs	0.22	8.132	8.470	0.139	
		(0.642)	(0.710)	(0.058)	
STs	0.08	8.146	8.029	0.070	
		(0.722)	(0.939)	(0.114)	
Muslims	0.08	8.262	8.458	0.435	
		(0.544)	(0.849)	(0.154)	
Others	0.03	8.651	9.286	0.259	
		(0.798)	(0.693)	(0.135)	
Pooled	**1.00**	**8.349**	**8.534**	**0.298**	
		(0.761)	(0.889)	(0.033)	

Table 7: Results from six groups analysis

Note: Standard Deviations are reported in parentheses for the columns titled 'Father's Income' and 'Son's Income'. Robust Standard Errors are reported in parentheses for the column titled $\boldsymbol{\beta_g}$. The between-group estimator is titled as $\boldsymbol{\delta}$. SCs stands for Scheduled Castes, STs stands for Scheduled Tribes, and OBCs stands for Other Backward Classes. Father's Income is the logarithm of scaled household income for father in 1994, and Son's Income is the logarithm of scaled income for son in 2012. α=0.3 and θ=0.9 are used for scaling HH income.

NSS Round Year	Age	Married	HH Size	Observations
A: SC/ST				
1983	40.63	0.91	6.16	20,899
	(0.11)	(0.002)	(0.03)	
2004-05	41.53	0.94	5.83	22,310
	(0.11)	(0.002)	(0.03)	
2011-12	42.31	0.95	5.29	16,770
	(0.17)	(0.003)	(0.04)	
B: Non-SC/ST				
1983	42.21	0.91	6.78	43,679
	(0.07)	(0.001)	(0.02)	
2004-05	43.27	0.95	6.12	42,315
	(0.08)	(0.001)	(0.02)	
2011-12	43.89	0.95	5.59	32,238
	(0.12)	(0.002)	(0.02)	

Table 8: Summary Statistics: National Sample Survey

Note: Standard Errors are reported in parentheses.

Group	Share	$LMPCE_{1983}$	$LMPCE_{2005}$	$LMPCE_{2012}$	$\boldsymbol{\delta}_{1983-2005}$	$\boldsymbol{\delta}_{1983-2012}$
A: SC/ST						
SC/ST	0.34	4.383	4.642	4.785	0.797	0.752
		(0.003)	(0.003)	(0.003)	(0.021)	(0.022)
Rest	0.66	4.645	4.853	4.983		
		(0.003)	(0.002)	(0.002)		
B: SC/ST/Muslims						
SC/ST/Muslim	0.44	4.410	4.659	4.809	0.858	0.762
		(0.003)	(0.002)	(0.003)	(0.018)	(0.021)
Rest	0.56	4.663	4.876	5.002		
		(0.003)	(0.002)	(0.003)		
Pooled	**1.00**	**4.565**	**4.781**	**4.917**		
		(0.002)	(0.002)	(0.002)		

Table 9: NSS Rural Sample: Results from Two-groups LMPCE analysis

Note: Standard Errors are reported in parentheses. Standard Errors of the between-groups estimator is calculated from Delta method. The between-groups estimator is titled as $\boldsymbol{\delta}$. SCs stands for Scheduled Castes and STs stands for Scheduled Tribes.

Group	Share	Rural	$LMPCE_{1983}$	$LMPCE_{2005}$	$LMPCE_{2012}$	$\boldsymbol{\delta}_{1983-2005}$	$\boldsymbol{\delta}_{1983-2012}$
A: SC/ST							
SC/ST	0.29	0.74	4.394	4.644	4.809	0.909	0.900
			(0.003)	(0.002)	(0.003)	(0.017)	(0.019)
Rest	0.71	0.59	4.653	4.879	5.042		
			(0.002)	(0.002)	(0.002)		
B: SC/ST/Muslims							
SC/ST/Muslim	0.4	0.69	4.418	4.656	4.830	1.000	0.951
			(0.003)	(0.002)	(0.002)	(0.017)	(0.017)
Rest	0.6	0.59	4.683	4.922	5.081		
			(0.002)	(0.002)	(0.002)		
Pooled	**1.00**		**4.585**	**4.811**	**4.976**		
			(0.002)	(0.001)	(0.002)		

Table 10: NSS Rural + Urban Sample: Results from Two-groups LMPCE analysis

Note: Standard Errors are reported in parentheses. Standard Errors of the between-groups estimator is calculated from Delta method. The between-groups estimator is titled as $\boldsymbol{\delta}$. SCs stands for Scheduled Castes and STs stands for Scheduled Tribes.

	(1)	(2)	(3)	(4)
	$\boldsymbol{\delta}_{1983-2012}$	$\boldsymbol{\delta}_{1983-2012}$	$\boldsymbol{\delta}_{1983-2012}$	$\boldsymbol{\delta}_{1983-2012}$
SC/ST vs Rest	0.752	0.749	0.766	0.727
	(0.018)	(0.022)	(0.025)	(0.033)
SC/ST/Muslims vs Rest	0.762	0.766	0.773	0.713
	(0.021)	(0.021)	(0.023)	(0.029)
Age Group:				
16 - 65	Yes	No	No	No
25 - 65	No	Yes	No	No
35 - 65	No	No	Yes	No
45 - 65	No	No	No	Yes
Observations	178,211	172,493	136,717	83,110

Table 11: NSS Rural Sample: Age Groups [Consumption]

Note: Standard Errors are reported in parentheses. Standard Errors of the between-groups estimator is calculated from Delta method. The between-groups estimator is titled as $\boldsymbol{\delta}$. The specified age group is applied in both Round 61 and Round 68. SCs stands for Scheduled Castes and STs stands for Scheduled Tribes.

Group	**Share**	EduYears1983	EduYears2005	EduYears2012	$\boldsymbol{\delta}_{1983-2005}$	$\boldsymbol{\delta}_{1983-2012}$
A: SC/ST						
SC/ST	0.34	1.271	2.680	3.513	1.147	1.104
		(0.023)	(0.036)	(0.056)	(0.042)	(0.044)
Rest	0.66	2.701	4.320	5.092		
		(0.022)	(0.030)	(0.048)		
B: SC/ST/Muslims						
SC/ST/Muslims	0.44	1.400	2.695	3.536	1.340	1.295
		(0.020)	(0.031)	(0.050)	(0.040)	(0.040)
Rest	0.56	2.823	4.605	5.381		
		(0.024)	(0.034)	(0.053)		
Pooled	**1.00**	**2.255**	**3.777**	**4.579**		
		(0.017)	(0.024)	(0.037)		

Table 12: NSS Rural Sample: Education

Note: Standard Errors are reported in parentheses. Standard Errors of the between-groups estimator is calculated from Delta method. The between-groups estimator is titled as $\boldsymbol{\delta}$. SCs stands for Scheduled Castes and STs stands for Scheduled Tribes.

	(1)	(2)	(3)	(4)
	$Panel\ HH_{94}$	$Nonpanel\ HH_{94}$	$Panel\ HH_{05}$	$Refresher\ HH_{05}$
Household (HH) Size	6.03	5.77	5.58	5.16
	(2.86)	(2.79)	(2.81)	(2.43)
Number of couples in the HH?	1.33	1.27	1.33	1.20
	(0.79)	(0.76)	(0.80)	(0.68)
Total HH Income[1,000 Rupees]	27.3	27.0	41.7	40.6
	(35.0)	(36.3)	(63.1)	(75.5)
Log HH Income	9.82	9.79	10.17	10.11
	(0.86)	(0.89)	(0.99)	(1.03)
Log HH Income (Scaled)	8.56	8.58	8.96	8.95
	(0.77)	(0.80)	(0.89)	(0.96)
22+ in HH			2.93	2.72
			(1.49)	(1.33)
Below poverty line			0.22	0.18
			(0.44)	(0.41)
working adults(15-59)	1.46	1.40		
	(0.90)	(0.90)		
Any cultivable land?	0.68	0.65		
	(0.47)	(0.48)		
Land Owned (Acres)	34.24	29.73		
	(77.17)	(73.48)		
Observations	10,792	22,438	13,081	13,733

Table 13: Robustness: Comparability of Panel Households

Note: Standard Deviation is reported in parentheses. Panel HH(94) refers to households that were interviewed in 1994 as well as subsequent rounds of IHDS. Non-Panel HH (95) refers to households that were interviewed only once in 1994. Panel HH(05) refers to households that were interviewed in 1994 as well as subsequent rounds of IHDS. Refresher HH (05) refers to households from rural India that were interviewed for the first time in 2005, and they were re-interviewed in the later rounds of IHDS. As these households were added as a refresher sample, they were not observed in 1994.

	(1)	(2)
Relative IGEE	0.465	0.461
	(0.026)	(0.025)
Absolute IGEE	0.389	0.390
	(0.022)	(0.021)
Includes Non-Resident Sons	**No**	**Yes**
Observations	1,774	1,917

Table 14: Robustness: IGEE Results

Note: Robust standard errors are reported in parentheses. Panel A present results pertaining to resident members, whereas Panel B present results when non-resident members are also included in the sample. IGEE stands for intergenerational educational elasticity. Relative measure regresses sons' years of schooling against fathers' years of schooling. When we run the same regression after dividing sons' and fathers' years of schooling with their corresponding standard deviations, we obtain the absolute measure of IGEE.

	(1)	(2)
	Income$_{Son2012}$	**Income**$_{Son2012}$
Income$_{Father1994}$	-1.366	-3.012
	(0.334)	(2.175)
Square Income$_{Father1994}$	0.099	0.298
	(0.020)	(0.267)
Cube Income$_{Father1994}$		-0.008
		(0.011)
Observations	1,776	1,776

Table 15: Robustness: Non-Linear Specifications

Note: Robust standard errors are reported in parentheses. Son's income is the logarithm of scaled household income for son in 2012. Father's income is the logarithm of scaled household income for father in 1994. The control variables are 'age of father', 'age of son', and their squares. α=0.3 and θ=0.9 are used for scaling HH income.

	(1)	(2)
	IGE (Quadratic)	**IGE (Cubic)**
Mean	0.290	0.308
	(0.028)	(0.038)
Quintile 1	0.087	0.080
	(0.042)	(0.041)
Quintile 2	0.209	0.223
	(0.027)	(0.037)
Quintile 3	0.287	0.305
	(0.027)	(0.038)
Quintile 4	0.365	0.380
	(0.036)	(0.039)
Quintile 5	0.502	0.494
	(0.059)	(0.062)
Observations	1,776	1,776

Table 16: Robustness: Marginal Effects

Note: Robust standard errors are reported in parentheses. α=0.3 and θ=0.9 are used for scaling HH income.