**Migration, Gender and Intergenerational Transmission of Education**

**Abstract**

**Purpose –** This study provides new evidence on the intergenerational educational mobility in India. It also examines the effects of gender and migration background on heterogeneous transmission of educational attainment.

**Design/Methodology/Approach –** We use survey data from the India Human Development Survey, which is nationally representative dataset based on a multi-topic survey of over forty thousand Indian households. We estimate intergenerational correlation coefficient of Hertz et al. (2007) as an absolute measure of intergenerational educational mobility. The sources of intergenerational persistence in education are examined using the decomposition approach proposed by Checchi et al. (2013). We use the novel identification strategy of Lewbel (2012), which constructs internal instruments to obviate potential endogeneity issues in the standard regression based approach to measure intergenerational mobility.

**Findings** – We find that the educational persistence has declined over time. The decomposition of intergenerational correlation coefficient reveals that the primary source of educational persistence is the intergenerational transmission of education from highly-educated fathers to their highly-educated children. However, the negative contribution of illiterate fathers and their college-educated children has increased over time. We find that the children of migrants are more likely to display downward education mobility as compared to those of non-migrants. In addition to the ordinary least squares (OLS) estimate of the intergenrational regeression coeffcient, we also use alternative two-stage instrumental variable (IV) based estimation to address the potential endogeneity of parental education. We find that the IV estimates much larger than the OLS estimates, and they suggest that father’s education is an important predictor of child’s education.

**Originality/Value** – This study is one of the earliest attempts to examine the effect of migration status on the intergenerational transmission of education in India. The comprehensive dataset used in analysis allows us to examine the trends in intergenerational educational mobility over a period of five decades (1947–1996). Methodological improvements ensure that the results are robust to the sample selection bias, as well as the endogeneity bias.

**Keywords**: educational mobility, gender, migration, persistence, India

**JEL Codes**: I2; J5; J71; J13; J15

**1. Introduction**

Intergenerational mobility refers to the changes in the social status across different generations within the same family. The intergenerational educational mobility measures the changes in the level of educational attainment of an individual with respect to the education level attained by their parents. We examine the evolution of intergenerational educational mobility in India over a period of five decades (1947–1996). In addition, we study the effects of gender and migration background on the intergenerational mobility of education. Since educational attainment is the key determinant of occupational status and lifetime earnings, educational mobility is one of primary means through which individuals can change their social position. Low intergenerational educational mobility (or high persistence) preserves the inequality in educational attainment, as the children of less educated parents continue to be less educated and vice versa.

The social mobility measures are generally estimated in the context of educational attainment, occupational status or income. Examining social mobility in terms of education attainment rather than income is preferable as it obviates several estimation issues associated with the latter approach. First, any measure of income mobility is likely to be affected by life cycle bias, wherein different individuals realize their peak income at different ages across generations. Since people are likely to complete their education in their mid-twenties; therefore measuring educational mobility is likely to circumvent the measurement problems emanating from life-cycle bias (Black and Devereux, 2011). Second, educational mobility is less prone to measurement error as compared to income mobility, as individuals generally tend to be more forthcoming in revealing their level of education as opposed to revealing their exact income. Due to these inherent advantages, even the studies on income mobility often use education as a proxy for earnings or for calculating some measure of imputed earnings (Björklund and Jäntti, 1997; Causa and Johansson, 2010; Dearden et al., 1997).

A number of arguments can be adduced to support a positive association between parental education and child education, that is, the children of highly educated parents tend to pursue higher education and vice versa. First, a direct effect of having highly educated parents is that they are likely to earn more than their less-educated counterparts, and therefore, they can provide better education to their children. Second, ceteris paribus, highly educated parents are likely to have better unobserved abilities than their less educated counterparts. The inheritance of such unobserved abilities has an indirect effect on the child’s educational attainment. Third, highly educated parents are also likely to do a better allocation of their time and material resources while raising their children (Becker and Tomes, 1986; Craig, 2006; Guryan et al., 2008). Fourth, education affects the bargaining power of an individual. For instance, educated mothers are more capable of channeling the household expenditure towards better development of their children (Baker and Stevenson, 1986; Currie and Moretti, 2003; Ware, 1984).

In the Indian context, most of the studies estimate the intergenerational educational coefficient using a simple bivariate regression, where the child's education is regressed on the parents’ education (Hnatkovska et al., 2013; Jalan and Murgai, 2008; Majumder, 2010). However, this approach does not account for the temporal changes in the distributions of child and parent education. In addition, the conventional regression scheme is prone to endogeneity bias. This is because several unobserved characteristics may be associated with both child’s and parents’ education. For instance, highly educated parents may have better skills and abilities that are transmitted to their children, which in turn affect their educational outcomes. Similarly, parental education may determine how they choose to allocate their time and financial resources in raising their children. These parental choices affect the educational outcomes of their children. Most of the previous studies do not account for the potential endogeneity problem due to a lack of valid instrumental variables that are sufficiently correlated with the parental education but uncorrelated with the children’s education.

This study examines the intergenerational transmission of education for different age cohorts using IHDS data. We account for potential endogeneity problem in the conventional regression specification by generating synthetic instrument variables using the novel identification strategy of Lewbel (2012). The remainder of the paper is organized as follows. Section 2 provides a brief review of the literature on the intergenerational educational mobility. Section 3 and 4 explains the data and the research framework used for this study, respectively. Section 5 presents the results. Section 6 discusses trends in education spending in India. Last section presents concluding remarks.

**2. Literature Review**

Recently, considerable research effort has been directed towards studying intergenerational educational mobility. Using socioeconomic panel data from Germany, Heineck and Riphahn (2009) measure intergenerational educational mobility in Germany for a period of five decades. They provide estimates of transition matrices, which measure the probability of children achieving a certain level of education conditional on the education level of the parent. They find that despite significant policy interventions and education reforms during their sample period, parental education continued to exert a strong effect on child’s parental outcomes. Using an international sample of 42 countries, Hertz et al. (2007) found large geographical differences in intergenerational transmission of education, wherein the Latin American countries exhibit the lowest mobility whereas the Nordic countries display highest mobility. The global average of the correlation between parent and child schooling was estimated at 0.420 over a period of 50 years. Daude (2011) measures the intergenerational transmission of education for 14 Latin American countries and found high persistence in education attainment across generations. Tverborgvik et al. (2013) found that children of highly educated parents are three times more likely to receive basic education as compared to the children of less educated parents.

While, most of the aforementioned studies report a high correlation between education level of an individual and the education level of their parents, other factors such as gender, religion, or membership of certain social groups, can confound this relation. For instance, Farre and Vella (2013) found that the education level of sons are likely to be correlated to the education level of their father, while education level of daughters are likely to be correlated to the education level of their mothers. Borjas (1992) found that apart from the parental education level, education level of immigrant children is affected by the human capital of the ethnic community to which they belong. Further, if the immigrant population faces difficulty in integrating with the local society, the parental education level generally plays a pivotal role in determining their offspring’s education level. Lack of access to public resources offered to the natives act as a barrier which prevents immigrant children from climbing up the social ladder; hence, they are likely to depend more on private investments, such as household assets or parental education level, rather than public investments (Ammermueller, 2007; Schneeweis, 2011). In the Indian context, a majority of migrants relocate from rural to urban regions in search of better employment opportunities. Since, the education levels in the rural areas are lower than those in urban areas, the general education level of the migrant population is lower than that of their native counterparts. Consequently, to the extent that parental education is positively associated with the education level of their children, the migrant children would be inherently disadvantaged in comparison to the children of the natives.

Several past works on intergenerational education mobility suffer from endogeneity problem that makes it difficult to distinguish between the nature and the nurture effects. More specifically, if the children of highly educated parents tend to be highly educated themselves, is it because of the genetic traits passed from parents to their children, or is it because of a better learning environment provided by more educated parents. Some studies attempt to account for the endogeneity problem to establish a true causal link between parental education level and child’s education level (see, for example, Checchi and Flabbi, 2007; Schnepf, 2002). In the Indian context, Jalan and Murgai (2008) found that intergenerational mobility had increased over time, irrespective of social classes based on caste and wealth. They attempt to control for endogeneity of parental education level by using the extent of prenatal care received by cohort of mothers in 1992-93 as a proxy. However, there are some limitations associated with this study: First, their analysis is restricted to young children aged between 15-19 years, to avoid the sample selection bias caused by female children moving out of their birth family for marriage. Second, due to the limitations of their dataset, they are unable to find an exact measure of the prenatal care received by the mothers of the current 15-19 year olds, when their mothers were pregnant with them. Maitra and Sharma (2009) accounts for the potential endogeneity of parental educational attainment by using birth year of parents and their original location as instrument variables for parental educational attainment. Interestingly, their results show that after controlling for endogeneity issue, the causal effect of parental education level on the education level of the children is insignificant. Using data from the National Sample Survey (NSS), Hnatkovska et al. (2013) investigate intergenerational mobility in terms of educational attainment, occupational choices and wages over the period 1983 to 2005. Their findings show a significant increase in intergenerational mobility among the disadvantaged section of the society, thereby moving towards their socially well-off counterparts in terms of education, occupation and wages. This analysis suffers from two issues. First, the intergenerational relationships are studied only for those parent-children pairs that are co-resident in the same household. Therefore, the estimated coefficients of intergenerational mobility suffer from potential sample selection bias. Second, unlike the IHDS dataset, the NSS data does not track the same household over time. This means that long term intergenerational comparisons are problematic as the authors only have one point in time observation for each parent-child pair. Emran and Shilpi (2015) found that family background and community factors play a significant role in affecting the educational outcomes of children in India. Azam and Bhatt (2015) found that intergenerational mobility in terms of educational attainment has increased, irrespective of one’s association with a particular social group. This study makes several contributions to the existing literature of intergenerational mobility of education in India. First, we provide robust estimates of intergenerational educational mobility by addressing the endogeneity problem using the novel two stage estimation procedure of Lewbel (2012). Second, whereas, Azam and Bhatt (2015) concentrate on intergenerational education mobility by analyzing only father­–son pairs, we extend their analysis by including all four intergenerational pairs­—father­s & sons, fathers & daughters, mothers & sons, and mothers & daughters. This allows us to gain unique insight in the role of gender in influencing intergenerational education mobility. This analysis is of particular interest as gender inequality is pervasive in India, and Indian females are distinctly disadvantaged in comparison to their male counterparts over most socio-economic criteria (see, for example, Ackerson and Subramanian, 2008; Arora, 2012; Behrman, 1988; Bhattacharya, 2006; Borooah, 2004; Dunn, 1993; Jacobs, 1996; Kishor, 1993; Murthi et al., 1995). Third, the panel dataset of IHDS surveys allows us to examine the role of migration on intergenerational education mobility. In this study, we define migrants as individuals who relocated between the intervening period of the two rounds of the IHDS surveys, i.e., between 2004 and 2011, and who were enrolled at the time of migration. To the best of our knowledge, this is the earliest attempt to examine the impact of migration on intergenerational education mobility in India. Fourth, this study also tries to identify the source of persistence in transmission of educational mobility among migrants and non-migrants by decomposition of intergenerational correlation coefficient.

**3. Data and variables definitions**

We use data from both rounds of IHDS surveys: IHDS-I surveys conducted in the year 2004-05 and IHDS-II surveys conducted in the year 2011-12. The IHDS-I (II) surveys are a nationally representative dataset acquired from multi-topic survey of 41,554 (42,152) Indian households. The IHDS surveys provide a panel dataset as 83% of the household interviewed under the IHDS-I surveys were re-interviewed under IHDS-II surveys. Our dataset provides information on children born in or after 1947. We divide our sample into five ten-year birth cohorts: 1947-1956, 1957-66, 1967-76, 1977-86, and 1987-96. The dependent variable used in this analysis is the completed years of education for children aged 15 and above. The education variable is a continuous variable ranging from 0 to 16 years, where 0 represents ‘no education’ and 16 represents highest educational qualification obtained by the child. Next, we decompose the overall intergenerational educational mobility measure into upward and downward mobility measures. In order to facilitate this decomposition, we define five ordinal categories for the level of educational attainment: No education=0 years; Primary=1-5 years; Middle= 6-8 years; Secondary=9-12 years; and College= 13-16 years. The primary explanatory variable in the regression analysis is the education level of parents, and its regression coefficient provides a measure of intergenerational educational mobility. In addition, we incorporate a number of control variables that can affect the education attainment of the child. These include age, household size, ethnicity, social status, city of residence (metro/non-metro), location (rural/urban), and state of residence. Overall, our sample comprises 16,253 father-son pairs, 15,528 mother-son pairs, 4987 father-daughter pairs, and 4,763 mother-daughter pairs.

Next, we measure the degree of intergenerational educational mobility among children of migrant households. This analysis identifies migrants as those children who migrated between the two survey rounds of IHDS: IHDS-I and IHDS-II, and those who were enrolled at the time of the IHDS-I surveys (2004-05) but completed their education by the time of IHDS-II surveys (2011-12). This sample of children allows us to measure the impact of migration on intergenerational mobility of education. Although, the IHDS surveys provide a binary variable that indicates whether an individual was enrolled at the time of the survey, there are a large number of missing observations for this variable. To overcome this limitation, we use the following approach. First, using the dataset of IHDS-I survey, we identify the starting year of education as the 2005 – number of years of education reported in IHDS-I survey. Second, we identify the ending year of education as 2005 + (number years of education in IHDS-II - number years of education in IHDS-I). Next, we select only those children for which the ending year of education is less than 2012, which implies that they finished their education before 2012. Third, we identify children whose families migrated when they were enrolled. IHDS-II includes a question “How many years ago did your family first come to this village/town/city?” This allows us to calculate the year of migration for the household. Finally, we identify children whose families migrated when they were enrolled using the following rule: the starting year of education is less than the year of migration, and the year of migration is less than the ending year of education. Table 1 reports some descriptive statistics for the father-son pairs in the full sample, and the migrant sample. In general, average years of educational attainment is higher for urban areas than their rural counterparts, for both full sample and migrant sample.

**4. Methodological framework**

*4.1 Intergenerational Correlation Coefficient (ICC)*

We estimate the following bivariate regression model of Hertz *et al.* (2007) to measure the intergenerational educational mobility:

(1)

where is the number of years of education attained by the child *i*. is the number of years of education attained by the parent of child *i*. is the intergenerational regression coefficient (IRC) that represents the marginal effect of the parent’s education on the child's educational attainment. Ceteris paribus, lower values of IRC would indicate higher intergenerational mobility of education, and vice versa.

The intergenerational correlation coefficient (ICC), , can be estimated as follows

(2)

Here the superscript refers to parents (fathers or mothers) and the superscript refers to children (sons or daughters). We divide our sample into five ten-year birth cohorts of chidren. For a specific birth cohort, is the standard deviation of the number of years of education of all children in that cohort. is the standard deviation of the number of years of education of their parents.

The value of ICC ranges between -1 and +1. The main difference between IRC and ICC is that the latter normalizes the educational attainment of a population by the standard deviation of their educational distribution (Black and Devereux, 2011). Note that in the estimation of ICC, the differences in the distribution of education of parents and their children are normalized by using a ratio of standard deviations of the years of education parents and their children, for each birth cohort. Therefore, ICC provides an absolute measure of intergenerational educational mobility. Conversely, for a given birth cohort, the magnitude of IRC depends on the differences in the distribution of education of parents and their children. A decrease in ICC over time may suggest that the correlation between the education of parents and their children has reduced. However, ICC measure could also fall over time because the variance of education in the child’s generation is lower than that in the parent’s generation following the introduction of educational reforms such as compulsory primary school education. Hertz *et al.* (2007) show that both IRC and ICC may change in opposite directions over time, and therefore, we report both of these measures of intergenerational educational mobility.

*4.2 Decomposition of intergenerational correlation coefficient*

From the perspective of policy makers, upward mobility may be viewed as a more preferable outcome than downward mobility. However, irrespective of the type of mobility, high levels of mobility would reduce ICC. Therefore, better understanding of intergenerational transmission of education can be gained by examining the source of ICC. The ICC measure can be decomposed as follows

(3)

Where, for a particular birth cohort, the superscript indicates the set of all children (either sons or daughters) in the same birth cohort; and the superscript indicates the set of all parents (either fathers or mothers) of the children in that cohort. and are the mean education level (measured as the number of years of education) for the children and parents, respectively. and are the standard deviations of the distribution of education level for the children and parents, respectively. is the covariance between the child’s and the parent’s levels of education. and are the education level of a particular parent-child pair. is the conditional probability of the child achieving a certain education level, , given that the parent’s education level is . is the probability that the parent has the education level, .

The decomposition in equation (3) implies that depends on the combined deviations of the child’s and the parent’s education from the mean, the conditional probability of child achieving a certain education level given the parent’s education level, and the probability of the parent achieving that level of education. As long as both generations of a parent-child pair achieve levels of education above (or below) the mean for their own cohort, this parent-child pair will contribute positively to , whereas if the parent is above (below) and the child is below (above) the mean levels of education for their own cohort, the pair will contribute negatively to .

*4.3 Two-stage instrumental variable (IV) estimates based on Lewbel (2012)*

A significant issue with standard regression based approach to measure intergenerational education mobility is that parental education variables are invariably endogenous variables. In the presence of endogeneity, the ordinary least squres (OLS) coefficients are biased and inconsistent. Unfortunately, thesecondary data sources rarely provide information about valid instruments, which are sufficiently correlated with parental education but not with children's education. To overcome this limitation, we employ the novel identification strategy of Lewbel (2012). The strategy involves a standard two stage estimation process. In the first stage, we generate synthetic instrumental variables (IV) for all endogenous regressors. These instruments are some linear combination of exogenous or control variables. In the second stage, the model is estimated by substituting instrumental variables for the endogenous regressors. This approach can be defined as follows.

We estimate the following three-equation model:

(4)

(5)

(6)

where and is the number of years of education for the father and mother of the *i*th child. are a set of control variables as listed in Section 3. Lewbel (2012) shows that in the presence of two endogenous variables, and , their regression coefficients and can be consistently identified if there exist some exogenous variables in , satisfying the following three conditions:

(i) E() = 0, E(, E(; (ii) Cov(,Cov; and (iii) Cov ( and Cov (, where can be a subset of the exogenous variables or itself.

Under the above assumptions, ( and ( can be used as the excluded instrument vector for and . The strength of the instruments corresponds to the degree of heteroscedasticity of and with respect to

**5. Results**

Table 2 reports the regression coefficients and the correlation coefficients for father-son pairs. The dependent variable is the years of education attained by the son and the independent variable is the number of years of schooling of the father. With the exception of the youngest cohorts, the average years of schooling of sons and their fathers has seen an upward trend. All regression and correlation coefficients are positive and statistically significant; however, there is a steady decline in their magnitude across birth cohorts. This implies that parental education still has a significant effect on child’s educational outcomes. The correlation coefficients are larger than the regression coefficients across all the five birth cohorts. This indicates that despite the increase in the average years of schooling of the child, parental education level still plays a pivotal role in affecting child’s educational outcomes. Though, the recent birth cohorts have seen a decline in the mean years of schooling. The urban sector exhibits higher degree of persistence (implying less mobility) as compared to the rural sector. Mobility has risen in the rural sector as regression coefficient has declined from 0.42 to 0.25 across the five birth cohorts. It may be argued that as the average years of schooling increases over time; the effect of parental education level will decline. However, results show that the correlation coefficient has also declined from 0.58 to 0.34 across the five birth cohorts. Here, correlation coefficient normalizes the average years of schooling of fathers and sons by the standard deviation of their educational distribution- implying that decline in persistence should not be attributed to the general increase in the average years of schooling over time.

In Table 3, we report OLS regression estimates for intergenerational mobility in terms of educational attainment. The results show that intergenerational coefficients are positive and statistically significant at 1 per cent level. In this analysis, both education variables are measured in number of years of schooling, therefore the magnitude of coefficient will reflect the rate at which the differences in education is likely to lessen across generations. Hence, our estimate shows that a one–year decline in the schooling of fathers results in a decline of 0.29 to 0.38 years in the educational level of the sons. The estimates for the control variables show that children from households residing in rural areas are more susceptible to less education than their urban counterparts. The size of household is negatively related to the child’s educational outcomes. A one-unit increase in the household size is associated with 0.014 to 0.051 years decline in children’s years of schooling. The association with one of the marginal groups (SC/ST/OBC) of the society decreases the child’s likelihood of attaining higher education. We also examine the question of whether the degree of intergenerational educational mobility varies across social groups. The coefficients of interaction terms for the marginal groups are negative and significant, whereas for others it is positive and significant. This suggests that the marginal sections of the society (SC/ST/OBC) are more mobile in terms of educational attainment than their non-marginalized counterparts.

The last two columns of table 3 report the estimates for the instrumental variable (IV) approach of Lewbel (2012). The two-stage IV specification of Lewbel (2012), represented by equations 4, 5 and 6, can be estimated using the standard two stage least squasres (2SLS) or the generalized method of moments (GMM). In the presence of heteroskedasticity, GMM estimates are considered more efficient than standard 2SLS estimates (Baum et al., 2003). We use the GMM estimation for the results reported in Table 3. The estimates provide strong and positive effect of parental education on their child educational attainment. The IV estimates indicate that a one-year difference in the schooling of fathers result in a difference of 0.60 to 0.80 years in the education level of the child.

The difference between the IV and OLS estimates shows that failing to account for the potential endogeneity of the explanatory variables can result in incorrect inference. The nature effects measure the transmission of unobserved ability and genetic endowments, while nurture effect measures the causal effect of parental education on children’s educational attainment. The OLS method captures the effect of nature as well as nurture effect, whereas the IV estimates capture only nurture effect by controlling for unobserved heterogeneity. Thus, after controlling for unobserved heterogeneity, we find very large causal impact of parental education on child’s schooling outcomes, which implies that there is a high degree of intergenerational educational persistence.

Intutition suggests that OLS estimates should be biased upwards, since any unobserved omitted variable should have similar relation with the education attainment of child and their parents. For instance, genetic ability can be one of the unobserved variables. Higher genetic ability is likely to positively affect parental education level as well as the education level of the child. This implies that this omitted variable would tend to have positive coefficient if included in the true model. Therefore, in the absence of omitted variable, the OLS estimates should have an upward bias. However, our IV estimates are larger than OLS estimates. This contradiction could be attributed to one or more of the following explanation. First, our education data relies on self-reported information by an individual; therefore, there is chance of some degree of measurement error. It is likely that downward bias due to measurement error overpower the omitted variable upward bias. Second, there could be some unobserved variables that are negatively correlated with father’s educational level but positively correlated with the educational attainment of the child, leading to downward bias in estimating intergenerational educational attainment. Third, it is argued that IV estimates are not capable of capturing the average effect on entire population, rather the local effects on the part of population which is affected by the instruments being used. For example, in our IV approach we have used social status of individual as one of the exogenous instruments. Therefore, it is assumed that the treatment effect of parental education is likely to differ depending on individual’s association with a particular social group.

Next, we perform some robustness tests to check the validity and relevance of the synthetic instruments. In the 2SLS specification, The Durbin-Wu-Hausman test rejects the null hypothesis of exogeneity for father’s educational attainment at 1 per cent level of significance. This result implies that estimates obtained though application of OLS would suffer from endogeneity bias. The Breush-Pagan heteroskedasticity test rejects the null hypothesis of constant error variance at 1 per cent level of significance, which indicates the presence of heteroskedasticity in the error process. We also run a Langrange-Multiplier(LM) test for underidentification using the Kleibergen and Paap rk (2006) statistic. The Kleibergen-Paap LM statistic rejects the null hypothesis that the model is under-identified and the generated instruments are weakly identified. The Hansen J statistic is statistically insignificant, which suggests that the overidentifying restrictions are valid.

Table 4 reports the measures of intergeneration mobility of education for all parent-child combinations. We find that irrespective of the way of measuring educational mobility, father’s education has a larger transmission effect on son’s education than the mother’s education. Interestingly, the effect of parental educational level on child’s educational outcomes is higher for daughters than sons. In other words, men are more mobile than women which could be attributed to men having better access to educational resources. We would like to add a caveat here: our sample for daughter over-represents young daughters aged 30 and below. Therefore, coefficients on daughters should be considered indicative, and would require further analysis. As earlier, we run a suite of diagnostic tests to check the validity of the IV approach, and we find that the instrumental variables are well specified, and the IV estimates are preferable over the OLS estimates.

Table 5 displays the intergenerational educational mobility coefficients between parent’s years of schooling and children’s schooling for children between the ages of 15-24 years. The age limit has been put in order to avoid sample selection bias, thus making comparisons between parent-child pairs more reliable. This is because both sons and daughters are generally likely to co-reside with their parents in the age-group 15-24, whereas the married daughters are generally do not reside with their parents. Table 5 reports some interesting results that contradict the aggregate results reported in Table 4. A comparison of coefficient estimates for sons and daughters shows that the father’s education exerts larger influence on the schooling of son’s, while the educational attainment of daughters is more strongly correlated with the education of their mothers. These findings are consistent with the role model hypothesis, which states that parent’s schooling exerts much larger influence on the schooling of their children with the same gender. This implies that father’s schooling has a higher correlation with the education level of the sons than daughters, while mother’s schooling has a higher correlation with the education level of the daughters than sons. In other words, transmission of education follows gender lines.

We also study the role of migration background in intergenerational transmission of education. Our migrant sample comprise of less than 2 per cent of the total sample size of father-son pairs. Table 6 presents the regression and correlation coefficients for the migrant sample. For the migrant sample, the correlation coefficient is 0.307 and statistically significant at 1 per cent level of significance. This indicates that, ceteris paribus, a one-year difference in schooling of fathers result in a difference of approximately 0.30 years in the years of schooling of the sons. Further, the IV estimates show that the importance of father’s years of schooling as a predictor of their son’s educational attainment increase significantly after controlling for potential endogeneity of parental schooling. The last column of Table 6 shows the estimates using the IV approach. The correlation coefficient of father’s educational attainment indicates that, ceteris paribus, a one-year difference in the schooling of fathers will result in a difference of approximately 0.65 years in the schooling of their sons.

As expected, other things remaining the same, the age of the child positively affects the average years of schooling attained by the child. The coefficient for the age of the child at the time of migration is negative and significant. This implies that, ceteris paribus, an increase in the age at the time of migration decreases the average years of schooling attained by the child. Therefore, age of the child at the time of migration is an important indicator of the level of education attinament of the child. Diagnostic statistics reveal that that the instrumental variables are well specified, and the IV estimates are preferable over the OLS estimates.

Table 7 shows the estimated transmission coefficient for all the parent-child pairs in the migrant sample. The correlation coefficient beteen the education level of parents and children are positive and significant for all four parent-child pairs. This shows that parental educational level is positively related to the educational level of the child. Additionally, we find that the paternal schooling is more strongly correlated to the years of schooling of sons than that of daughters, whereas maternal schooling is more strongly correlated to the years of schooling of daughters than that of sons. Therefore, gender role is significant in intergenerational education transmission the migrant population.

As espected, the age of the child is positively related to her educational attainment. The estimated association between the age at the time of migration and the educational level of the child is negatively significant at 1 per cent level of significance. The decline in the average educational level of the child due to a unit increase in the age at the time of migration is much larger for daughters than sons. The results show that daughters experience double disadvantage, being less mobile than their male counterparts and experiencing proportionately more negative child outcomes associated with the age at the time of migration. However, as the coefficients on intergenerational educational mobility for daughters have been drawn from small sample size, interpretation must be taken as purely suggestive.

Table 8 compares the degree of intergenerational mobility among the migrant and the non-migrant population. The results indicate that migrants are more mobile than their native counterparts. However, the intergenerational transmission coefficient doesn’t account for the fact that migrants are likely to experience upward mobility as their parents are often less-educated than their native counterparts. In other words, a lower value of intergenerational transmission coefficient implies greater mobility, but does not reveal the direction of such mobility.

We identify the sources of intergenerational persistence in by decomposing the intergenerational correlation coefficients following the approach of Checci et al. (2008). Table 9 presents the results of a decomposition of correlation coefficient for father–son pairs across different birth cohorts of sons. Using Equation (3), the correlation coefficient for each child cohort has been divided into 25 different components, which correspond to associations between to the five education levels of fathers, each associated with five education levels of sons. For example, the first panel of the Table 9 decomposes the intergenerational correlation coefficient of the group of sons having non-literate fathers in five categories corresponding to five education levels of sons. The vertical sum of these figures in each column gives the correlation coefficient for the specific child cohort. The correlation coefficient has declined from 0.621 for the oldest cohort to 0.373 for the youngest cohort. This implies that the degree of persistence in terms of educational attainment has declined over time, or an increase in mobility. However, it is interesting to identify the source of persistence as this increase in mobility could be attributed to two reasons. First, it could be upward mobility due to fathers who are illiterate or who have attained primary education and whose children have completed high school and college education. Second, it could be downward mobility due to fathers who have completed high school education or have obtained their college degree and whose children have achieved education level which is below the mean education level for their own cohort. Generally, upward mobility should be desirable as compared to downward mobility. However, upward mobility can exacerbate inequality in education attainment if it is primarily displayed by children of higly educated fathers but not by the children of fathers having low levels of education.

The results of table 9 indicate that the main source of persistence in intergenerational transmission of education is the positive contribution of highly-educated fathers having children who are either equally well-educated of better-educated than their fathers. There is an increasing proportion of low-educated fathers having children who are also low-educated and declining proportion of low-educated fathers having children who are highly-educated. The proportion of total positive contribution to the correlation coefficient of the group of sons with illiterate fathers has increased from 48 per cent for the oldest birth cohort to 62 per cent for the youngest birth cohort, whereas the proportion of positive contribution to correlation coefficient of the group of sons with fathers who have completed college education has declined from 37 per cent for the oldest cohort to 27 per cent for the youngest cohort. Although, the decline in the proportion of positive contribution to correlation coefficient at the upper end of the educational distribution is compensated by the increase in the proportion of positive contribution to correlation coefficient of the group of sons with fathers who have obtained secondary education has increased from 19 per cent for the oldest birth cohort to 40 per cent for the youngest birth cohort. Therefore, the biggest challenge that we are facing in terms of intergenerational educational mobility is the presence of high degree of positive persistence at the lower end of the educational distribution.

Table 10 shows similar decomposition for the sample of migrant households. the first point to note at the upper end of the educational distribution is the increasing positive contribution of fathers with college education and children who are also college-educated, which is a good sign. But, also note the increasing negative persistence of fathers who are highly educated and who have sons who are moving down the social ladder relative to their parents.

We find that the IV estimates are far larger than the corresponding OLS estimates, regardless of the specification and the sample. As discussed earlier, one of the reasons for IV estimates to be larger than OLS estimates could be attributed to the presence of some unobserved variables that are negatively (positively) correlated to the educational level of the parents and positively (negatively) correlated to the educational level of the child. In the Indian context, this could be attributed to the fact that the proportion of public expenditure on education towards primary education has been increasing as compared to expenditure on higher education. For instance, the proportion of expenditure on elementary education by the central government has increased from 13.74% in 1990-91 to 61.19% in 2005-06. The concentration of financial resources on primary education might help the less-resourceful children to gain primary education, but at the same time it inhibits other to pursue higher education due to high cost associated with higher education. This implies that it would be easier for the children of non-literate parents to gain primary education, whereas, rising cost of higher education may deter some children of highly educated parents to attain higher education. Therefore, unobserved institutional environment may such as the allocation of public expenditure on education, may result in a negative association between the education levels of the parents and their children.

**6. Conclusion**

We draw three major conclusions from our analyses. First, we find that although the degree of persistence in terms of education has declined steadily, implying increasing mobility; parental education still plays a very significant role in affecting child’s education level. Restricting the sample to youth population aged between 15 to 24 years, we find that paternal education level exerts larger influence on sons than on daughters, whereas maternal education level exerts larger influence on daughters than on sons. This shows that education level of parents is highly correlated to the education level of their children with same gender.

Second, decomposing the correlation coefficient between father’s years of schooling and child’s years of schooling reveals that although the intergenerational persistence has declined over time, implying an increasing mobility; it is still significant. The positive persistence at the lower end of the education distribution has increased from 43 per cent to 61 per cent over time, while negative persistence at the lower end has increased from 5 per cent to 22 per cent. In other words, the proportion of highly-educated fathers with sons who are equally well-off in terms of education has increased over time and the proportion of illiterate fathers with sons who are highly-educated has declined over time. The intergeneration correlation of educational level among migrants shows that migrants are more mobile than their native counterparts. However, decomposing the correlation coefficient between father’s years of schooling and child’s years of schooling among the migrants show that higher mobility in terms of educational attainment observed among migrant population is primarily due to the migrant children moving down the education ladder relative to their parents.

Third, the traditional instrumental variable approach is of limited use with secondary databases, as it is difficult to identify variables which are independent of error term and which do not affect the dependent variable when independent variable is held constant (exclusion restriction). We use an alternative identification strategy proposed by Lewbel (2012) which replaces endogenous regressors, such as parental education, with synthetic instrumental variables constructed using linear combinations of exogenous regressors. The major advantage of this identification strategy is that it does not rely on the standard exclusion restriction. after controlling for potential endogeneity of parental education level, we find that the IV estimates are considerably larger than the corresponding OLS estimates. There are two reasons why the OLS estimates may be downward biased relative to the IV estimates. First, there could be some unobserved variables that are negatively correlated with father’s educational level but positively correlated with the educational attainment of the child, leading to a downward bias in OLS coefficients of intergenerational education transmission. Second, the coefficients for the instrumental variables may not capture the average effect on entire population, but the local effects on the part of population which is affected by the instruments being used. For example, in our instrumental variables approach we have used social status of individual as one of the exogenous instruments. Therefore, it is assumed that the treatment effect of parental education is likely to differ depending on individual’s association with a particular social group.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1: Descriptive Statistics for the father-son pairs** | | | | |
|  | **Full Sample** | | **Migrant Sample** | |
|  | Rural | Urban | Rural | Urban |
| ***Child Educational Outcomes*** |  |  |  |  |
| Mean Years of Schooling | 9.02 | 10.39 | 11.01 | 12.75 |
| *Proportion of Children with:* |  |  |  |  |
| No schooling: 0 years | 0.00 | 0.00 | 0.00 | 0.00 |
| Up to primary level: 1-5 years | 0.16 | 0.11 | 0.07 | 0.01 |
| Middle school completion: 6-8 years | 0.25 | 0.19 | 0.12 | 0.05 |
| Up to Higher secondary: 9-12 years | 0.47 | 0.43 | 0.53 | 0.43 |
| Graduate and above: 13-16 years | 0.11 | 0.26 | 0.27 | 0.51 |
| Mean age | 26.67 | 27.67 | 23.56 | 23.91 |
| ***Parent's characteristics*** |  |  |  |  |
| Mean years of schooling | 4.31 | 6.98 | 5.86 | 9.83 |
| Father's age | 57.58 | 58.2 | 53.43 | 52.94 |
| ***Proportion of children from different caste groups*** | |  |  |  |
| Scheduled Castes (SC) | 0.21 | 0.18 | 0.13 | 0.19 |
| Scheduled Tribes (ST) | 0.09 | 0.03 | 0.08 | 0.04 |
| Other backward castes (OBC) | 0.40 | 0.41 | 0.33 | 0.31 |
| Forward caste (FC) (Including Brahmins) | 0.28 | 0.01 | 0.43 | 0.45 |
| Others | 0.01 | 0.35 | 0.01 | 0.00 |
| ***Proportion of children from different ethnic groups*** | |  |  |  |
| Hindus | 0.82 | 0.75 | 0.65 | 0.85 |
| Muslims | 0.11 | 0.18 | 0.14 | 0.09 |
| Other religion | 0.07 | 0.05 | 0.21 | 0.05 |
|  |  |  |  |  |
| Number of observations | 10411 | 5842 | 58 | 224 |

Notes: Only children for whom child-parent relationships can be identified in the dataset, as described in the text, are included in the sample.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 2: Regression and Correlation Coefficients for father-son pairs, All-India** | | | | | | |
| **Birth cohort** | **Regression coefficient (β)** | **Correlation coefficient (ρ)** | **Average years of schooling (child)** | **Average years of schooling (parent)** | **S.D. of education (child)** | **S.D. of education (parent)** |
| **All India** | | | | | | |
| 1947-1956 | 0.51 | 0.66 | 10.08 | 6.34 | 3.76 | 4.95 |
| 1956-1966 | 0.40 | 0.52 | 10.41 | 5.77 | 3.70 | 4.84 |
| 1967-1976 | 0.34 | 0.47 | 10.48 | 6.03 | 3.51 | 4.91 |
| 1977-1986 | 0.35 | 0.47 | 10.18 | 5.91 | 3.50 | 4.70 |
| 1987-1996 | 0.29 | 0.39 | 8.47 | 4.35 | 3.14 | 4.24 |
| ***Overall*** | **0.35** | **0.36** | **9.51** | **5.27** | **3.48** | **4.61** |
| **Rural** | | | | | | |
| 1947-1956 | 0.42 | 0.58 | 8.13 | 4.16 | 3.22 | 4.44 |
| 1956-1966 | 0.29 | 0.36 | 9.56 | 4.27 | 3.48 | 4.32 |
| 1967-1976 | 0.30 | 0.38 | 9.92 | 4.75 | 3.47 | 4.49 |
| 1977-1986 | 0.29 | 0.38 | 9.57 | 4.77 | 3.35 | 4.36 |
| 1987-1996 | 0.25 | 0.34 | 8.25 | 3.77 | 2.99 | 4.00 |
| ***Overall*** | **0.29** | **0.36** | **9.02** | **4.30** | **3.29** | **4.25** |
| **Urban** | | | | | | |
| 1947-1956 | 0.38 | 0.54 | 12.17 | 8.67 | 3.16 | 4.43 |
| 1956-1966 | 0.48 | 0.61 | 11.75 | 8.12 | 3.65 | 4.67 |
| 1967-1976 | 0.36 | 0.51 | 11.42 | 8.19 | 3.39 | 4.82 |
| 1977-1986 | 0.39 | 0.52 | 11.12 | 7.63 | 3.51 | 4.68 |
| 1987-1996 | 0.36 | 0.47 | 8.95 | 5.61 | 3.39 | 4.45 |
| ***Overall*** | **0.41** | **0.36** | **10.39** | **6.98** | **3.62** | **4.73** |
|  |  |  |  |  |  |  |
| Notes: All β & ρ coefficients are significant at a 5% level. | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3: Intergenerational transmission of human capital for father-son pairs** | | | | |
| **Variables** | **Ordinary Least Square (OLS)** | | **IV-Lewbel** | |
| ***Regression coefficient*** | ***Correlation coefficient*** | ***Regression coefficient*** | ***Correlation coefficient*** |
| *father\_edu* | 0.293\*\*\* | 0.389\*\*\* | 0.604\*\* | 0.801\*\*\* |
|  | (0.005) | (0.007) | (0.040) | (0.053) |
| *age\_child* | 0.068\*\*\* | 0.019\*\*\* | 0.035\*\*\* | 0.010\*\*\* |
|  | (0.004) | (0.001) | (0.006) | (0.002) |
| *age\_square* | -0.005\*\*\* | -0.001\*\*\* | -0.004\*\*\* | -0.002\*\*\* |
|  | (0.000) | (0.000) | (0.000) | (0.000) |
| *age\_father* | 0.006\* | 0.001\* | 0.022\*\*\* | 0.006\*\*\* |
|  | (0.003) | (0.001) | (0.004) | (0.001) |
| *hhsize* | -0.051\*\*\* | -0.014\*\*\* | -0.017\* | -0.005\* |
|  | (0.008) | (0.002) | (0.010) | (0.003) |
| *rural* | -0.474\*\*\* | -0.134\*\*\* | -0.238\*\*\* | -0.068\*\* |
|  | (0.054) | (0.015) | (0.108) | (0.031) |
| *OBC* | -0.588\*\*\* | -0.166\*\*\* | -0.182\*\* | -0.052\*\* |
|  | (0.059) | (0.017) | (0.083) | (0.024) |
| *SC* | -0.887\*\*\* | -0.249\*\*\* | -0.049 | -0.014 |
|  | (0.071) | (0.020) | (0.132) | (0.038) |
| *ST* | -0.653\*\*\* | -0.184\*\*\* | -0.332\* | 0.095\*\* |
|  | (0.102) | (0.029) | (0.169) | (0.048) |
| *Others* | -0.626\*\*\* | -0.176\*\*\* | -0.234 | -0.067 |
|  | (0.206) | (0.058) | (0.225) | (0.064) |
| *Muslim* | -1.112\*\*\* | -0.313\*\*\* | -0.481\*\*\* | -0.138\*\*\* |
|  | (0.076) | (0.022) | (0.115) | (0.033) |
| *Other\_reg* | -0.162 | -0.046 | -0.256\*\* | -0.073\*\* |
|  | (0.109) | (0.031) | (0.123) | (0.035) |
| *non\_metro* | -0.433\*\*\* | -0.123\*\*\* | -0.103 | -0.029\* |
|  | (0.106) | (0.029) | (0.123) | (0.035) |
| *OBC\*father\_edu* | -0.019\* | -0.005\* | -0.013\* | -0.002\*\* |
|  | (0.0116) | (0.003) | (0.009) | (0.001) |
| *SC\*father\_edu* | -0.060\*\*\* | -0.017 \*\*\* | -0.042\*\*\* | -0.014\*\*\* |
|  | (0.014) | (0.004) | (0.003) | (0.002) |
| *ST\*father\_edu* | -0.038\*\* | -0.010\* | -0.025\*\* | -0.008\*\* |
|  | (0.022) | (0.066) | (0.013) | (0.004) |
| *Others\*father\_edu* | 0.054 | 0.015 | 0.061 | 0.028 |
|  | (0.047) | (0.013) | (0.051) | (0.041) |
| No. of obs. | 16,253 | | 16,253 | |
| Adj. R-squared | 0.3198 | | 0.3121 | |
| **Diagnostic statistics** | |  |  |  |
| Breush-Pagan LM statistic | | | 46.34\*\*\* | |
| Kleibergen-Paap rk LM statistic | | | 360.87\*\*\* | |
| Hansen J statistic | | | 1.21 | |
| Durbin-Watson statistic | | | 72.95\*\*\* | |

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Robust standard errors in parentheses

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 4: Intergenerational correlation coefficients by parent-child pairs** | | | | | | | | | |  |
|  | **Father-Son** | | | **Father-Daughter** | | **Mother-Son** | | | **Mother-Daughter** | |
|  | **OLS** | | **IV-Lewbel** | **OLS** | **IV-Lewbel** | **OLS** | | **IV-Lewbel** | **OLS** | **IV-Lewbel** |
| *Regression coefficient* | 0.292\*\*\* | | 0.604\*\* | 0.324\*\*\* | 0.746\*\*\* | 0.305\*\*\* | | 0.309\*\*\* | 0.371\*\*\* | 0.656\*\*\* |
|  | (0.005) | | (0.040) | (0.002) | (0.202) | (0.007) | | (0.025) | (0.002) | (0.104) |
| *Correlation coefficient* | 0.388\*\*\* | | 0.801\*\*\* | 0.394\*\*\* | 0.851\*\*\* | 0.331\*\*\* | | 0.334\*\*\* | 0.381\*\*\* | 0.675\*\*\* |
|  | (0.008) | | (0.053) | (0.002) | (0.262) | (0.007) | | (0.026) | (0.002) | (0.107) |
| No.of observations (N) | 16,253 | | | 4,987 | | 15,528 | | | 4,763 | |
| **Diagnostic statistics** |  |  | |  |  |  |  | |  |  |
| Breush-Pagan LM statistic | - | 46.34\*\*\* | | - | 43.55\*\*\* | - | 901.74\*\*\* | | - | 407.42\*\*\* |
| Kleibergen-Paap rk LM statistic | - | 360.87\*\*\* | | - | 9.72\*\* | - | 347.62\*\*\* | | - | 62.92\*\*\* |
| Hansen J statistic | - | 1.21 | | - | 6.51\* | - | 2.83 | | - | 0.12 |
| Durbin-Watson statistic | - | 72.95\*\*\* | | - | 4.14\*\* | - | 0.035 | | - | 7.36\*\*\* |
| Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Robust standard errors in parentheses. | | | | | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 5. Regression and Correlation coefficient for children aged between 15-24 years** | | | | |
|  | **Father-Son** | **Father-Daughter** | **Mother-Son** | **Mother-Daughter** |
| *Regression coeffcient* | 0.223\*\*\* | 0.219\*\*\* | 0.271\*\*\* | 0.305\*\*\* |
|  | (0.008) | (0.001) | (0.011) | (0.015) |
| *Correlation coefficent* | 0.295\*\*\* | 0.265\*\*\* | 0.272\*\*\* | 0.314\*\*\* |
|  | (0.0108) | (0.014) | (0.011) | (0.015) |

Notes: The table gives the intergenerational educational regression and correlation coefficient for children aged between 15-24 years. Standard errors are in parenthesis. \*\*\* Statistically significant at p<0.01 level.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 6. Intergenerational transmission of education among migrants (Dependant variable: son's years of schooling)** | | | | |
| **Variables** | **Ordinary Least Square (OLS)** | | **IV-Lewbel** | |
| **Regression coefficient** | **Correlation coefficient** | **Regression coefficent** | **Correlation coefficient** |
| father\_edu | 0.215\*\*\* | 0.336\*\*\* | 0.607\*\*\* | 0.651\*\* |
|  | (0.027) | (0.042) | -0.226 | (0.261) |
| age\_child | 0.371\*\*\* | 0.127\*\*\* | 0.305\*\*\* | 0.118\*\*\* |
|  | (0.036) | (0.012) | -0.07 | (0.018) |
| age\_migration | -0.179\*\*\* | -0.061\*\*\* | -0.149\*\* | -0.046\*\*\* |
|  | (0.040) | (0.013) | (0.066) | (0.017) |
| No.of observations (N) | 282 | | 282 | |
| Adjusted R-squared | 0.3198 | | 0.3121 | |
| **Diagnostic statistics** |  |  |  | |
| Breush-Pagan LM statistic | | | 7.57\*\*\* | |
| Kleibergen-Paap rk LM statistic | | | 10.804\*\*\* | |
| Hansen J statistic | | | 3.713 | |
| Durbin-Watson statistic | | | 7.742\*\*\* | |
| Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Robust standard errors in parentheses. | | | | |

|  |  |  |
| --- | --- | --- |
| **Table 7. Transmission coefficients among the migrants, All India** | | |
| **Variables** | **Ordinary Least Square (OLS)** | |
|  | **father-son pairs** | |
|  | *Regression coefficient* | *Correlation coefficient* |
| *parent\_edu* | 0.215\*\*\* | 0.336\*\*\* |
|  | (0.027) | (0.042) |
| *age\_child* | 0.371\*\*\* | 0.127\*\*\* |
|  | (0.036) | (0.012) |
| *age\_migration* | -0.179\*\*\* | -0.061\*\*\* |
|  | (0.040) | (0.013) |
| *Number of observations* | 282 | |
|  | **father-daughter pairs** | |
|  | *Regression coefficient* | *Correlation coefficient* |
| *parent\_edu* | 0.195\*\* | 0.249\*\* |
|  | (0.081) | (0.104) |
| *age\_child* | 0.432\*\*\* | 0.131\*\*\* |
|  | (0.091) | (0.027) |
| *age\_migration* | -0.271\*\*\* | -0.082\*\*\* |
|  | (0.101) | (0.030) |
| *Number of observations* | 66 | |
|  | **mother-son pairs** | |
|  | *Regression coefficient* | *Correlation coefficient* |
| *parent\_edu* | 0.191\*\*\* | 0.282\*\*\* |
|  | (0.042) | (0.062) |
| *age\_child* | 0.427\*\*\* | 0.133\*\*\* |
|  | (0.057) | (0.017) |
| *age\_migration* | -0.217\*\*\* | -0.067\*\*\* |
|  | (0.059) | (0.018) |
| *Number of observations* | 148 | |
|  | **mother-daughter** | |
|  | *Regression coefficient* | *Correlation coefficient* |
| *parent\_edu* | 0.257\*\*\* | 0.342\*\*\* |
|  | (0.002) | (0.103) |
| *age\_child* | 0.447\*\*\* | 0.136\*\*\* |
|  | (0.093) | (0.028) |
| *age\_migration* | -0.319\*\*\* | -0.097\*\*\* |
|  | (0.104) | (0.031) |
| *Number of observations* | 62 | |
| Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Robust standard errors in parentheses. | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 8. Transmission coefficients for father-son pairs** | | |  |  |
|  | **Migrants** | | **Non-migrants** | |
|  | ***OLS*** | **IV-Lewbel** | ***OLS*** | **IV-Lewbel** |
| Regression coefficient | 0.168\*\*\* | 0.627\*\*\* | 0.291\*\*\* | 0.575\*\*\* |
|  | (0.031) | (0.225) | (0.005) | (0.103) |
| Correlation coefficient | 0.140\*\*\* | 0.520\*\*\* | 0.378\*\*\* | 0.749\*\* |
|  | (0.026) | (0.187) | (0.007) | (0.135) |
| No.of observations (N) | 282 | | 16,423 | |
| **Diagnostic statistics** |  |  |  |  |
| Breush-Pagan LM statistic | 7.57\*\*\* | | 77.14\*\*\* | |
| Kleibergen-Paap rk LM statistic | 10.804\*\*\* | | 24.709\*\*\* | |
| Hansen J statistic | 3.713 | | 6.122 | |
| Durbin-Watson statistic | 7.742\*\*\* | | 10.301\*\*\* | |
| Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; Robust standard errors in parentheses. | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 9. Decomposition of intergenerational correlation coefficient for father-son pairs** | | | | | | |
| **Child's education/birth cohort** | **1947-1956** | **1957-1966** | **1967-1976** | **1977-1986** | **1987-96** | **Overall** |
| ***Father: No Education*** | | | | | | |
| C:No education | 0.000 | 0.000 | 0.013 | 0.003 | 0.013 | 0.009 |
| C: Primary | 0.180 | 0.124 | 0.113 | 0.134 | 0.166 | 0.148 |
| C:Middle | 0.118 | 0.084 | 0.068 | 0.077 | 0.050 | 0.068 |
| C:Secondary | 0.001 | 0.017 | 0.023 | 0.009 | -0.055 | -0.018 |
| C:College | 0.000 | -0.028 | -0.048 | -0.046 | -0.027 | -0.036 |
| **Total contribution** | **0.299** | **0.197** | **0.169** | **0.176** | **0.147** | **0.171** |
| ***Father: Primary*** | | | | | | |
| C:No education | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | 0.024 | 0.020 | 0.017 | 0.015 | -0.002 | 0.009 |
| C:Middle | 0.006 | 0.009 | 0.010 | 0.011 | -0.001 | 0.005 |
| C:Secondary | 0.000 | 0.003 | 0.005 | 0.002 | 0.001 | -0.002 |
| C:College | -0.015 | -0.016 | -0.015 | -0.013 | 0.001 | -0.006 |
| **Total contribution** | **0.015** | **0.017** | **0.017** | **0.016** | **-0.001** | **0.007** |
| ***Father: Middle*** | | | | | | |
| C:No education | 0.000 | 0.000 | -0.001 | -0.001 | 0.000 | -0.001 |
| C: Primary | 0.000 | -0.005 | -0.006 | -0.008 | -0.023 | -0.013 |
| C:Middle | -0.007 | -0.007 | -0.005 | -0.009 | -0.011 | -0.011 |
| C:Secondary | 0.000 | -0.003 | -0.004 | -0.002 | 0.024 | 0.006 |
| C:College | 0.018 | 0.025 | 0.017 | 0.020 | 0.021 | 0.022 |
| **Total contribution** | **0.010** | **0.009** | **0.001** | **0.000** | **0.010** | **0.003** |
| ***Father: Secondary*** | | | | | | |
| C:No education | 0.000 | 0.000 | -0.001 | -0.001 | -0.001 | -0.001 |
| C: Primary | 0.000 | -0.017 | -0.009 | -0.009 | -0.018 | -0.012 |
| C:Middle | -0.008 | -0.002 | -0.011 | -0.013 | -0.011 | -0.013 |
| C:Secondary | -0.003 | -0.009 | -0.014 | -0.007 | 0.051 | 0.015 |
| C:College | 0.119 | 0.140 | 0.111 | 0.126 | 0.099 | 0.121 |
| **Total contribution** | **0.108** | **0.112** | **0.075** | **0.097** | **0.120** | **0.111** |
| ***Father: College*** | | | | | | |
| C:No education | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | 0.000 | -0.008 | -0.002 | -0.001 | -0.001 | -0.001 |
| C:Middle | 0.000 | 0.000 | -0.004 | -0.003 | -0.001 | -0.002 |
| C:Secondary | -0.001 | -0.004 | -0.007 | -0.002 | 0.013 | 0.005 |
| C:College | 0.189 | 0.161 | 0.175 | 0.149 | 0.087 | 0.134 |
| **Total contribution** | **0.189** | **0.150** | **0.162** | **0.143** | **0.097** | **0.135** |
| **Correlation Coefficient** | **0.621** | **0.474** | **0.413** | **0.431** | **0.373** | **0.427** |
| Note: No education: 0 years; Primary: 1-5 years; Middle: 6-8 years; Secondary: 9-12 years; and College: 13-16 years. | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 10. Decomposition of intergenerational correlation coefficient for father-son pairs in migrant households** | | | | | | |
| **Child's education/birth cohort** | **1947-1956** | **1957-1966** | **1967-1976** | **1977-1986** | **1987-96** | **Overall** |
| ***Father: No Education*** | | | | | | |
| C:No education | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | - | - | 0.000 | 0.000 | 0.080 | 0.056 |
| C:Middle | - | - | 0.000 | 0.074 | 0.079 | 0.074 |
| C:Secondary | - | - | 0.221 | 0.000 | 0.033 | 0.054 |
| C:College | - | - | -0.049 | -0.058 | -0.018 | -0.047 |
| **Total contribution** | - | - | **0.172** | **0.016** | **0.174** | **0.137** |
| ***Father: Primary*** | | | | | | |
| C:No education | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | - | - | 0.000 | 0.000 | 0.011 | 0.008 |
| C:Middle | - | - | 0.070 | 0.000 | 0.023 | 0.024 |
| C:Secondary | - | - | 0.089 | 0.121 | 0.039 | 0.070 |
| C:College | - | - | -0.039 | -0.050 | -0.008 | -0.035 |
| **Total contribution** | - | - | **0.120** | **0.071** | **0.065** | **0.068** |
| ***Father: Middle*** | | | | | | |
| C:No education | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | - | - | 0.000 | 0.000 | 0.004 | 0.005 |
| C:Middle | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C:Secondary | - | - | 0.000 | 0.067 | 0.005 | 0.014 |
| C:College | - | - | -0.004 | -0.011 | -0.009 | -0.011 |
| **Total contribution** | - | - | **-0.004** | **0.056** | **0.000** | **0.008** |
| ***Father: Secondary*** | | | | | | |
| C:No education | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | - | - | 0.000 | 0.000 | -0.004 | -0.002 |
| C:Middle | - | - | 0.000 | 0.000 | -0.008 | -0.004 |
| C:Secondary | - | - | -0.021 | -0.001 | -0.026 | -0.026 |
| C:College | - | - | 0.047 | 0.001 | 0.064 | 0.045 |
| **Total contribution** | - | - | **0.026** | **0.000** | **0.027** | **0.012** |
| ***Father: College*** | | | | | | |
| C:No education | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C: Primary | - | - | 0.000 | 0.000 | 0.000 | 0.000 |
| C:Middle | - | - | 0.000 | 0.000 | -0.019 | -0.014 |
| C:Secondary | - | - | 0.000 | -0.025 | -0.045 | -0.056 |
| C:College | - | - | 0.193 | 0.181 | 0.172 | 0.216 |
| **Total contribution** | - | - | **0.193** | **0.157** | **0.108** | **0.146** |
| **Correlation Coefficient** | **-** | **-** | **0.507** | **0.304** | **0.373** | **0.372** |

Note: No education: 0 years; Primary: 1-5 years; Middle: 6-8 years; Secondary: 9-12 years; and College: 13-16 years.

**References**

Ackerson, L.K. and Subramanian, S.V. (2008), “State gender inequality, socioeconomic status and intimate partner violence (IPV) in India: a multilevel analysis”, *Australian Journal of Social Issues*, Vol. 43 No. 1, p. 81.

Ammermueller, A. (2007), “Poor Background or Low Returns? Why Immigrant Students in Germany Perform so Poorly in the Programme for International Student Assessment”, *Education Economics*, Vol. 15 No. 2, pp. 215–230.

Arora, R.U. (2012), “Gender Inequality, Economic Development, and Globalization: A State Level Analysis of India”, *The Journal of Developing Areas*, Vol. 46 No. 1, pp. 147–164.

Baker, D.P. and Stevenson, D.L. (1986), “Mothers’ Strategies for Children’s School Achievement: Managing the Transition to High School”, *Sociology of Education*, Vol. 59 No. 3, pp. 156–166.

Baum, C.F., Schaffer, M.E., Stillman, S. and others. (2003), “Instrumental variables and GMM: Estimation and testing”, *Stata Journal*, Vol. 3 No. 1, pp. 1–31.

Becker, G.S. and Tomes, N. (1986), “Human Capital and the Rise and Fall of Families”, *Journal of Labor Economics*, Vol. 4 No. 3, Part 2, pp. S1–S39.

Behrman, J.R. (1988), “Intrahousehold Allocation of Nutrients in Rural India: Are Boys Favored? Do Parents Exhibit Inequality Aversion?”, *Oxford Economic Papers*, Vol. 40 No. 1, pp. 32–54.

Bhattacharya, P.C. (2006), “Economic Development, Gender Inequality, and Demographic Outcomes: Evidence from India”, *Population and Development Review*, Vol. 32 No. 2, pp. 263–292.

Björklund, A. and Jäntti, M. (1997), “Intergenerational income mobility in Sweden compared to the United States”, *The American Economic Review*, Vol. 87 No. 5, pp. 1009–1018.

Black, S.E. and Devereux, P.J. (2011), “Recent Developments in Intergenerational Mobility”, in Ashenfelter, D.C. and O. (Ed.), *Handbook of Labor Economics*, Vol. 4, Part B, Elsevier, pp. 1487–1541.

Borjas, G.J. (1992), “Ethnic Capital and Intergenerational Mobility”, *The Quarterly Journal of Economics*, Vol. 107 No. 1, pp. 123–150.

Borooah, V.K. (2004), “Gender bias among children in India in their diet and immunisation against disease”, *Social Science & Medicine*, Vol. 58 No. 9, pp. 1719–1731.

Causa, O. and Johansson, A. (2010), “Intergenerational Social Mobility in OECD Countries”, *OECD Journal: Economic Studies*, Vol. 2010 No. 1, pp. 1–44.

Checchi, D., Fiorio, C.V. and Leonardi, M. (2013), “Intergenerational persistence of educational attainment in Italy”, *Economics Letters*, Vol. 118 No. 1, pp. 229–232.

Checchi, D. and Flabbi, L. (2007), *Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks*, IZA Discussion Paper No. 2876, Institute for the Study of Labor (IZA), available at: https://ideas.repec.org/p/iza/izadps/dp2876.html (accessed 11 March 2017).

Craig, L. (2006), “Parental education, time in paid work and time with children: an Australian time-diary analysis”, *The British Journal of Sociology*, Vol. 57 No. 4, pp. 553–575.

Currie, J. and Moretti, E. (2003), “Mother’s Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings”, *The Quarterly Journal of Economics*, Vol. 118 No. 4, pp. 1495–1532.

Daude, C. (2011), “Ascendance by descendants? On intergenerational education mobility in Latin America”, *OECD Development Centre Working Papers*, No. 297, p. 1.

Dearden, L., Machin, S. and Reed, H. (1997), “Intergenerational mobility in Britain”, *The Economic Journal*, pp. 47–66.

Dunn, D. (1993), “Gender inequality in education and employment in the scheduled castes and tribes of India”, *Population Research and Policy Review*, Vol. 12 No. 1, pp. 53–70.

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

Farré, L. and Vella, F. (2013), “The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labour Force Participation”, *Economica*, Vol. 80 No. 318, pp. 219–247.

Guryan, J., Hurst, E. and Kearney, M. (2008), “Parental Education and Parental Time with Children”, *The Journal of Economic Perspectives*, Vol. 22 No. 3, p. 23.

Heineck, G. and Riphahn, R.T. (2009), “Intergenerational Transmission of Educational Attainment in Germany — The Last Five Decades”, *Jahrbücher Für Nationalökonomie Und Statistik / Journal of Economics and Statistics*, Vol. 229 No. 1, pp. 36–60.

Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Verashchagina, A. (2007), “Intergenerational economic mobility around the world”, *The BE Journal of Economic Analysis & Policy*, Vol. 7 No. 2, pp. 1–46.

Hnatkovska, V., Lahiri, A. and Paul, S.B. (2013), “Breaking the Caste Barrier Intergenerational Mobility in India”, *Journal of Human Resources*, Vol. 48 No. 2, pp. 435–473.

Jacobs, J.A. (1996), “Gender Inequality and Higher Education”, *Annual Review of Sociology*, Vol. 22 No. 1, pp. 153–185.

Jalan, J. and Murgai, R. (2008), “Intergenerational Mobility in Education in India’. Paper presented at the 5th annual conference of the Indian Statistical Institute, Delhi, December”.

Kishor, S. (1993), “‘May God Give Sons to All’: Gender and Child Mortality in India”, *American Sociological Review*, Vol. 58 No. 2, pp. 247–265.

Lewbel, A. (2012), “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models”, *Journal of Business & Economic Statistics*, Vol. 30 No. 1, pp. 67–80.

Maitra, P. and Sharma, A. (2009), “Parents and children: Education across generations in India”, *5th Annual Conference on Economic Growth and Development, ISI Delhi, Delhi*, available at: http://www.isid.ac.in/~pu/conference/dec\_09\_conf/Papers/PushkarMaitra.pdf (accessed 12 March 2017).

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*, Vol. 4 No. 4, pp. 463–494.

Murthi, M., Guio, A.-C. and Drèze, J. (1995), “Mortality, Fertility, and Gender Bias in India: A District-Level Analysis”, *Population and Development Review*, Vol. 21 No. 4, pp. 745–782.

Oreopoulos, P., Page, M.E. and Stevens, A.H. (2006), “The Intergenerational Effects of Compulsory Schooling”, *Journal of Labor Economics*, Vol. 24 No. 4, pp. 729–760.

Schneeweis, N. (2011), “Educational institutions and the integration of migrants”, *Journal of Population Economics*, Vol. 24 No. 4, pp. 1281–1308.

Schnepf, S.V. (2002), *A Sorting Hat That Fails? The Transition from Primary to Secondary School in Germany, UNICEF Innocenti Research Centre, Florenz*.

Ware, H. (1984), “Effects of Maternal Education, Women’s Roles, and Child Care on Child Mortality”, *Population and Development Review*, Vol. 10, pp. 191–214.