MPRA
Munich Personal RePEc Archive

Gender, Geography and Generations: Intergenerational Educational Mobility in Post-reform India

Emran, M. Shahe and Shilpi, Forhad

3 September 2014

Online at https://mpra.ub.uni-muenchen.de/58316/
MPRA Paper No. 58316, posted 09 Sep 2014 05:32 UTC

Gender, Geography and Generations:
Intergenerational Educational Mobility in Post-reform India

M. Shahe Emran[1]

IPD, Columbia University

Forhad Shilpi

DECRG, World Bank

ABSTRACT

The existing studies report substantial improvements in educational mobility in post-reform India using intergenerational regression coefficient (IGRC) across age cohorts in a cross-section survey. In contrast, our estimates of sibling (SC) and intergenerational (IGC) correlations for the same age cohort from two surveys show strong persistence, stronger than in Latin America, which remained largely unchanged from 1991/92-2006. Only the women in urban areas experienced substantial improvements, with the lower caste urban women benefitting the most. As measures of mobility, IGC and SC are more informative and robust than IGRC, and the widely accepted conclusions based on IGRC alone may be misleading.

Key Words: Intergenerational Mobility, Education, Equality of Opportunity, Sibling Correlation, Intergenerational Correlation, Economic Liberalization, Rural-Urban Inequality, Gender Gap, India

JEL Codes: O12, J62

[1] We would like to thank Pranab Bardhan, Mtthew Lindquist, Anders Bjorklund, Will Martin, Rukmini Banerjee and Giovanna Prennushi and the participants at MIE conference on intergenerational mobility at the Becker-Friedman Institute at the University of Chicago in May 2012 and IGC-ISI conference at Delhi in July 2012 for helpful comments on earlier drafts. The views and conclusions are authors' own and should not be attributed to the institutions they are affiliated with. Email for correspondence: shahe.emran@gmail.com (M. Shahe Emran) and fshilpi@worldbank.org (Forhad Shilpi).

Introduction

The increasing inequality in income distribution at a time of considerable economic growth during the last couple of decades has rekindled interests in intergenerational mobility in both developed and developing countries.[2] Following wide ranging economic liberalization in the early 1990s, India experienced sustained high economic growth; per capita GDP grew at a 4 percent rate over the two decades after liberalization. The evidence indicates that while growth led to a significant poverty reduction, it was also associated with a rise in inequality (World Bank (2011)).[3] There is increasing concern that the benefits of economic growth were not shared broadly, and remained especially concentrated in urban areas, thus widening the rural-urban gap (Bardhan (2007, 2010), Dreze and Sen (2011), Basu (2008), Prasad (2012)). The estimates of top incomes by Banerjee and Piketty (2005) show that the share of top 0.01, 0.1, and 1 percent in total income has increased substantially from a trough in the mid1980s, and that this increase coincided with the move away from 'Socialist' to more market oriented economic policies. According to their estimates, in 1999-2000, per capita income gap between the 99th and 99.5th percentiles was four times as large as the gap between the median and the 95th percentile.[4] Dreze and Sen (2011) argue that Indian economic reform has been an "unprecedented success" in terms of economic growth, but an "extraordinary failure" in terms of improvements in the living standards of general people and social indicators.

However, an important question is whether the observed increase in cross-sectional inequality is a natural outcome of efficient incentive structure in a liberalized and market oriented economy that rewards hard work and entrepreneurial risk taking, or is it primarily due to inequality of opportunity arising from differential access, for example, to education, markets and political power?[5] The rise in

---

[2] Among the developing countries, China and India are two prominent examples where impressive economic growth has been accompanied by an increase in inequality. For a discussion on rising inequality in Asia, see Jushong and Kanbur (2012). The recent decline in intergenerational mobility in USA and UK has also attracted a lot of attention; see, for example, Deparle in New York Times (January 4, 2012) and Mazumder (2012) on USA, and Dearden et al. (1997) and Blanden et al. (2005) on UK.

[3] For evidence on rising inequality in India after 1991, see Ravallion (2000), Deaton and Dreze (2002), Sen and Himanshu (2004). A recent survey of the available evidence shows that consumption inequality has increased slightly, but the income inequality in India is much higher than what is usually thought of (close to Brazil) (World Bank (2011)). It is now widely appreciated that the available estimates of consumption and income inequality may be significantly biased downward, because the household surveys fail to cover the top income households.

[4] The common perception about a significant increase in inequality is reinforced by spectacular conspicuous consumption by the super-rich: Mukesh Ambani, the chairman of Reliance Industries in India owns and lives in the first billion dollar house in the world (Woolsey, M, Forbes.com, April 30, 2008), and in the mega wedding of two sons of Subrata Roy, the 'chief guardian' of Sahara Group, $ 250,000 was spent on candles alone (Srivastava, S, BBC online, February 11, 2004)!

[5] The observed income inequality may also reflect inequality in endowments, especially in land in the context of rural areas. High land inequality can hinder economic development as it restricts access to credit for landless and

cross-sectional inequality becomes a serious concern especially when it is a result of inequality of opportunity, i.e., the inability of children born in poorer families and disadvantaged social groups such as low castes to move beyond their parents' position in economic ladder by their own effort and choices.[6] The goal of this paper is to analyze the trends in and levels and patterns of educational mobility over a period of almost a decade and a half after the liberalization in 1991 (1993-2006), with a special focus on possible gender and spatial differences (rural-urban and village/neighborhood fixed effect). Education is used as an indicator of economic status in the absence of suitable data on permanent income.[7] The role of education may be especially important in post-reform India where growth has been concentrated in skill intensive sectors: the software industry and call centers being iconic examples (Kochhar et al. (2006), Bardhan (2010), Kotwal et al. (2011)).[8] The focus on education is also appropriate from a policy perspective, because it is among the few policy levers that enjoy wide popular support in many countries.

To understand the educational mobility in post-reform India, we use two related measures: (i) sibling correlation (SC) and (ii) intergenerational correlation (IGC) in educational attainment. In contrast, most of the available evidence on intergenerational educational mobility in India is based on variants of intergenerational regression coefficient (IGRC). A second important difference is that while we analyze the same age cohort (16-27 years) from surveys in 1992 and 2006, most of existing studies rely on different age cohorts from a single cross-section survey. Our results show that the choice of the measure and data matters for the substantive conclusions: while the existing studies conclude that educational mobility has improved substantially in India in recent decades, the evidence presented in this paper paints a more sober picture: educational persistence is very high in India and it has remained largely unchanged in post reform period. A potentially important issue in understanding intergenerational educational mobility using household survey data is possible biases due to the coresidency restriction standard in

land poor households. Note also that even if the observed high inequality is an efficient outcome in terms of resource allocation, a society might find it unacceptable on ethical grounds.

[6] Higher inequality of opportunity is likely to lead to a higher cross-sectional inequality (Atkinson (1981)). Many observers believe that inequality in India reflects inequality in opportunity. For example, Basu (2008) comments that "A certain amount of inequality may be essential to mitigate poverty….But the extent of Inequality in India seems to be well above that".

[7] Reliable data on children's and parents' income over the life cycle are not available in a developing country such as India. As emphasized in the recent literature, good quality income data over a number of years at appropriate phase of the lifecycle are needed to tackle the attenuation bias in the estimated intergenerational correlation in income (Solon (1999), Mazumder (2003)). The analysis of intergenerational persistence in income in India is also complicated by the fact that a majority of population especially during parent's generation were engaged in family farming as self-employed workers making it difficult to attribute income to individual members. For a discussion on the limitations of income data in household surveys in developing countries, see Deaton (1997). Another important problem relates to the fact noted before that the household surveys do not adequately represent the top end of the income distribution, and thus estimates based on income will tend to underestimate the inequality.

[8] This is in contrast to the Chinese experience where growth has been dominated by agriculture and labor intensive manufacturing. Bardhan (2010) and Datt and Ravallion (2010) emphasize low and unequal human capital as an important constraint on poverty reduction in India.

household surveys. Recent analysis shows that the bias due to coresidency restriction in intergenerational regression coefficient (IGRC), the most widely used measure of intergenerational persistence in the literature, is severe, it can be as high as 52 percent (see Emran and Shilpi (2014)). Using data from India and Bangladesh, Emran and Shilpi (2014) show that, in contrast, the bias in the normalized measures such as IGC and Sibling correlation is very low (less than 2 percent difference in IGC estimates in most cases), and the conclusions based on IGC and Sibling Correlations from coresident samples are reliable and robust. For an extended discussion and additional evidence on this issue, please see section (3.1) below.

An important finding from the sibling studies in developed countries is that gender or geographic location (as measured by neighborhood effect) does not exert any significant influence on educational or income mobility of children (Solon (1999), Bjorklund and Salvanes (2010)). Are gender and geography also largely irrelevant for educational mobility of children in developing countries? One can argue that the role of gender and geography is likely to be much more prominent in a developing country such as India, because gender bias against women is more common and stronger, geographic mobility is lower, and many areas (especially rural) are not integrated with the urban growth centers because of underdeveloped transport infrastructure.[9] One might also worry that the disadvantaged social groups (e.g., low caste) in India may not be able to take advantage of the opportunities offered by economic reform and globalization, and there might be complex interactions among gender, geography and social identity.[10]

The data used in this paper come from the1992/93 and 2006 rounds of the National Family Health Survey (NFHS) in India. The first period of our sample nearly overlaps with the timing of economic liberalization (1991-1992), and thus provides a plausible benchmark for understanding the nature of mobility over a period of 15 years after liberalization. We focus on the role of family background in educational attainment of the youth (16 to 27 year olds at the time of the survey) who constitute the bulk of the new entrants into the labor market.[11] Thus we compare the estimated effects of family background on the educational attainment of the 'youth of 1991' (i.e., the 16-27 years age cohorts at the time of the 1991-92 survey) to that of the 'youth of 2006' (i.e., the 16-27 years age cohorts at the time of 2006

[9] There is evidence that geographic location may be important for economic opportunities faced by households in developing countries. For example, Jalan and Ravallion (1999) show that there are geographic poverty traps in China. Emran and Hou (2013) find that better access to markets increases household consumption in rural China in a significant way. They also find that the effects of domestic market centers are much larger than that of international market access. Emran and Shilpi (2012) find that pattern of agricultural specialization in Nepal depends on distance from the urban markets in a non-linear fashion.

[10] As Luke and Munshi (2011) observe: "(W)hen new opportunities presented themselves under British colonial rule, the upper castes were quick to gain access to western education and with it coveted administrative and professional jobs. The concern is that the higher castes might once again seize the new opportunities that are made available by globalization, widening the existing caste-gap even further."

[11]

survey). To examine the spatial aspects, the empirical analysis is done separately for families residing in rural and urban areas, and also estimates of the neighborhood fixed effects are provided. To discern any possible gender bias, we implement the empirical analysis separately for male and female samples. Following Bjorklund et al. (2010), we use the mixed effects model to estimate the sibling correlation. An advantage of this approach is that both the family and community level covariates can be included in the analysis to examine their relative influence on sibling correlation (Mazumder (2008), Bjorklund et al. (2011)). We examine the influence of two types of covariates on sibling and intergenerational correlations: the first relates to caste and religion of the household which have been identified as important determinants of educational attainment in India, and the second relates to the geographic location as measured by neighborhood fixed effect.[12]

Our estimates of sibling and intergenerational correlations suggest no significant change in educational mobility for a large proportion of the relevant population in India from 1992/93 to 2006. Sibling (and intergenerational) correlations in our full sample have declined only marginally from 0.64 (0.57) in 1992/93 to 0.62 (0.54) in 2006 respectively.[13] The estimates indicate that a decade and a half after the economic liberalization in 1991, the absolute magnitudes of sibling and intergenerational correlations in India in 2006 are still very large, larger than the available estimates for the Latin American countries (for sibling correlations) and Asian countries (for intergenerational correlations).[14] The aggregate picture of stagnation, however, hides important gender and spatial differences. While the evidence indicates that the sibling correlation among men (brothers) has remained effectively unchanged (it increased slightly from 0.614 in 1993 to 0.624 in 2006), it experienced a moderate decline for women, (sisters) from 0.780 to 0.696. Geographic location is important, both in 1992/93 and 2006; the neighborhood effect accounts for about 40 percent of the sibling correlation among women and a third

[12] One might argue that we should also include parental income as a control to isolate the income effect. We believe that it could be potentially interesting when appropriate data on parental *permanent* income are available. Unfortunately, in most of the survey data sets such as NSS, NFHS, IHDS used to study India, the data on income correspond to single year or a few years at long time intervals. This makes it impossible to estimate the permanent income with any measure of confidence. In this context, education and occupation are much better indicators of permanent income (Bjorklund, 2011).

[13] Note that a formal test of equality of the estimates in 1993 and 2006 rejects the null because of very small standard errors due to the large sample sizes (number of observations is 34000 in 1993 and 38000 in 2006). However, statistical precision is largely irrelevant here, because the difference in the numerical magnitude of the estimates between 1992/93 and 2006 is very small in most of the cases, suggesting the lack of any substantial change in intergenerational mobility over a period of almost a decade and a half of impressive economic growth.

[14] According to the estimates in Hertz et al. (2007), the estimates of intergenerational correlation in education in many Latin American countries are higher than our estimate for India. One may thus conclude that whether India is less or more mobile than Latin America cannot be judged by our estimates; it is less mobile according to sibling correlation, but more mobile according to intergenerational correlation. This conclusion, however, ignores that fact that sibling correlation is a much broader measure and thus one should rely on it for the ranking of different countries (see equation (6) below). The fact that intergenerational correlation is higher only implies that the direct role played by parents' education is more important in Latin America, but the over-all level of educational immobility is higher in India.

among men. In terms of geographic pattern, we find that sibling correlation remained essentially unchanged in rural areas, but declined marginally in urban areas. Perhaps the most interesting trends and patterns emerge when we partition the data using both gender and geography. The sibling correlations among men (brothers) in rural areas have increased a bit, but the correlation has in fact declined marginally in the urban areas. In contrast, the sibling correlations among women (sisters) registered a decline irrespective of geographic partitioning of the data. However, geography matters for women also, the women in urban areas experienced much more substantial decline in sibling correlations. As a result, the gender gap in sibling correlation has virtually disappeared in urban areas. Despite moderate improvements in mobility among women, the gender gap in rural areas remains substantial. We also find that among the urban women, it is the lower caste women who experienced the largest decline in the sibling correlation. The evidence on improvements in educational mobility of women in India is similar to the available evidence on China and Malaysia (see Emran and Sun (2011) on China and Lillard and Willis (1994) on Malaysia).[15] The broad trends in and patterns of educational persistence as measured by sibling correlations and discussed above are also observed in the estimates of intergenerational correlations in education between parents and children. The importance of geographic location for educational mobility in India is also evident from the strong role of the 'neighborhood effect' in explaining the sibling correlations. This is in sharp contrast to the case of developed countries where there is little or no evidence of a significant neighborhood effect.

The rest of the paper is organized as follows. The conceptual framework underpinning empirical work is described in section 2. Data and empirical strategy are elaborated in section 3. Section 4 organized in different subsections presents the main empirical results, and section 5 reports as set of robustness checks. Some preliminary conjectures for explaining the observed trends in and patterns of educational mobility in post-reform India are offered in section 6. The paper concludes with a summary of the findings.

Related Literature

The literature on intergenerational economic mobility in developed countries is large, most of which focuses on intergenerational correlation between parents' and children's incomes (for reviews, see Solon (1999, 2002), Black et al. (2010)).[16] However, economic analysis of intergenerational mobility in the context of developing and transitional countries remains a largely unexplored area of research. The

---

[15] The positive evidence on women may seem puzzling given the fact that son preference is prevalent in all three countries. We provide a set of possible explanations for the observed trend later in the paper.

[16] See, among others, Arrow et al. (2000), Dearden et al. (1997), Mulligan (1999), Solon (1999, 2002), Birdsall and Graham (1999), Fields et al. (2005), Bowles et al. (2005), Blanden et al. (2005), World Development Report (2006), Mazumder (2003), Hertz (2005), Bjorklund et al. (2006), and Lee and Solon (2009).

available contributions on developing countries focus on intergenerational regression coefficient (IGRC), but do not estimate intergenerational (IGC) or sibling correlations (SC); see, for example, Jalan and Murgai (2008) and Maitra and Sharma (2010) on India, Lillard and Willis (1994) on Malaysia, Emran and Shilpi (2011) on Nepal and Vietnam, and Emran and Sun (2011) on China. The only study known to us that uses sibling correlation in the context of developing countries is Dahan and Gaviria (2001) who provide estimates of sibling correlations in education for 16 Latin American countries.[17] They find that El Salvador, Mexico, Colombia and Ecuador are the least mobile (high sibling correlation) countries, with sibling correlation explaining almost 60 percent of the variation in educational outcomes.

Most of the existing studies on intergenerational educational mobility in developing countries use IGRC and IGC to provide estimates of persistence between parents' and children's educational attainments.[18] However, it has been increasingly appreciated in the literature that the IGRC and IGC are partial and incomplete measures at best, and the influence of family background on children extends much beyond what is implied by parental characteristics (Corcoran et al. (1976), Mazumder (2008), Bjorklund et al. (2010)). There is now a substantial literature in economics that uses sibling correlation in economic outcomes as an omnibus measure of immobility (for early contributions, see among others, Corcoran et al. (1976, 1990), Solon et al. (1991); for a recent discussion on the advantages of sibling correlation for understanding intergenerational mobility, see Bjorklund and Jantti (2012)).[19] Sibling correlation provides a summary measure of all the common family and community background factors shared by siblings, but not chosen by children themselves. The available evidence in the context of developed countries shows that the factors common to siblings explain from 40 to 65 percent of variation in educational outcomes (Bjorklund and Salvanes (2010)). In contrast, the intergenerational correlation between parents and children– the traditional measure of intergenerational persistence -- explains only 9 to 21 percent of variations in children's educational outcome. To the best of our knowledge, there is no study in the literature on developing countries that exploits estimates of both sibling and intergenerational correlations to trace out the levels, trends in and patterns of intergenerational mobility.[20]

[17] It is important to appreciate that sibling correlation as a measure of immobility is equally valid for developed and developing countries.

[18] For a survey of this literature, see Black and Devereux (2010), Bjorklund and Salvanes (2010) for developed countries, and Hertz et al. (2009) for both developed and developing countries.

[19] We discuss in details later the relationship between intergenerational and sibling correlations and also why sibling correlation provides a comprehensive and intuitive measure of immobility. Please see pp. 10-11 below.

[20] The recent literature on intergenerational mobility has emphasized the importance of robustness checks. There are two dimensions to this robustness issue: (i) alternative indicators of economic status and (ii) alternative measures of mobility for a given indicator of mobility. While some papers have analyzed multiple indicators of economic status using only a single measure of mobility, we concentrate on one salient indicator of economic status, i.e., education, and use two alternative measures of mobility to reach robust conclusions.

(2) Conceptual Framework

(2.1) Sibling Correlation (SC)

For the estimation and interpretation of sibling correlations, we adopt a conceptual framework that has been the workhorse in the empirical literature on sibling correlations (see, Solon et al. (1991), Solon (1999), Bjorklund et al. (2002), Bjorklund and Lindquist (2010), Bjorklund and Salvanes (2010), Mazumder (2008) and (2011)). Following Solon (1999) and Bjorklund et al. (2010), we begin with a simple model of children's educational attainment:

$$S_{ij} = \mu + a_i + b_{ij} \qquad (1)$$

Where S_{ij} is the years of schooling of sibling j in family i, μ is the population mean, a_i is a family component which is common to all siblings in family i and b_{ij} is the individual specific component for sibling j which captures j's deviation from the family component. Conceptually, sibling correlation is a measure of the variance in the household specific component a_i across different households relative to the variance in children's educational attainment. It thus nets out the population mean μ which captures the factors common to all households that determine the average educational attainment in a society.

Assuming that the components a_i and b_{ij} are independent, the variance of S_{ij} can be expressed as the sum of variances of the family and individual components as:

$$\sigma^2{}_s = \sigma^2{}_a + \sigma^2{}_b \qquad (2)$$

The sibling correlation in education then can be expressed as:

$$\rho_s = \frac{\sigma^2{}_a}{(\sigma^2{}_a + \sigma^2{}_b)} \qquad (3)$$

Thus ρ_s is the share of variance of children's education that can be attributed to common family background, and it is equal to the correlation in educational attainment among the siblings in a randomly selected household from the population. This is why this measure is called sibling correlation (Bjorklund et al. (2010)).

Sibling correlation can be thought of as a summary statistic of the importance of common family and community effects which include anything and everything shared by the siblings. This is a measure of immobility because all these factors affecting the educational attainment of the siblings are not chosen by the children themselves, but they 'are born into it'. To appreciate `sibling correlation' as a measure of immobility, it is instructive to consider the implications of credit market imperfections. We first consider the polar case of perfect markets leading to perfect educational mobility, and then contrast it with a society composed of rich and poor households where only the poor households face credit constraint. In a perfectly mobile society (no household faces credit constraint), every child has access to education, and the optimal level of education depends only on individual ability. Family background is irrelevant because marginal benefit and marginal cost of education do not depend on family index i in equation (1) above; every family has access to credit at a given interest rate $r_i = r$ for all i, and the wage rate in the labor market does not depend on family connections, i.e., $w_{ij}(S_{ij}) = w(S)$. In terms of equation (1) above the family component a_i does not play any role in explaining the variance in educational outcome $\sigma^2{}_S$, because after we net out the population mean μ (determined by the common interest rate r and labor market returns w), the correlation between siblings in educational outcome is same as the correlation between two randomly chosen children's education of same age from the population, i.e., equal to zero (consider equation (3) above with $\sigma_a^2 = 0$). The variance in observed educational outcome is thus completely determined by the variance in individual ability (the idiosyncratic component b_{ij}). Now consider credit market imperfections so that the poor parents face a higher interest rate $r_l > r_h$ where subscript l denotes poor (low income) and h denotes rich (high income), possibly because of lack of collaterals. Facing higher costs, poor parents optimally invest less in education of a child compared to a rich household, holding child ability constant.[21] Under the plausible assumption that the distribution of ability does not depend on family background, the *average* education of children in a randomly drawn poor household is now lower than that in a randomly drawn rich household, implying $a_{il} < a_{ih}$. This increases the variance in the family component ($\sigma^2{}_a$) in equation (3), making the sibling correlation larger in magnitude.

One may also find it useful to consider the difference between inequality in *outcomes* and inequality in *opportunities* in this context. Consider the polar case of a society with prefect equality in educational *outcome*, with $S_{ij} = \underline{S} = \mu$, i.e., the education level of every child is same, which is also the population mean, and the variance in educational *outcome* is zero. Also, trivially the variances in both the family and individual components in equation (3) are equal to zero, i.e., $\sigma_a^2 = \sigma_b^2 = 0$. This implies that even with equality in educational *opportunities* across different families ($\sigma_a^2 = 0$), a perfect equality in educational *outcomes* is possible only if every child is endowed with the same ability and chooses to exert exactly the same effort ($\sigma_b^2 = 0$). In such a 'clone society', the sibling correlation as in equation (3) cannot be defined, because both the denominator and numerator in equation (3) above are zero in this case. However, this polar case is of little relevance for the analysis of real world data which are always characterized by non-zero variance in ability, effort, and educational outcomes.

It is useful to distinguish among different types of family and community factors that are commonly experienced by siblings. The family level variables include observable factors such as parental education and occupation as well as unobserved factors such as common genetic traits, parental aspirations, child rearing ability and style, cultural inheritances and interaction among siblings. The

[21] This assumption may, however, be less tenable in the case of higher education, because there is substantial evidence that family environment in early childhood may affect both cognitive and non-cognitive ability of a child significantly.

community effects include factors such as school availability and quality as well as peer effects within the neighborhood. Though sibling correlation captures most of the family background influences, it does not capture all of them. For instance, genetic traits not shared by siblings, differential treatment of siblings and time dependent changes in family and neighborhood factors will show up in the individual component of outcome variance, though they might be part of family background. As a result, the estimate of sibling correlations can be taken as a *lower bound estimate* of the total influence of the common family background on children's education outcome (for a discussion on this point, see Bjorklund and Salvanes (2010)).[22]

(2.2) Intergenerational Correlation (IGC)

In this subsection, we discuss the differences and interrelationships between sibling correlation and intergenerational correlation as measures of the influence of family background and intergenerational persistence in economic outcomes. The standard regression model to estimate intergenerational correlation between parents and children can be written as:

$$S_{ij} = \mu + \beta {S^p}_i + e_{ij} \quad (4)$$

Where ${S^p}_i$ is the parental year of schooling in family i, and β is the *intergenerational regression coefficient*. Because individual component in equation (1) is orthogonal to the family component, one can express the family component as:

$$a_i = \beta {S^p}_i + z_i \quad (5)$$

Where z_i denotes family factors that are orthogonal to parental education. It follows from equation (5) that:

$$\rho_s = \frac{{\sigma^2}_a}{{\sigma^2}_s} = \beta^2 \frac{{\sigma^2}_{sp}}{{\sigma^2}_s} + \frac{{\sigma^2}_z}{{\sigma^2}_s} = (\rho_{IG})^2 + family\ factors\ orthogonal\ to\ parental\ eduction \quad (6)$$

Where ρ_{IG} is the *intergenerational correlation* in education. The above equation is widely known in the literature (see, for example, Solon (1999)). It shows clearly that sibling correlation is a broader measure of the impact of family background than the squared intergenerational correlation. This has important implications, especially when the conclusions based on sibling and intergenerational correlations conflict

[22] We emphasize here that it does not contradict the point made earlier that sibling correlation is a broad measure of immobility. As noted before and explained in detail below, the sibling correlation is composed of two components: one due to parental influence and the other due to common factors experienced by siblings such as schooling and peer effects etc. For an excellent discussion on the advantages of using sibling correlation as a measure of immobility and inequality of opportunity, see Bjorklund and Jantti (2012).

with each other. The fact that sibling correlation is a broader measure of the effects of family background means that one should in general rely on it rather than the intergenerational correlation, especially when there is a conflict. We emphasize here again that the intergenerational correlation parameter (ρ_{IG}) is different from intergenerational regression coefficient (β) used in the existing studies on educational mobility in India such as Jalan and Murgai (2008) and Maitra and Sharma (2010).[23]

(2.3) Estimating Equations

To estimate the sibling correlations, we extend the regression model in equation (1) and specify the following mixed effects model:

$$S_{ij} = \mu + a_i + b_{ij} + Z_{ij}\gamma \qquad (7)$$

Where Z_{ij} is a vector of control variables.

To estimate the intergenerational correlation in education, we augment equation (4) to estimate the following regression specification:

$$S_{ij} = \mu + \beta S^{p}{}_{i} + Z_{ij}\pi + e_{ij} \qquad (8)$$

Equations (7) and (8) can be estimated as soon as Z_{ij} vector is specified. Following Bjorklund et al. (2010) and Mazumder (2008, 2011), we take a sequential approach in introducing variables to Z_{ij} vector. As is standard in this literature, the benchmark model includes age as a control.[24] This is to ensure that the cohort effects in mean educational attainment do not contaminate our analysis. A focus here is on two types of explanatory variables: caste and religion, and neighborhood fixed effects. Evidence from India suggests that educational outcomes vary systematically across different caste and religion groups. We add a village/neighborhood level fixed effect as a part of Z_{ij} to capture any common community level factors faced by the children growing up in the same locality. A comparison of sibling correlations estimated using alternative specifications can shed light on the importance of caste and religion as well as geographic location as captured by the neighborhood effect.[25] As noted in earlier studies (summarized in Bjorklund and Salvanes (2010)), if households are sorted across neighborhoods according to their attributes (well-off families living in better neighborhoods), then the estimate of neighborhood effect is

[23] The estimates of Hnatkovoska et al. (2011) also do not adjust for changing variance in education. Thus their estimates are more like IGRC and are likely to be subject to similar biases arising from a co-resident sample, among other things.

[24] We also add a gender dummy when estimating the pooled sample.

[25] This approach follows Mazmuder (2008, 2011) and Bjorklund et al. (2010). The basic idea is that if the estimated sibling correlation is primarily driven by factors such as neighborhood effects, caste and religion, then the estimate would decline significantly once these factors are included in the regression.

biased upward. So the comparison will provide an upper bound estimate of neighborhood effect. In contrast, the estimate of intergenerational correlation can be biased upward (due to correlation in genetic traits) or downward (due to measurement error).

Also, it is important to appreciate that the conclusions regarding the role played by different factors can depend on the order in which different variables are added to a specification, a point discussed in the literature on "step-wise regressions" and recently emphasized by Gelbach (2009). To ensure that the conclusions in this paper are not affected by the "order of addition" problem, we checked the robustness of the results with respect to alternative ordering of the control variables. The evidence shows that ordering matters very little in this context, and all the conclusions reached here remain unchanged if we reverse the order of neighborhood fixed effects and caste and religion as controls.

We compare the estimated sibling correlations with the estimates of intergenerational correlations and neighborhood effects. This allows us to deduce the extent of sibling correlations that can be accounted for by the parent-child link and the neighborhood effect. The part of sibling correlations that remains unaccounted for by these two factors is mainly due to common family environment such as family structure (e.g. divorced/separated parents) and parental skills and patience in child rearing etc. Note that if the strong sibling correlation observed in the data is due mainly to intergenerational correlations in education and common neighborhood effects, then it indicates higher inequality in opportunities than if it were due to parents' child rearing skills. [26]

(2.4) Estimation Approaches

The intergenerational correlation can be estimated by first using OLS regression for equation (8) to estimate the intergenerational regression coefficient β and then using the following formula for intergenerational correlation that adjusts for the changes in the variance in education: $\rho_{IG} = \beta \frac{\sigma_{sp}}{\sigma_s}$. For the estimation of sibling correlation in equation (7), the family and individual components need to be estimated. The available literature on sibling correlations relies on two alternative estimation methods. Mazumder (2006, 2011) uses the Restricted Maximum Likelihood (REML) method which has better small sample properties under the normality assumption. Bjorklund et al. (2010) instead uses Stata's GLLAMM to estimate the family and individual variance components $\sigma^2{}_a$ and $\sigma^2{}_b$ in the mixed effects

[26] Bjorklund, Lindahl and Lindquist (2010) find a sibling correlation of around 0.21 for Sweden. Almost 70 percent of sibling correlation in Sweden can be explained by parental involvement in school work and mother's patience (willingness to postpone benefits into the future and propensity to plan ahead). Intergenerational correlations in education as well as neighborhood effects are found to have small influence on sibling correlations. Sweden however is characterized by nearly universal access to quality education, generous child care assistance and low income inequality.

model and the NLCOM command to estimate the sibling correlation ρ_s.[27] A limitation of this procedure is that its small sample properties are not well understood. One practical difficulty in implementing the REML in our application is that we include more than 3000 dummies for the neighborhood fixed effects which creates convergence problems. The large number of fixed effects can be handled easily in the two step procedure suggested by Bjorklund et al (2010) to estimate the individual and family variance components. The estimates of variance components used in this paper are thus from the two-step procedure. In the first stage, the residual from the OLS regression of children's education on the set of controls Z_{ij} is retrieved. In the second stage, the residual is passed on to the GLLAMM to get estimates of ${\sigma^2}_a$ and ${\sigma^2}_b$. The estimation of sibling correlation is implemented in the final stage, using Stata's NLCOM command.[28] We utilized this three step procedure to estimate the sibling correlation. However, we note here that in all the specifications without the neighborhood fixed effects, the estimates of sibling correlation from the REML are slightly larger than the estimates from the three stage procedure discussed above. The estimates reported in this paper can be taken as conservative estimates. The estimates using Restricted Maximum Likelihood are available from the authors.[29]

(3) Data and Empirical Issues

The data for our analysis come from the National Family Health Survey (NFHS), 1992/93 and 2006. The NFHS is a large-scale and nationally representative survey of nearly all of Indian states. The main target group for this survey is women in their reproductive years. While both surveys followed similar sampling methodology, the surveys differ somewhat in terms of sample size and questionnaires. The NFHS 2006 used three separate questionnaires to interview 109,041 households, 124,385 unmarried and ever married women between 15 to 49 years of age and 74,369 unmarried and ever married men in the age group 15-54 years. The NFHS 1992/93 on the other hand collected information from 88,562 households and 89,777 ever-married women in the age group 13-49 years. Data for our analysis are drawn from the household and women's questionnaires which are common to both surveys.

To define the estimation sample, we follow the literature and restrict our sample to young adult siblings between the age of 16 and 27 years. The argument for estimating sibling correlations from closely spaced siblings rests on the fact that there may be important changes in the family structure as well as shocks to family life over a longer time horizon diluting the already conservative estimate of family background on children's outcome. To check the sensitivity of our results, we report the estimates of sibling and intergenerational correlations for other similar age groups also.

[27] The GLLAMM uses an iterative GLS to estimate the mixed effect model.
[28] The Bjorklund et al (2010) approach thus is a three-stage procedure.
[29] For details of the estimation method using GLLAMM, please see Rabe-Hesketh et al. (2002).

(3.1) Coresdiency Restriction and Its Implications

It is well-recognized in the literature that the co-resident sample is the appropriate one for estimation of the sibling correlation, as we would like to capture the factors siblings face while *growing up together* in a family (see, for example, the discussion in Bjorklund et al. (2010) and Mazumder (2008, 2011). If a sibling leaves the household (say goes to a boarding school in another state), by definition he/she shares very little common with the sample of co-resident children in terms of home, school and community environment. If a sibling completes education at home, but is much older than the other siblings, he also faces different school and community environment, because of increased schooling supply over time, for example.

However, the co-resident sample may bias the estimate of intergenerational link in education between parents and children, especially when the metric is intergenerational regression coefficient (IGRC). For example, among older sons, the best educated ones tend to leave household earlier than less educated ones, which is likely to bias intergenerational regression coefficient β downward. But it does not necessarily bias the estimate of intergenerational correlation, because such exit of better educated sons from the household also reduces the variance in their education, thus offsetting the decline in the intergenerational regression coefficients. Similar arguments hold if the girls leave parent's house after marriage and also drop out of school. It is well-known in the literature that truncation of the sample from below (due to marriage) or from above (due to job and education related migration) not only affects the mean, it also affects the variance (see, for example, Hausman and Wise (1977, 1978)).

Now consider the relation between IGC and IGRC: $\rho_{IG} = \beta \frac{\sigma_{sp}}{\sigma_s}$, where ρ_{IG} denotes IGC, β is the IGRC, σ_{sp} and σ_s are the standard deviation of parental and children's education respectively . Note that there is no sample selection for the parents, so the estimate of σ_{sp} is not biased. Because of coresidency restriction, both β and σ_s are downward biased when the sample is truncated from above, or below, or both. As a result, the bias due to coresidency is at least partly offset in the IGC estimate. Emran and Shilpi (2014) show that if the estimated IGRC (β) from a coresident sample is $\delta\,\beta$ ($\delta \neq 1$), then the corresponding estimate for IGC (θ) is $\sqrt{\delta}\,\theta$; thus the bias in the IGC estimate is lower by an order of magnitude. The evidence presented in Emran and Shilpi (2014) show that in case of India the estimates of IGC from coresident and full sample are virtually identical. According to their estimates the bias in IGC is less than 7 percent (in many cases less than 2 percent), while the bias in IGRC can be as high as 52 percent. For example, consider their estimates for male children in the 13-50 years age range which is subject to sample selection both due to the fact that some children did not finish schooling yet and also that the older children who stay back in rural areas with parents are likely to be less educated. The estimates of IGRC are 0.42 and 0.32 for full and coresident samples respectively, but the IGC estimates are 0.37 and 0.35 respectively. Emran and Shilpi (2014) also provide estimates of bias in coresident samples for sibling correlations. The evidence shows that the bias in sibling correlation estimates is much smaller than that in the IGRC estimates, but slightly higher than the IGC estimates. However, the age range used for sibling correlation in Emran and Shilpi (2014) does not match exactly the age range used in this paper. We use the data from Emran and Shilpi (2014) for 16-27 years age range and estimate the sibling correlations for the coresident and full samples. Interestingly, the estimates are virtually identical for the female sample (0.671 for coresident sample and 0.670 for full sample), while the estimate for male sample is biased downward by about 10 percent.[30] It is thus clear that our estimates of intergenerational and sibling correlations are not likely to suffer from any significant sample selection biases due to the co-resident sample.[31]

Tracking the same younger age cohort [16-27] between 1991 and 2006 has the added advantage that our estimates are comparable and are not unduly influenced by changes in co-residency pattern over the life cycle. Finally, we check sensitivity of our results with two alternative samples: 16-24 years and 19-24 years of age for children. The 16-24 years sample consists of young children and thus it reduces the selection because of older children leaving the household. The second sample excludes the youngest from our sample because some of the younger children (e.g. 16-18 years old) may not have completed schooling biasing the estimates of intergenerational correlations and regression coefficients downward. We perform an additional robustness check by repeating our estimation for the age group 19-24 years. For this age group, less than one percent of children are still in school.

(3.2) Empirical Approach

As noted in the introduction, our empirical approach differs in some important ways from that of the existing studies on educational mobility in India. With the exception of Azam and Bhatt (2012) who provide estimate of intergenerational correlations (only for sons-fathers), the available studies on educational mobility in India use variants of intergenerational regression coefficient (IGRC) as the only metric. Jalan and Murgai (2008) use the NFHS 1998/99 data to estimate intergenerational regression

[30] It is interesting and important to note here that although the sibling correlation estimate is biased for the sons' in coresident sample, the conclusion that the sibling correlation is higher for daughters remain robust across coresident and full samples. Emran and Shilpi (2014) provide more extensive evidence that comparison of different groups based on coresident sample with IGC and sibling correlation estimates do not conflict with the conclusions based on the full sample.

[31] We hasten to add an important caveat here about the bias estimate for the sibling correlation. As noted earlier, sibling correlation estimate is not biased if a child left the house at an early age and went to a boarding school in Delhi, for example. In that case, the inclusion of the nonresident children may actually bias the estimate of the effects of common environment. So part of the "bias" in the estimate may not be bias at all.

coefficients for different age cohorts and reach the conclusion that educational mobility has improved substantially over time (younger cohorts). Maitra and Sharma (2010) also rely on cohort based analysis of IGRC, but use 2005 India Human Development Survey, and arrive at a similar conclusion that educational mobility has improved.[32] Hnatkovska et al. (2011) examine the probability of children having a different level of education compared with their parents among the socially dis-advantaged Scheduled Caste/Tribes relative to rest of the population using different rounds of NSS data.[33] In contrast to Jalan and Murgai (2008) and Maitra and Sharma (2010), we track influences of family background and parental education directly for the same age cohort between 1992/93 (year immediately following economic liberalization) and 2006 (15 years after liberalization). For the reasons mentioned above, we restrict our sample to younger age cohort (16-27 years).

Estimation was carried out for all children and separately for brothers and sisters. Since an important objective of our study is to uncover spatial differences in intergenerational mobility, we also estimate the sibling and intergenerational correlations for sub-samples defined on the basis of geographical location such as rural and urban areas, and developed and less developed regions/states. Following Mazumder (2011 and 2008) and Solon et al (1991), our main estimation samples include the singleton households. We check robustness of our results by repeating the estimations for samples that exclude singleton households. The number of observations for different sub-samples is reported in Table 1. The samples for all children consist of 34,585 observations in 1992/93 and 39,562 observations in 2006. The average numbers of children per family are 2.35 in 1992/93 and 1.98 in 2006. The shares of singleton families in our sample of 16 to 27 years olds are 25 percent in 1992/93 and 36 percent in 2006. More than a third of the families have two children in both survey years. About 63 percent and 59 percent of our total samples are brothers in 1992/93 and 2006 respectively. As reported in Table 1, sample sizes for different sub-samples are considerable, the smallest sample size being 4,892 for sisters in urban areas in 1992/93. The large sample sizes ensure precision of our estimates of sibling and intergenerational correlations for both survey years.

Summary statistics from our main samples are presented in Table 2. The education levels of both boys and girls have improved between 1992/93 and 2006. Average education of boys increased from 7.63

[32] Trying to uncover trends in intergenerational correlations on the basis of estimates from different age cohorts is problematic when co-residency pattern of children and parents changes over life cycle. The coefficients tend to be underestimated for younger age-cohorts who may still be in school and tend to be over-estimated when parents co-reside with better educated children. Thus intergenerational regression coefficients may suggest a spurious decrease in intergenerational persistence across age cohorts simply due to changes in co-residency pattern over the life cycle.

[33] Hnatkovska et al. (2011) do not estimate intergenerational correlations. Instead they regress the probability of education switching (defined as children having different education level than parents) on scheduled caste and scheduled tribe (SC/ST) dummies for various rounds of NSS data between 1983 and 2005. The magnitudes of coefficients of SC/ST dummy are then compared to find the trend in intergenerational persistence among SC/ST compared with non-SC/ST population. Thus the estimates are similar to intergenerational regression coefficients in the sense that they are not standardized for changing variance across different generations.

years in 1992/93 and 8.76 years in 2006. The gains in girl's education have been more dramatic: it increased from 6.9 years in 1992/92 to 8.67 in 2006. As a result, the gap between boys and girls has narrowed considerably between these two survey years.[34] There are improvements in educational attainments in mother and father's generations as well. Yet, a substantial gap in father's and mother's education in both survey years indicates lack of a convergence in male and female education for parent's generation. Average education of father increased from 5.33 years to 6.43 years between the two survey years, while that of mother increased from 2.63 years to 3.75 years. The improvements in years of education were associated with a decline in the standard deviation of education levels between the survey years. Consistent with international evidence in Hertz et al. (2009), the variances of education levels are higher in parent's generation compared with the kids in both the survey years. This decline in variance implies that relying on measures of immobility that are not standardized by the variance such as intergenerational regression coefficient (IGRC) alone to understand intergenerational mobility may be misleading.

The summary statistics for the rural sample are also reported in Table 2. As expected, average education levels are lower in rural areas compared with our full sample. Consistent with national trends, average years of schooling have increased for both boys and girls in rural areas. The gender gap in education has also narrowed though the gap is still larger in rural areas compared with our full sample. Summary statistics for other sub-samples also confirm improvements in education attainment of children during this period. The trends in education levels reported here are consistent with those reported in other studies (ASER reports, World Bank (2011)).

In addition to education levels, Table 2 provides summary statistics for age and caste and religion composition of our sample. Overall, the samples from two years appear to be comparable to each other in terms of age though caste-religion composition indicates some change perhaps due to change in geographical coverage of the two surveys.[35]

(4) Empirical Results

Equations (7) and (8) form the basis of empirical estimation of sibling and intergenerational correlations respectively. To estimate the individual and family components of equation (7), we followed the Bjorklund et al. (2010) three-step procedure discussed earlier.[36] Unless otherwise noted, all standard

[34] Similar convergence in educational attainment between boys and girls over the reform period is observed in China (see, for example, Behrman et al. (2008)).

[35] The NFHS 1992/93 covered 24 states and Delhi (the Capital city) whereas 2006 survey covered all of the 29 states. The sample sizes in the NFHS are comparable to more widely used National Sample Surveys.

[36] Equation (7) can be estimated directly (without the first two steps) using Stata GLLAMM procedure when the set of control variables is small. However, it becomes unmanageable when we introduce neighborhood fixed effects.

errors are clustered at the family level. All sibling pairs are given equal weights in all estimation results presented in this paper.

(4.1) Results from the Full Sample

Table 3 reports the results for the full sample. The sibling and intergenerational correlations estimated from our simplest specification of equations (7) and (8) are reported in panel A. In this simplest specification, age dummies are introduced to control for children's age, and in the 'all children sample', a female dummy to account for gender difference in education level. The sibling correlation is estimated to be 0.642 in 1993 which declines slightly to 0.616 in 2006. Both of these parameters are estimated with great precision (t-statistics greater than 95).[37] The estimates imply that the influence of the factors common to siblings on their educational attainment is very high (more than 60 percent) and has remained remarkably stable over more than a decade. Interpreting it from a different angle, the estimates of sibling correlations suggest that individual effort and other idiosyncratic factors account for less than 40 percent of variations in schooling years, both in 1992/93 and 2006. The absolute magnitude of the sibling correlation in 2006 is quite high, higher than the available estimates for Latin American countries including Brazil and El Salvador.[38]

The third row in Table 3 reports the estimates of the intergenerational correlations between children and parents in education. We define the parent's education variable as the maximum of father's and mother's years of schooling. We, however, note that the results and conclusions in this paper are not sensitive to alternative definitions of parental education such as average of mother's and father's years of schooling. The intergenerational correlations reported in panel A are estimated from a simple specification that controls only for age and gender.[39] The estimates for all children show a slight decline in intergenerational correlations between two survey years: it declined from 0.574 in 1992/93 to 0.540 in 2006. The absolute magnitude of intergenerational correlation for India is, however, much larger than the

For the sake of comparability, results reported in this paper are from the procedure that estimated equation (7) in two steps. The results from single step estimation do not differ from that of two step procedure when applied to specifications that does not include neighborhood fixed effects.

[37] The REML estimates show a decline of sibling correlations from 0.659 in 1992/93 to 0.627 in 2006. The REML estimates are slightly larger than those from the three step procedure followed in this paper.

[38] The highest estimate is 0.60 among 16 Latin American countries, for El Salvador (Dahan and Gaviria (2001). Among developed countries, sibling correlations are found to be highest in USA. The estimates range between 0.6 (Mazumder (2008) for biological siblings in the same household for age cohort born during 1957-1969 and 0.63 (Conley and Glauber(2008) for siblings with same biological mother for age cohort 1958-76). In contrast, the Nordic and European countries are much more mobile, the average sibling correlation is around 0.4 (see Bjorklund and Salvanes (2010)).

[39] We follow the existing literature here. See for example, Bjorklund et al. (2010).

average for other Asian countries reported by Hertz et al (2009) (average=0.39).[40] Among 10 Asian countries covered by Hertz et al. (2009), only Indonesia has intergenerational correlation in education (0.55) which is comparable to that for India.[41] In contrast to the sibling correlations, the Latin American countries such as Brazil, Chile, Peru, and Colombia have lower intergenerational correlation (the estimate is around 0.60 as reported by Hertz et al. (2007)). As noted before, this implies that while parental education plays a less important role in India compared to Latin American countries, the overall educational mobility is lower in India as is evident from higher sibling correlation. This also vindicates our argument that conclusions based on a partial measure such as IGRC alone may be incomplete and even misleading.

(4.1.1) Gender and Intergenerational Mobility in Education

To understand any possible gender bias in the intergenerational educational mobility, we report estimates of sibling correlations for brothers and sisters separately in columns 3 to 6 of Table 3.[42] The estimates show that while the sibling correlation among men (brothers) did not change perceptibly between 1992/93 and 2006, it experienced a moderate decline in the case of women (sisters). The estimated sibling correlations are: 0.614 (1992/93) and 0.624 (2006) for men and 0.780 (1992/93) and 0.696 (2006) for women. Compared with men, the magnitude of sibling correlation among women is thus significantly higher in both survey years. This is in contrast with evidence from developed countries where there is no significant gender differences in sibling correlations (Bjorklund and Salvanes (2010)). Despite the moderate decline from 1992/93, the estimate for women in 2006 (0.696) is well above the upper bound estimates for sibling correlations among women found in developed countries.[43]

We also analyze the trend in intergenerational correlations between parents and children across gender (columns 3-6, Table 3). The intergenerational correlations for men remained stable (0.541 in 1992/93 and 0.523 in 2006), but for women, it declined moderately from 0.622 to 0.559 between the two survey years, 1993 and 2006. Consistent with our findings regarding sibling correlations,

[40] While intergenerational correlations for India here are estimated for 16-27 age cohorts, the estimates in Hertz et al are for adults in age range 20-69 years. As noted by Hertz et al (2009), with increase in the level of education for younger cohorts, the intergenerational correlations for younger cohorts have either become smaller or not change at all. In that sense, our estimates for the intergenerational correlations for India are likely to be on the conservative side as a metric for overall educational immobility.

[41] The intergenerational correlations in Latin American countries are higher than that of India. The average for 7 Latin American countries in Hertz et al (2009) is 0.60.

[42] The sibling correlations among sisters (brothers) are estimated by keeping only female (male) children in the sample. The sample sizes for sisters and brothers add up to that of all children because singleton households are included in all subsamples.

[43] The estimates of sibling correlation among sisters for developed countries fall within the range [0.46-0.6]. Only one study reported a significant difference in sibling correlations between brothers and sisters for USA (Conley and Glauber (2008)).

intergenerational correlations are stronger for women than men in both the years. The higher intergenerational persistence observed for women compared with men in India is consistent with the recent findings on intergenerational economic mobility from other developing countries. For instance, Emran and Shilpi (2011) find occupational immobility (as measured by intergenerational regression coefficient between parents and children) to be much higher for women in Nepal and Vietnam.

As discussed in the conceptual framework above, the square of intergenerational correlation provides an estimate of the share of total variance in schooling that can be explained by parent's education alone. The estimates (5th row in Table 3) show that parent's education alone can explain between 27 to 29 percent of variations in years of education for men (brothers) and 31 to 39 percent variations for women (sisters). In contrast, for developed countries, parental education explains only 9 to 21 percent of total variations in schooling years (see Bjorklund and Salvanes (2010)).

(4.1.2) Role of Caste and Religion

A potentially important determinant of educational attainment in India is the caste and religious identity of a household. Studies on education in India show that the average level of education is much lower among children from socially disadvantaged scheduled caste (SC) and scheduled tribes (ST) (Jalan and Murgai (2008), Kajima and Lanjouw (2006), Aslam et al. (2011)). In the next specification of our regressions, we include dummies for SC, ST and other backward castes. We also include a dummy for households whose head is a Muslim, as Muslim are among the most economically lagging groups in India (Sachar Committee (2006), World Bank (2011)). The effects of the caste and religion dummies on the estimated sibling correlation is minimal; the estimates in panel B of Table 3 are only slightly smaller compared with those reported in panel A. The inclusion of caste and religion dummies also does not affect the magnitudes of intergenerational correlations in any significant way. The results thus suggest that sibling and intergenerational correlations do not vary across caste groups in any significant way in both of the survey years, 1993 and 2006. The conclusion that the sibling and intergenerational correlations do not depend in any significant way on caste or religious identity is confirmed by the estimates from the sub-samples based on caste and religion reported in Table 6, panel A. The only exception is the urban women, where lower caste women experienced significantly higher mobility compared to the upper caste women (see section 4.2.1 below).

(4.1.3) Role of Geographic Location: Neighborhood Effect

As noted in the introduction, a focus of this study is to analyze the potential spatial aspects of intergenerational educational mobility in India, and whether the role of geography has changed over time in the post-reform period. A simple but powerful way to gauge the importance of geographic location is to include neighborhood fixed effects in the estimating equations, and then compare the estimates of sibling correlations and intergenerational correlations with and without the neighborhood fixed effect. Note that the fixed effect captures all the factors shared by the children growing up in a neighborhood which include peer effects and school availability and quality, among other things.

Panel C of Table 3 presents the estimates that include neighborhood fixed effects where neighborhood is defined as the sample cluster (PSU). A typical PSU consisted of 24 households in 2006 and 16 households in 1992/93 where households are usually chosen to be in close proximity to each other. Our full samples include 3799 and 3400 such clusters (PSUs) in 2006 and 1993 respectively. The results show that geographic location as measured by PSU level fixed effect matters a lot for intergenerational mobility in education. The estimates for sibling correlations become substantially smaller when neighborhood fixed effects are taken into account: the sibling correlations in the full sample decline from 0.642 to 0.395 in 1992/93 and from 0.616 to 0.385 in 2006. The implied neighborhood correlations (after netting out caste and gender effects) are 0.23 in 1993 and 0.20 in 2006 for the full sample, 0.22 (1993) and 0.19 (2006) for men, and 0.30 (1993) and 0.29 (2006) for women. These estimates of neighborhood correlations are substantially larger than those found for developed countries (Bjorklund and Salvanes (2010)).[44] The neighborhood correlations account for nearly a third of sibling correlations among men and 40 percent of that among women. This can be interpreted as strong evidence in favor of geographic location as a first order mediating factor for the influence of family background on education of children.[45]

The estimates of sibling correlations in panel C of Table 3 can be considered to be lower bound estimates of family background's influence on children's educational outcomes (net of caste and religion, and neighborhood effects), because the estimates of neighborhood effects are biased upward due to sorting of similar families in a neighborhood. These lower bound estimates imply that about 40 percent of variations in children's education can be explained by family background (net of neighborhood, caste and religion) alone. For women, the net influence of family background declined from 0.46 to 0.39 between 1992/93 and 2006, whereas it increased slightly for men from 0.38 to 0.41 over the same period.

The inclusion of neighborhood fixed effects reduces the estimates of intergenerational correlations between parents and children also. But compared with sibling correlations, the magnitudes of the reductions are much smaller. For instance, estimates for men (brothers) declined from 0.52 to 0.48 in 1992/93 and 0.49 to 0.45 in 2006. The estimates of intergenerational correlations in panel C indicate that more than 20 percent of variations in total schooling of children and nearly a third of raw sibling correlations in education (shown in panel A) can be explained by parent's education alone.

---

[44] The largest estimate for neighborhood correlation is 0.15 for USA (Solon et al. (2000)).

[45] The relatively larger role of location for women probably reflects lower geographic mobility among them.

To provide a sense of relative importance of different factors in explaining the variations in children's education, we use estimates from Table 3 and plot them in Figure 1. Figure 1 show that in 2006, individual effort and other idiosyncratic factors explain about 40 percent of variations in education outcomes of men, and about 30 percent for women, the rest are due to common factors experienced by siblings. The sibling correlation can be decomposed into three components: (i) parental education, (ii) geographic location (i.e., neighborhood effect), and a residual family environment shared by siblings which presumably capture the parental child rearing skills, among other things. The neighborhood effect is computed as the difference between sibling correlations estimates in panel A and C (e.g. 0.23 for all children sample in 2006). The contribution of parent's education to variance of children's education is squared of intergenerational correlation estimates in panel C (e.g. 0.196 for all children in 2006). The estimates show that neighborhood effect and parent's education together can explain more than 70 percent of sibling correlations (in panel A). The common neighborhood factors and parental education are particularly important for sisters (women): their share in sibling correlations is large -- 0.81 in 1992/93 and 0.78 in 2006. For men, contribution of these two factors to sibling correlations decreased from 0.78 in 1992/93 to 0.69 in 2006 due mainly to decrease in the neighborhood correlations (Figure1). For women, intergenerational persistence has declined but neighborhood correlations remain nearly unchanged.

(4.2) Geography of Educational Mobility: Rural vs. Urban

The evidence on strong neighborhood effects in sibling and intergenerational correlations discussed above brings the focus on geographic location as an important factor in understanding educational mobility in post-reform India. This raises the question whether the levels, time trends and gender patterns of sibling and intergenerational correlations differ significantly across different geographic areas; for example, are there any significant differences between rural and urban areas? The recent academic literature and reports in popular press in India give a strong impression that the rural areas have been largely bypassed by the positive effects of economic liberalization and strong economic growth that followed (World Bank (2011)). In this subsection, we provide additional analysis of the role of geographic location in intergenerational educational mobility by focusing on the rural-urban divide.

The estimates of sibling and intergenerational correlations for rural and urban areas are reported in Tables 4 and 5 respectively. Consistent with the format of Table 3, we represent estimates from three different specifications of equations (7) and (8) in three panels (A, B and C) of Table 4 and 5. These specifications correspond exactly to the specifications in Table 3 and are not discussed here again for the sake of brevity.

The sibling and intergenerational correlations for all children and for men are larger in magnitudes in urban areas compared with rural areas (Tables 4 and 5). For instance, for all children (men), the sibling correlation is 0.579 (0.556) in rural areas compared with 0.664 (0.652) in urban areas in 1992/93. The corresponding intergenerational correlations are 0.482 (0.465) and 0.572 (0.553) in rural and urban areas respectively. For women in 1992/93, there is practically no difference in sibling correlations between urban and rural areas (both approximately=0.74), though intergenerational correlation is higher in urban areas (0.593 vs. 0.523). Between 1992/93 and 2006, intergenerational correlations remained nearly unchanged for both men and women in rural areas. There was a marginal decline in the sibling correlation for rural women in the same period, but the sibling correlation among rural men increased slightly. For men in urban areas, both sibling and intergenerational correlations remained effectively unchanged between 1992/93 and 2006. In contrast, sibling and intergenerational correlations have decreased substantially for urban women during the same period. As a result, gender difference in sibling and intergenerational correlations effectively vanished in urban areas.

Using estimates from Table 4 and 5, we decompose the total variance in children's education into individual and common 'family background' components using the raw correlations in panel A. The family background component is further decomposed into three separate parts accounted for by parental education, common neighborhood environment and other common family factors (by comparing panel A and C estimates). The relative contributions of these different factors to total variance of children's education are plotted in Figures 2 and 3. Consistent with results from full sample, influences of parent's education and common neighborhood factors are important in both rural and urban areas, accounting for more than 60 percent of sibling correlations for men and 65 percent for women. The contribution of common neighborhood factors to variance in education is larger for rural women whereas parental education is relatively more important for both men and women in urban areas. Overall, common family backgrounds are perhaps most important factor for rural women for whom less than 30 percent of variations in education can be explained by individual effort, choices and other unobserved idiosyncratic factors.

To summarize, though the influences of parental education and common family background are smaller in magnitude in rural areas, there has been little or no progress over a decade and a half after the economic liberalization in 1991. The largest improvement in educational mobility has been experienced by urban women while men experienced effectively no improvement regardless of their location. We find common neighborhood environment and parental education as the most important source of sibling correlations in both urban and rural areas. The influences of common neighborhood factors are particularly important for rural women.

(4.2.1) Caste and Educational Mobility: Does Geography Matter?

Our main results presented in Tables 3-5 suggest no substantial differences in sibling and intergenerational correlations across caste groups. Readers may be curious if this conclusion holds true across geographical areas and gender groups. Table 6 reports sibling correlations different caste status for men, women and all children across rural and urban areas, and full sample. The estimates of sibling correlations are slightly smaller in magnitudes for lower caste than upper castes in the case of men and all children samples. This is mostly true for women as well with the exception of urban areas in 1992/93. Consistent with our earlier results, the sibling correlations for men and all children remained stable between 1992/93 and 2006 across all subsamples. Though sibling correlations among women of both upper and lower castes declined in all subsamples, the decline is substantial only in urban areas. The decline is particularly large for low caste women (from 0.77 to 0.56) compared with upper caste women (from 0.72 to 0.66).

(5) Robustness Checks

We check the sensitivity of our estimates in two ways. For younger age cohorts, estimates can be biased downward if a substantial proportion of them have not completed schooling.[46] To check the sensitivity of our estimates, we repeat our entire analysis for two age cohorts: younger age cohorts [16 to 20 years] who are less likely to be exiting household, and slightly older age cohorts [19 to 24 years] most of whom have already completed schooling.

Panels A and B of Table 7 report the results for the full sample for age cohorts [16-20] and [19-24] respectively. These results are based on the specifications that control for age and gender. For intergenerational correlations, we find no significant differences in the estimates for age cohorts [16-20 years] reported in Table 7 from those reported in Table 3 for any of the survey years. The estimates for 19-24 years olds are slightly larger than comparable estimates in Table 3. The changes in the sibling and intergenerational correlations between 1992/93 and 2006 implied by Table 7 for both age groups are similar to those implied by Table 3. For women, the decline in sibling correlations between 1992/93 and 2006 is somewhat smaller for the age group 19-24 years compared with full sample and sample of 16-20 years age cohort. However, our overall results regarding larger decline in sibling correlations for urban women between these two survey years are confirmed for the 19-24 years age cohorts as well. Consistent with Table 3, the estimates from Table 7 suggest large gender differences in sibling and intergenerational correlations.

[46] More than a half of the 16-17 year olds had not completed schooling in 2006. Less than 1 percent of children older than 19 years were still in school in 2006.

The estimates of sibling correlations for the age cohort [16 to 20 years, and 19-24 years] are larger in magnitudes both for men and women (see Table 7). This is consistent with the evidence in literature which finds higher sibling correlations among closely spaced children compared with widely spaced children (Bjorklund and Salvanes (2010)).[47] Such higher correlations arise from the fact that for more widely spaced children, family background may change substantially over time.

The results reported so far are based on samples which include singleton households in the sample. In the next robustness check, we repeat our entire analysis for the samples which excluded singleton households from the sample. Table 8 reports the results for full samples from our simplest regression specification that controls only for age and gender. While excluding singletons leads to considerable drops in sample sizes, the estimates for sibling and intergenerational correlations in Table 8 are almost indistinguishable from the comparable estimates in Table 3 (panel A). All other conclusions about trend and pattern of sibling and intergenerational correlations hold true for these truncated samples as well.

(6) Reconciling the Conflicting Evidence

Our main empirical results indicate that the sibling and intergenerational correlations in education in India remained largely unchanged over a period of almost a decade and a half (1993-2006) after the economic liberalization in 1991. The only group that experienced significant decline in the sibling and intergenerational correlations are women in urban areas. These results seem to contradict the evidence presented by Jalan and Murgai (2008) and Maitra and Sharma (2010) who find substantial improvements in educational mobility in India over time. They find that the magnitude of the *intergenerational regression coefficient (IGRC)* declined substantially for younger age cohorts. We noted earlier that the estimates of IGRC from co-resident sample are likely to be substantially biased. We also discussed some of the pitfalls in relying on cohort based analysis when data consists of only the co-resident children. In fact, Jalan and Murgai (2008) are well aware of the limitations of the cohort analysis and discuss many of the same points we raised earlier. Our results can differ from the earlier studies on three grounds: (i) intergenerational correlation takes into account the declining variance in education in children's generation, and (ii) we compare the same age cohorts (16-27 years) across two surveys, instead of relying on the different age cohorts in a single survey round, and (iii) both of these studies cover a much longer time period using different cohorts (our analysis focuses on changes over a decade and a half), and part of the difference in the results may arise from the difference in time horizons.

It is, however, important to check if we get estimates similar to those of Jalan and Murgai (2008) and Maitra and Sharma (2010) from a cohort based analysis. We thus report the estimated

[47] This is in contrast to IGRC where the estimates for younger cohorts are likely to be biased downward. See below.

intergenerational regression coefficients (IGRC) and intergenerational correlations (IGC) for both rural and urban areas for three age cohorts [15-19, 20-24 and 25-29 year] in Table 9. It is interesting that the estimates of IGRC show a declining effect of parental education for the younger age cohorts for both the survey years which is consistent with the estimates in Jalan and Murgai (2008) and Maitra and Shrama (2010). However, a comparison of estimates of IGRC for each age cohort between two survey years shows a different picture: only a small decline in 2006 compared with 1992/93. This reinforces the conclusion that no significant improvements have occurred in the post-reform period when comparing similar age cohorts in two different points of time.

The estimates of IGC also show small differences across age cohorts for both years and no substantial change across years for each age cohorts. A comparison of IGRC and IGC estimates shows that our results are different from that of Jalan and Murgai (2008) and Maitra and Sharma (2010) partly due to the differences in the ways IGC and IGRC are estimated: IGC accounts for change in both means and variances of education distributions of both generations whereas IGRC ignores the changes in variances of children's education. Thus while the IGRC estimates suffer from the coresident sample bias significantly, the IGC estimates have very little bias in a coresident sample.

To put the conflicting estimates in perspective, we emphasize that the intergenerational regression coefficient (IGRC) and intergenerational correlation (IGC) are designed to answer different questions. If one is interested in the question: how much does a one year more schooling on average in parental generation matter for the average schooling attainment in the children's generation, IGRC is the appropriate metric (assuming no sample selection problem). However, if one is interested in understanding how much of the observed cross-sectional inequality in educational outcome (years of schooling) of children can be attributed to inequality in educational opportunity as measured by educational inequality in the parental generation, the relevant metric is IGC. Since the focus of most of the recent policy and academic debate has been on understanding the role of inequality of opportunity in the observed inequality of outcomes, the IGC, rather than IGRC, seems to be the more appropriate metric. Another important advantage of IGC as a measure is that it is less sensitive to the details of empirical implementation, and thus lends to comparison across different studies across different countries, a point emphasized by Hertz et al. (2007), among others. For example, the magnitude of IGRC can change significantly depending on how parental educational attainment is measured (the maximum, or the average of parents schooling), but the IGC is much more robust.[48] It is also important to remember that IGC is only a partial measure compared to sibling correlation (SC), because parental education is only

[48] We follow Behrman et al. (2001) and use maximum of mother's and father's schooling. Hertz et al. (2007) use average of mother's and father's schooling. Similar to Hertz et al. (2007), we find that the estimated IGC do not vary significantly, if we use average schooling instead as parental educational attainment. An earlier version of this paper reported the estimates of IGC and SC using average parental schooling, which are omitted for the sake of brevity.

one element in the set of factors shared by the siblings. Thus both IGC and SC need to be estimated to understand the role of family background in the observed cross-sectional schooling inequality.

(7) Toward an Understanding of the Trends in and Pattern of Educational Persistence

The objective of this study is to provide robust evidence on the trends in and pattern of educational mobility in post-reform India with special emphasis on the roles played by gender and geography. In other words, the goal here is to help establish the "facts" about educational mobility in India over a period of a decade and a half after extensive economic liberalization. In this section, we attempt a first pass at understanding the observed trends in and pattern of educational mobility in post-reform India delineated in earlier sections. We bring together the existing evidence on the many different pieces of the puzzle to provide some plausible explanations, which can be thought of as building blocks for future in-depth analysis of educational mobility in India. We, however, hasten to add that our discussion is only a small step in a major research program that needs to be undertaken to understand the nature of educational mobility in post-reform India. We also emphasize a caveat widely understood in the literature that although the estimates of sibling and intergenerational correlations are important as measures of educational mobility over time, they do not imply causality. Among other things, the literature has emphasized the difficulties in causal interpretations because of correlations in genetic endowment (ability) and preference among the siblings and also between the parents and children (see, for example, Bjorklund and Salvanes (2010)).[49] We, however, emphasize that our study enjoys an advantage in this regard. This is because the changes observed over time are not likely to be driven by *changes* in genetic correlations among siblings and between parents and children, as a decade and a half is a short span of time for any significant changes in genetic correlations. Thus when one observes large changes over a relatively short period of time, as we do in the case of women in urban areas, for example, it is likely that they reflect changes in the 'environmental factors' in the household and the community. The evidence in this paper thus can be helpful in identifying potential causal factors, and thus constitutes an essential first step to policy-relevant economic analysis of educational mobility. Our results indicate that the focus of a causal analysis of the observed educational persistence should primarily be on the geographic location, parental education and their correlates. The large impact of geography including the neighborhood effect points to the importance differences in school availability and quality and access to

[49] It is, however, important not to push the distinction between "nature" and "nurture" too far, because there are important interactions between the two, a point emphasized in the Behavioral Genetics literature (see, for example, Plomin et al. (2001)). For interesting discussions on the limitations of the nature vs. nurture debate, see Goldberger (1979).

urban markets (returns to education). The importance of parental education, on the other hand, suggests the possible importance of credit constraints and role model effects as potential causal channels.[50]

The recent literature has underscored the importance of schooling expansion and returns to education as major factors in determining trends in educational mobility (and more generally economic mobility including income mobility).[51] The evidence indicates that educational mobility improves when government invests heavily in educational infrastructure to ensure access at low costs.

(7.1) Rural-Urban Gap: Understanding the Higher Correlations in Urban Areas

The magnitude of sibling and intergenerational correlations are in general larger in urban areas. This is true in 1993 for both men and women, also for men in 2006. It seems puzzling, because the schooling infrastructure and financial sector are expected to be more developed in urban areas. However, there are a number of factors that may help explain the observed higher persistence in educational attainment in urban areas compared to the rural areas.

Most of the schools, both primary and secondary, in rural areas are public schools and thus tuition free. The public primary schools also provide mid-day meals. The absence of tuition costs and provision of mid-day meals help the poorer households (parents with lower education) to send their children to schools. Moreover, the private market for supplementary tutoring is not developed in rural areas. A private market for quality tutoring could potentially give an advantage to the children of richer (and more educated) parents, creating inequality in educational opportunities. The above factors combined together weaken the link between parental income and children's educational attainment in the rural areas. In contrast, there has been dramatic growth in private schools and supplementary private tutoring in urban areas in India in last couple of decades (Kingdon (2007), World Bank (2009)). According to one estimate the share of enrollment in private secondary schools in urban India was about 30-40 percent in 2002. Thus parental income and access to credit have become increasingly important in urban India for children's education, creating more prominent role for parental education and family background.

Another important factor is the differences in returns to education. The available estimates for 1993/94 shows that while returns to primary education were higher in rural areas, returns to higher education were higher in urban areas (Duraisamy (2002)). The recent estimates indicate that the rural-urban gap in returns to education has increased after the liberalization (Aslam et al. (2011)). For wage employment for men, it is 6.3 percent in rural areas, but 32 percent in urban areas. For female wage

[50] For evidence that parent's may be important role models in occupation choices, especially for women in Nepal, see Emran and Shilpi (2011).
[51] See, for example, the discussion on the role of inequality in access of higher education and increasing wage premium for higher education in explaining the observed decline in mobility in UK by Blanden et al. (2008).

employment, the corresponding returns are 8 percent (rural) and 44 percent (urban) (see Aslam et al. (2011)).[52] As noted by many observers, the economic growth in India after economic liberalization has been both skill-biased and urban-biased, driven by service sector growth including information technology (Kotwal et al. (2011), Bardhan (2010)).[53] The higher returns to education in urban areas make the investment in children's education more attractive for all parents. The potential positive effect of higher returns to education on disadvantaged children's educational mobility is, however, counteracted by an important dynamic interaction between parental education and higher returns following liberalization. After the liberalization, the more educated parents could take advantage of the emerging opportunities in the urban labor market and they experienced higher income growth. The higher income allowed them to invest in children's education to reap the benefits of increasing returns to education. But the poor (and relatively uneducated) parents were less successful in taking advantage of the skill intensive growth process. Thus while the children of relatively educated parents in urban India continued to receive more and better education, the children of less educated parents failed to move beyond their parent's ranks, resulting in persistence between children and parent's education in the urban areas. Our evidence suggests that this is especially true for men in urban areas, but the experience of urban women requires additional explanations as they had substantial improvements in educational opportunities in the face of the forces discussed above. We turn to possible resolution to this puzzle in the next section.

(7.2)The Curious Case of Urban Women

Although the factors discussed above are expected to tighten the link between parental education and children's education in urban areas, the evidence in this paper shows that women in urban areas experienced substantial improvements in educational mobility. This comes across as especially counterintuitive in the context of a country where son preference is strong. However, note that even though the sibling correlation among sisters has gone down the most over the sample period, even in 2006 the magnitude of both sibling and intergenerational correlations remain significantly higher for women, indicating lower educational mobility compared to men. A related important finding is that the lower caste women experienced a larger decline in sibling correlations compared to the upper caste women. This may seem doubly puzzling as in addition to gender bias the lower caste women face significant disadvantages both in social and market interactions.

---

[52] The skill biased nature of economic growth can also be seen from growth of wages across different schooling levels. The average wage in 1999/2000 for someone with a college degree was 73 percent higher than someone with high school degree, and 67 percent higher for someone with high school degree compared to someone with middle school education (based on NSS data).

[53] A substantial part of the readymade garments industry is located in large cities including Delhi and Bangalore.

Interestingly, contrary to the conventional view, son preference in education in fact implies that as incomes grow (and/or credit market access improves), parents find it acceptable to invest in a daughters' education. Given the high perceived returns to a son's education (family lineage, old age support, dowry, social prestige etc.) the parents try to invest in son's education even if they face poverty. They start to invest in lower return assets such as a daughters' education only when they have more income and/or face lower credit constraints. It is thus only natural that a growing number of urban parents began to invest more in girls' education when their income grew following the liberalization. As noted above, the returns to education for women in urban areas has increased substantially over the reform period which makes it more attractive to invest in daughters' education irrespective of parental education.

Improvements in schooling facilities in urban areas might also have played a role. The absence of restrooms can be an important constraint for girls to attend middle and high schools; many girls drop out when they reach puberty because there is no restroom. Most of the urban schools in India now have separate restrooms for girls.[54] Note that the absence of girl's restroom affects the schooling of girls across the distribution, and thus elimination of this constraint would increase mobility.

The finding that lower caste urban women experienced more mobility compared to the upper caste women may in part reflect lower social constraints on their labor market participation. The caste difference can also be understood in terms of the insightful analysis of Munshi and Rosenzweig (2006) who find that the lower caste women were able to adapt better to the new occupations, as they are not expected to follow in their father's footprint. This 'freedom by neglect' helps the daughters achieve better occupational mobility which is likely to feed into higher educational attainment. Using survey data from Mumbai, Munshi and Rosezweig (2006) find that the sons in lower caste families were channeled into local language schools with established parental network and entered into traditional parental occupations. The daughters, on the other hand, enrolled in English medium schools and were better prepared to take advantage of non-traditional jobs, especially those with a premium for English proficiency.

A straightforward implication of the above discussion is that the same set of factors is likely to be responsible for lack of improvements in educational mobility for women in rural areas. Low parental income, low returns to education, lack of skilled jobs, lack of girl's restroom, and stronger social norms against women's participation in the labor market, all combined together can result in little or no improvements in educational mobility for rural women

[54] We thank Pranab Bardhan for pointing out to us the importance of girl's restroom in this context.

Conclusions

The Indian economy grew at a robust pace since its economic liberalization in 1991and achieved significant reduction in poverty. At the same time, the evidence indicates an increase in inequality (World Bank (2011), Deaton and Dreze (2002), Datt and Ravallion (2010)). This paper examines the trends in and patterns of intergenerational mobility in education among new entrants in the labor force (16-27 year olds) between 1992/93 – a year immediately following the economic liberalization – and 2006 – nearly 15 years after liberalization. To the best of our knowledge, this is the first paper in the literature to employ both intergenerational and sibling correlations to study the evolution of educational mobility in a developing country. An important advantage of using sibling and intergenerational correlations instead of intergenerational regression coefficient used in much of the earlier literature is that the conclusions are likely to be robust in data sets with coresidency restriction.

The empirical results indicate that educational mobility remained effectively unchanged for a large proportion of Indian youth after a decade and a half of high economic growth. This evidence is important because it shows that the earlier conclusions in the literature indicating substantial improvements in educational mobility in post-reform India depend critically on the data and the measure of mobility used. According to our estimates of sibling and intergenerational correlations, between 1992/93 and 2006, the only group that experienced significant improvements in educational mobility is women in urban areas. We find that estimates of sibling and intergenerational correlations among men stayed almost the same over the reform period in urban areas, but may have increased slightly in rural areas. In contrast, the sibling and intergenerational correlations among women have declined irrespective of geographic location, but only women in urban areas have experienced a substantial decline. Interestingly, among the urban women, there are significant caste differences: the lower caste women have experienced significantly better educational mobility compared to the upper caste women. Similar improvements in the educational mobility of women during a period of high growth and schooling expansion have been observed in other countries, for example, China and Malaysia. We discuss a number of possible reasons behind the observed patterns in educational persistence in post-reform India which can motivate future analysis of intergenerational educational mobility in India.

Although the trend in women's educational mobility, especially in urban areas, shows clear improvements, it is important to put the changes in perspective by looking at the magnitudes of the correlations. While the improvements have effectively eliminated gender gaps in urban areas, the magnitudes of the sibling correlations remain high in 2006 compared to other countries. For example, the sibling correlation for both men and women in 2006 is approximately 0.64 which is higher than the available estimates for Latin American countries including Brazil. The estimates, however, indicate that intergenerational correlation in education is lower in India compared to most of the Latin American countries. This implies that while the over-all educational mobility is lower in India compared to Latin American countries, the direct role played by parents in India seems to be less important. It is thus misleading to conclude that educational mobility is higher in India on the basis of estimates of intergenerational regression coefficients. Another important finding is that neighborhood effect constitutes an important part of inequality of opportunity as measured by sibling correlation, which is different from the available evidence on developed countries. In rural areas, the gender gap remains substantial in 2006; for example, sibling correlation among women is 0.70 in rural areas compared with 0.57 for men respectively. The high levels of educational persistence across generations are also evident in intergenerational correlations where the estimates for India are amongst the largest for Asian countries (Hertz. et al (2009)).

References

Aslam. M. , G. Kingdon, A. De, and R. Kumar (2010), Economic Returns to Schooling and Skills: An Analysis of India and Pakistan, Working Paper.

Arrow, K, S. Bowles, S. Durlauf (2000). *Meritocracy and Economic Inequality*, Princeton University Press.

Atkinson, A (1981). “On Intergenerational Income Mobility in Britain”, *Journal of Post Keynesian Economics*, 3, 194-18.

Banerjee, Abhijit, and Thomas Piketty (2005). “Top Indian Incomes, 1922–2000.” *World Bank Economic Review*, 19(1): 1–20.

Bardhan, P (2010). *Awakening Giants, Feet of Clay: Assessing the Economic Rise of China and India*, Princeton University Press.

Bardhan, P (2007). “Poverty and Inequality in China and India: Elusive Link with Globalization”, *Economic and Political Weekly*, 42 (38).

Basu, K (2008), “The Enigma of India’s Arrival: A review of Arvind Virmani’s Propelling India: From Socialist Stagnation to Global Power”, *Journal of Economic Literature*, 46 (2).

Behrman, Jere R. and Paul Taubman (1986). “Birth Order, Schooling and Earnings”, *Journal of Labor Economics*, 4(3).

Behrman, J., E. Hannum, M. Wang, and J. Liu (2008): Education in the Reform Era, in Brandt and T. Rawski Ed: China's Great Economic Transformation, Cambridge university Press.

Bjorklund A and K. Salvanes. (2011). “Education and Family Background: Mechanisms and Policies,” *Handbook in the Economics of Education* vol 3, EA Hanushek, S Machin and L Woessmann (es.), The Netherlands: North Holland, 2011, pp. 201-247.

Bjorklund A, Lindahl L, Lindquist M. (2010). “What More Than Parental Income, Education and Occupation? An Exploration of What Swedish Siblings Get from Their Parents.” *B.E. Journal of Economic Analysis & Policy*, 10 (1) (Contributions), Article 102.

Bjorklund A, Jantti, M, Lindquist, M. (2009). “Family background and income during the rise of the welfare state: trends in brother correlations for Swedish men born 1932-1968.” *Journal of Public Economics*, 93 (5-6), 671-680.

Björklund, Anders & Jäntti, Markus, 2012. "How important is family background for labor-economic outcomes?," Labour Economics, Elsevier, vol. 19(4), pages 465-474.

Black, S. E. and P. Devereux (2010). “Recent Developments in Intergenerational Mobility”, Forthcoming *Handbook of Labor Economics*.

Black, S. E., P. Devereux and K. Salvanes (2005). “The More the Merrier? The Effect of Family Composition on Children’s Outcomes”. *The Quarterly Journal of Economics*, 120(2), 669–700.

Blanden, J, P. Gregg, S. Machin (2005). Intergenrational Mobility in Europe and North America, Center For Economic Performance Report, April, 2005.

Blanden J., A. Goodman, P. Gregg, and S. Machin (2004). Changes in Intergenerational Mobility in Britain, in M. Corak (ed.) Generational Income Mobility, Cambridge University Press.

Blanden, J. P. Gregg and L. Macmillan (2008). Accounting for Intergenerational Persistence: Non-Cognitive Skills, Ability and Education, Economic Journal, 117, C43-C60.

Conley, Dalton and Rebecca Glauber (2008). “All In the Family? Family Composition, Resources, and Similarity in Socioeconomic Status”. *Research in Social Stratification and Mobility*, 26, 297–306.

Dahan, M and A. Gaviria (2001), “Sibling Correlations and Intergenerational Mobility in Latin America”, *Economic Development and Cultural Change*”, Vol 49, no.3.

Datt, G and M. Ravallion (2010), “Shining for the Poor, Too?” *Economic and Political Weekly*, February 13 2010.

Dearden, L and S. Machin, and H. Reed (1997): Intergenerational Mobility in Britain, Economic Journal, January, pp.47-66.

Deaton, A, and J. Dreze (2002). “Poverty and Inequality in India: A Re-Examination”, *Economic and Political Weekly*, Vol. 37, No. 36 (Sep. 7-13, 2002), pp. 3729-3748.

Dreze, J and A. Sen (2011), “ Putting Growth In Its Place: It has to be but a means to development, not an end in itself”, Otlook India.com, Nov 14, 2011.

Duraisamy, P. (2002). “Changes in the returns to education in India, 1983–94: by gender, age-cohort and location”, *Economics of Education Review,* 21(6), 609–622.

Emran, M. Shahe and F. Shilpi. 2011. “Intergenerational Occupational Mobility in Rural Economy: Evidence from Nepal and Vietnam”, *Journal of Human Resources*, 46(2), p.427-458.

Emran. M. Shahe and Yan Sun (2011). “Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China: 1988-2002”, Paper presented at American Economic Association Annual Conference, Denver, 2011.

Emran, M Shahe and F. Shilpi (2014), “It Depends on How You Measure it: The Co-resident Sample Bias in Estimating Intergenerational Economic Mobility,” Working Paper, World Bank.

Gelbach, J (2009), When Do Covariates Matter? And How Much?, Working paper, ASU.

Goldberger, A. (1989), "Economic and Mechanical Models of Intergenerational Transmission," *American Economic Review*, vol. 79, no.3, 504-513.

Green, F, S. Machin, R. Murphy, and Y Zhu (2010), The Changing Economic Advantage From Private Schools, Working Paper.

Hnatkovska, V., A. Lahiri and S. Paul. 2013. “Breaking the Caste Barrier: Intergenerational Mobility in India,” Journal of Human Resources.

Hertz Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith and Alina Veraschagina. 2007. “The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends”. *The B.E. Journal of Economic Analysis and Policy* (Advances), 7(2), Article 10.

Kotwal, A, B. Ramaswami and W. Wadhwa (2011). "Economic Liberalization and Indian Economic Growth: What's the Evidence?," *Journal of Economic Literature*, American Economic Association, vol. 49(4), pages 1152-99, December.

Jalan, J. and R. Murgai. 2008. “Intergenerational Mobility in Education in India”, Manuscript, World Bank, Delhi.

Jalan, J and M. Ravallion (1999), "China's Lagging Poor Areas," *American Economic Review,* vol. 89(2).

Jushong, Z and R. Kanbur (2012), “Confronting Rising Inequality in Asia”, Chapter in Asian Development Outlook 2012, ADB.

Kingdon, G (2007). "The progress of school education in India," *Oxford Review of Economic Policy*, vol. 23(2), pages 168-195, Summer.

Kochhar, K., U. Kumar, R. Rajan, A. Subramanian and I. Tokatlidis. 2006. “India’s pattern of development: What happened, what follows?" *Journal of Monetary Economics,* 53 (5): 981–1019.

Lillard, L, and R. J. Willis, 1994. "Intergenerational Educational Mobility: Effects of Family and State in Malaysia," *Journal of Human Resources*, vol. 29(4), pages 1126-1166.

Lindahl, L. 2008. “A Comparison of Family and Neighborhood Effects on Grades, Test Scores, Educational Attainment and Income—Evidence from Sweden”. Essay III in L. Lindahl, *Family Background and Individual Achievement,* dissertation no.75, Swedish Institute for Social Research.

Lindahl L (forthcoming) “Does the Childhood Environment Matter for School Performance, Education and Income? - Evidence from a Stockholm Cohort.” *Journal of Economic Inequality.*

Luke, N and K. Munshi (2011), “Women as Agents of Change: Female Income and Mobility in India”, *Journal of Development Economics*, vol. 94, 2011.

Maitra, P and A. Sharma (2010), “Parents and Children: Education Across Generations in India”, Working paper, Monash University.

Mazmuder, B (2012), “Is intergenerational economic mobility lower now than in the past?”, *Chicago Fed Letter*, April 2012.

Mazumder, B. 2011. “Family and Community Influences on Health and Socioeconomic Status: Sibling Correlations Over the Life Course,” *The B.E. Journal of Economic Analysis & Policy:* Vol. 11: Iss. 3 (Contributions), Article 1.

Mazumder, B. 2008. “Sibling Similarities and Economic Inequality in the US” *Journal of Population Economics*, 21, p. 685-701.

Munshi, Kaivan and Mark Rosenzweig. 2006. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, vol. 96(4), pages 1225-1252, September.

Plomin, R, DeFries, J, P. McGuffin, G. McClearn, and (2001). *Behavioral Genetics*, Worth Publishers, New York, 4th Edition.

Prasad, E (2012), “A Clarion Call for Emerging Markets”, Project Syndicate, February 14, 2012.

Pratham. 2010. “ASER 2009 – Annual Status of Education Report (Rural) 2009”, New Delhi, Pratham Resource Center.

Rabe-Hesketh, Sophia, Anders Skrondal and Andrew Pickles. 2002. “Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature,” *Stata Journal,* 2(1), 1-21.

Ravallion, Martin (2000): “Should Poverty Measures be Anchored to the National Accounts?”, *Economic and Political Weekly*, 34 (35 and 36), 26 August, 3245-52.

Sen, A. and Himanshu (2004), “Poverty and Inequality in India-II: Widening Disparities during the 1990s”, *Economic & Political Weekly,* 39

Solon, Gary (1999). “Intergenerational Mobility in the Labor Market,” in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* 3A, Elsevier, Amsterdam, North Holland.

Solon, G, M. Page, and G. Duncan (2000). “Correlations between Neighboring Children in Their Subsequent Educational Attainment,” *Review of Economics and Statistics* 82(3), 383-392.

Solon, G, M. Corcoran, R. Gordon and D. Laren. 1991. “A Longitudinal Study of Sibling Correlations in Economic Status,” *Journal of Human Resources* 26 (3), 509-534.

Srivastava, S (2004), “India mega-wedding justifies hype”, BBC Online, February 11, 2004.

Strauss, John and Thomas, Duncan, 1995. "Human resources: Empirical modeling of household and family decisions," Handbook of Development Economics, in: Hollis Chenery† & T.N. Srinivasan (ed.), Handbook of Development Economics, edition 1, volume 3, chapter 34, pages 1883-2023 Elsevier.

Woolsey, M (2008), World's Most Expensive Homes: Inside the World's First Billion-Dollar Home”, Forbes Com, April 30, 2008.

Walton, M (2010). Inequality, Rents and the Long-run Transformation of India, Working Paper, KSG, Harvard University.

World Bank (2011). *Perspectives on Poverty in India: Stylized Facts from Survey Data*, World Bank, Washington DC.

Table 1: Number of Observations for different samples

	All children		Brothers (Men)		Sisters (Women)	
	1993	2006	1993	2006	1993	2006
Full sample	34,585	39,562	21,895	23,625	12,690	15,937
Rural	22,308	20,191	14,510	12,247	7,798	7,944
Urban	12,277	19,371	7,385	11,378	4,892	7,993

Table 2: Summary Statistics

	Full Sample				**Rural Sample**			
	1992/93		2006		1992/93		2006	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Children's Schooling Years								
All Children	7.36	4.51	8.72	3.92	6.28	4.44	7.71	3.94
Brothers	7.63	4.32	8.76	3.77	6.77	4.31	7.97	3.75
Sisters	6.90	4.77	8.67	4.13	5.36	4.54	7.32	4.18
Parent's Schooling Years								
Father	5.33	4.94	6.43	5.09	3.91	4.26	4.90	4.56
Mother	2.63	3.91	3.75	4.58	1.46	2.82	2.17	3.41
Parent's[1]	5.54	4.93	6.82	5.03	4.07	4.27	5.19	4.54
Children's Age (years)								
All Children	19.55	2.98	19.22	2.96	19.52	2.98	19.08	2.95
Brothers	19.91	3.07	19.67	3.12	19.87	3.07	19.55	3.12
Sisters	18.93	2.70	18.57	2.56	18.86	2.69	18.36	2.51
Caste and Religion Composition								
Proportion Scheduled Caste	0.11	0.31	0.17	0.38	0.12	0.33	0.18	0.38
Proportion Scheduled Tribe	0.12	0.33	0.11	0.31	0.16	0.36	0.16	0.37
Proportion Backward Caste	-	-	0.33	0.47			0.34	0.47
Proportion Muslim	0.17	0.37	0.16	0.36	0.19	0.39	0.12	0.33

Note: 1: Maximum of Father and Mother's schooling years

Table 3: Sibling and Intergenerational Correlations: Full Sample

	All children		Brothers		Sisters	
	1993	2006	1993	2006	1993	2006
			PANEL A			
Sibling Correlation (SC)	0.642	0.616	0.614	0.624	0.780	0.696
	(110.90)***	(97.43)***	(71.20)***	(67.21)***	(89.72)***	(57.45)***
Intergenerational Correlation (IGC)	0.574	0.540	0.541	0.523	0.622	0.559
	(107.42)***	(105.48)***	(84.35)***	(82.41)***	(82.91)***	(77.53)***
IG squared	0.329	0.292	0.293	0.274	0.387	0.312
Proportion of SC explained by IGC	0.514	0.474	0.476	0.439	0.496	0.449
Controlling for Caste and religion			PANEL B			
Sibling Correlation (SC)	0.624	0.586	0.598	0.597	0.764	0.675
	(102.88)***	(87.49)***	(66.67)***	(61.09)***	(81.99)***	(51.83)***
Intergenerational Correlation (IGC)	0.551	0.505	0.521	0.491	0.594	0.522
	(99.15)***	(94.70)***	(77.93)***	(74.12)***	(75.73)***	(69.17)***
IGC squared	0.304	0.255	0.271	0.241	0.353	0.272
Proportion of SC explained by IGC	0.487	0.435	0.454	0.404	0.462	0.404
Controlling for Neighborhood FE			PANEL C			
Sibling Correlation (SC)	0.395	0.385	0.375	0.406	0.460	0.389
	(45.04)***	(42.96)***	(29.36)***	(30.42)***	(21.85)***	(17.43)***
Intergenerational Correlation (IGC)	0.479	0.443	0.474	0.452	0.519	0.455
	(97.66)***	(95.66)***	(76.56)***	(75.56)***	(70.65)***	(66.39)***
IGC squared	0.229	0.196	0.225	0.204	0.269	0.207
Proportion of SC explained by IGC	0.580	0.509	0.599	0.503	0.585	0.532
No. of observations	34,585	39,562	21,895	23,625	12682	15,937

Robust t statistics in parentheses. Standard errors corrected for clustering at family level
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Sibling and Intergenerational Correlations: Rural Sample

	All children		Brothers		Sisters	
	1993	2006	1993	2006	1993	2006
			PANEL A			
Sibling Correlation (SC)	0.579	0.569	0.556	0.571	0.746	0.698
	(72.58)***	(60.83)***	(47.60)***	(40.09)***	(57.81)***	(41.99)***
Intergenerational Correlation (IGC)	0.482	0.481	0.465	0.458	0.523	0.518
	(69.33)***	(68.16)***	(57.70)***	(52.62)***	(48.72)***	(51.02)***
IGC squared	0.232	0.231	0.216	0.210	0.274	0.268
Proportion of SC explained by IGC	0.401	0.407	0.389	0.367	0.367	0.385
Controlling for Caste and religion			PANEL B			
Sibling Correlation (SC)	0.565	0.543	0.543	0.548	0.732	0.680
	(68.69)***	(55.76)***	(45.38)***	(37.07)***	(53.94)***	(38.51)***
Intergenerational Correlation (IGC)	0.464	0.450	0.449	0.428	0.500	0.484
	(65.11)***	(62.26)***	(54.03)***	(48.13)***	(45.46)***	(46.44)***
IGC squared	0.215	0.203	0.202	0.183	0.250	0.234
Proportion of SC explained by IGC	0.381	0.373	0.371	0.335	0.341	0.345
Controlling for Neighborhood FE			PANEL C			
Sibling Correlation (SC)	0.379	0.327	0.361	0.355	0.480	0.396
	(35.60)***	(25.77)***	(23.35)***	(18.40)***	(18.07)***	(12.68)***
Intergenerational Correlation (IGC)	0.414	0.400	0.414	0.398	0.449	0.430
	(69.31)***	(62.63)***	(54.39)***	(47.67)***	(48.21)***	(45.05)***
IGC squared	0.171	0.160	0.171	0.158	0.202	0.185
Proportion of SC explained by IGC	0.452	0.490	0.474	0.446	0.420	0.466
No. of observations	22,308	20,191	14,510	12,247	7798	7,944

Robust t statistics in parentheses. Standard errors corrected for clustering at family level
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Sibling and Intergenerational Correlations: Urban Sample

	All children		Brothers		Sisters	
	1993	2006	1993	2006	1993	2006
			PANEL A			
Sibling Correlation (SC)	0.664	0.625	0.652	0.648	0.743	0.634
	(67.33)***	(68.34)***	(46.99)***	(51.55)***	(43.96)***	(30.25)***
Intergenerational Correlation (IGC)	0.572	0.526	0.553	0.534	0.593	0.508
	(54.47)***	(65.49)***	(44.41)***	(54.60)***	(40.06)***	(45.63)***
IGC squared	0.327	0.277	0.306	0.285	0.352	0.258
Proportion of SC explained by IGC	0.493	0.443	0.469	0.440	0.474	0.407
Controlling for Caste and religion			PANEL B			
Sibling Correlation (SC)	0.649	0.588	0.638	0.614	0.730	0.600
	(62.87)***	(59.05)***	(44.49)***	(45.19)***	(41.26)***	(25.97)***
Intergenerational Correlation (IGC)	0.554	0.482	0.537	0.493	0.573	0.461
	(51.19)***	(57.40)***	(41.89)***	(47.84)***	(37.64)***	(39.79)***
IGC squared	0.307	0.232	0.288	0.243	0.328	0.213
Proportion of SC explained by IGC	0.473	0.395	0.452	0.396	0.450	0.354
Controlling for Neighborhood FE			PANEL C			
Sibling Correlation (SC)	0.440	0.451	0.409	0.470	0.474	0.409
	(29.55)***	(36.93)***	(19.13)***	(26.58)***	(14.45)***	(13.17)***
Intergenerational Correlation (IGC)	0.488	0.441	0.483	0.461	0.531	0.431
	(59.91)***	(68.37)***	(46.24)***	(55.09)***	(43.78)***	(44.44)***
IGC squared	0.238	0.194	0.233	0.213	0.282	0.186
Proportion of SC explained by IGC	0.541	0.431	0.570	0.452	0.595	0.454
No. of observations	12,277	19,371	7,385	11,378	4,892	7,993

Robust t statistics in parentheses. Standard errors corrected for clustering at family level
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Sibling Correlations across Caste groups and Regions

	Upper Caste		Lower Caste	
	1993	2006	1993	2006
All children		**Panel A (Full Sample)**		
Sibling Correlation (SC)	0.644	0.596	0.610	0.592
	(86.93)***	(58.64)***	(79.55)***	(49.69)***
Intergenerational Correlation (IGC)	0.574	0.539	0.500	0.491
	(83.84)***	(86.96)***	(53.48)***	(51.09)***
Brothers				
Sibling Correlation (SC)	0.607	0.591	0.615	0.609
	(54.33)***	(40.54)***	(54.76)***	(35.07)***
Intergenerational Correlation (IGC)	0.546	0.521	0.471	0.475
	(65.65)***	(67.39)***	(42.90)***	(40.93)***
Sisters				
Sibling Correlation (SC)	0.777	0.744	0.702	0.653
	(73.29)***	(41.57)***	(48.59)***	(27.79)***
Intergenerational Correlation (IGC)	0.611	0.558	0.562	0.509
	(64.09)***	(64.09)***	(40.16)***	(37.08)***
All children		**Panel B (Rural)**		
Sibling Correlation (SC)	0.587	0.563	0.534	0.543
	(55.62)***	(48.23)***	(41.47)***	(32.66)***
Intergenerational Correlation (IGC)	0.483	0.478	0.428	0.440
	(53.06)***	(53.34)***	(37.95)***	(37.26)***
Brothers				
Sibling Correlation (SC)	0.553	0.562	0.537	0.546
	(35.41)***	(31.37)***	(29.32)***	(22.03)***
Intergenerational Correlation (IGC)	0.469	0.453	0.419	0.411
	(43.66)***	(40.88)***	(32.61)***	(28.25)***
Sisters				
Sibling Correlation (SC)	0.767	0.707	0.682	0.654
	(52.28)***	(35.78)***	(26.88)***	(20.92)***
Intergenerational Correlation (IGC)	0.517	0.516	0.468	0.482
	(37.76)***	(39.97)***	(24.70)***	(28.42)***
All children		**Panel C (Urban)**		
Sibling Correlation (SC)	0.659	0.629	0.642	0.591
	(55.03)***	(59.62)***	(33.77)***	(30.88)***
Intergenerational Correlation (IGC)	0.569	0.535	0.513	0.457
	(45.23)***	(57.84)***	(25.28)***	(26.65)***
Brothers				
Sibling Correlation (SC)	0.633	0.641	0.661	0.656
	(37.14)***	(43.50)***	(26.21)***	(27.69)***
Intergenerational Correlation (IGC)	0.557	0.542	0.478	0.480
	(37.73)***	(48.08)***	(19.76)***	(23.32)***
Sisters				
Sibling Correlation (SC)	0.722	0.656	0.769	0.561
	(33.90)***	(27.63)***	(25.88)***	(12.76)***
Intergenerational Correlation (IGC)	0.581	0.519	0.565	0.423
	(32.99)***	(40.68)***	(19.24)***	(17.31)***

Robust t statistics in parentheses. Standard errors corrected for clustering at family level

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Estimates are from regression specification that controls for age and gender (all children sample only).

Table 7: Sibling and Intergenerational Correlations: Full Sample (16-20 years and 19-24 years age cohorts)

	All children		Brothers		Sisters	
	1993	2006	1993	2006	1993	2006
			PANEL A [16-20 year]			
Sibling Correlations (SC)	0.688	0.638	0.697	0.644	0.856	0.724
	(49.19)***	(69.59)***	(31.07)***	(40.91)***	(47.49)***	(41.22)***
Intergenerational Correlations (IG)	0.575	0.530	0.531	0.510	0.631	0.552
	(78.28)***	(92.81)***	(55.90)***	(68.41)***	(59.55)***	(69.30)***
Proportion of SC explained by IGC	0.519	0.403	0.404	0.404	0.465	0.421
No. of observations	13,140	27,966	8,098	15,274	5,042	12,683
			PANEL B [19-24 year]			
Sibling Correlations (SC)	0.672	0.641	0.656	0.641	0.819	0.806
	(65.04)***	(55.97)***	(40.95)***	(36.76)***	(53.99)***	(51.78)***
Intergenerational Correlations (IG)	0.593	0.575	0.554	0.550	0.652	0.609
	(83.15)***	(77.88)***	(64.36)***	(60.98)***	(60.46)***	(54.07)***
Proportion of SC explained by IGC	0.523	0.516	0.468	0.472	0.519	0.460
No. of observations	15,182	16,003	10,196	10,275	4,986	5,728

Robust t statistics in parentheses. Standard errors corrected for clustering at family level
* significant at 10%; ** significant at 5%; *** significant at 1%
Note: Estimates are from regression specification that controls for age and gender (all children sample only).

Table 8: Sibling and Intergenerational Correlations: Full Sample (excluding singleton families)

	All children		Brothers		Sisters	
	1993	2006	1993	2006	1993	2006
Sibling Correlation (SC)	0.643	0.620	0.629	0.640	0.775	0.693
	(108.00)***	(95.57)***	(71.48)***	(67.13)***	(83.38)***	(54.51)***
Intergenerational Correlation (IG)	0.572	0.533	0.530	0.511	0.606	0.549
	(86.04)***	(75.84)***	(56.84)***	(48.48)***	(43.07)***	(36.27)***
IG squared	0.327	0.284	0.281	0.261	0.367	0.301
Proportion of SC explained by IG	0.509	0.458	0.447	0.408	0.474	0.435
No. of observations	25,833	25,681	12,637	11,194	5,384	5,494

Robust t statistics in parentheses. Standard errors corrected for clustering at family level
* significant at 10%; ** significant at 5%; *** significant at 1%
Note: Estimates are from regression specification that controls for age and gender (all children sample only).

Table 9: Intergenerational Correlations (IGC) and Intergenerational Regression Coefficients (IGRC): By Age cohorts

	Rural			Urban		
	15-19 yr	20-24yr	25-29yr	15-19 yr	20-24yr	25-29yr
2006						
IGC	0.394	0.460	0.461	0.449	0.535	0.583
	(40.33)***	(32.50)***	(20.87)***	(37.24)***	(35.30)***	(23.85)***
IGRC	0.281	0.424	0.468	0.259	0.427	0.511
	(40.33)***	(32.50)***	(20.87)***	(37.24)***	(35.30)***	(23.85)***
1993						
IGC	0.435	0.467	0.451	0.507	0.567	0.592
	(46.92)***	(37.94)***	(24.43)***	(32.52)***	(32.23)***	(20.87)***
IGRC	0.389	0.510	0.565	0.324	0.489	0.545
	(46.92)***	(37.94)***	(24.43)***	(32.52)***	(32.23)***	(20.87)***

Note: 1: Maximum of Father and Mother's schooling years. Regressions include control for caste and religion, children's age dummies and state fixed effects.
Robust t statistics in parentheses. Standard errors corrected for clustering at family level
* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1: Decomposition of Variance of Siblings' Years of Schooling: Full sample

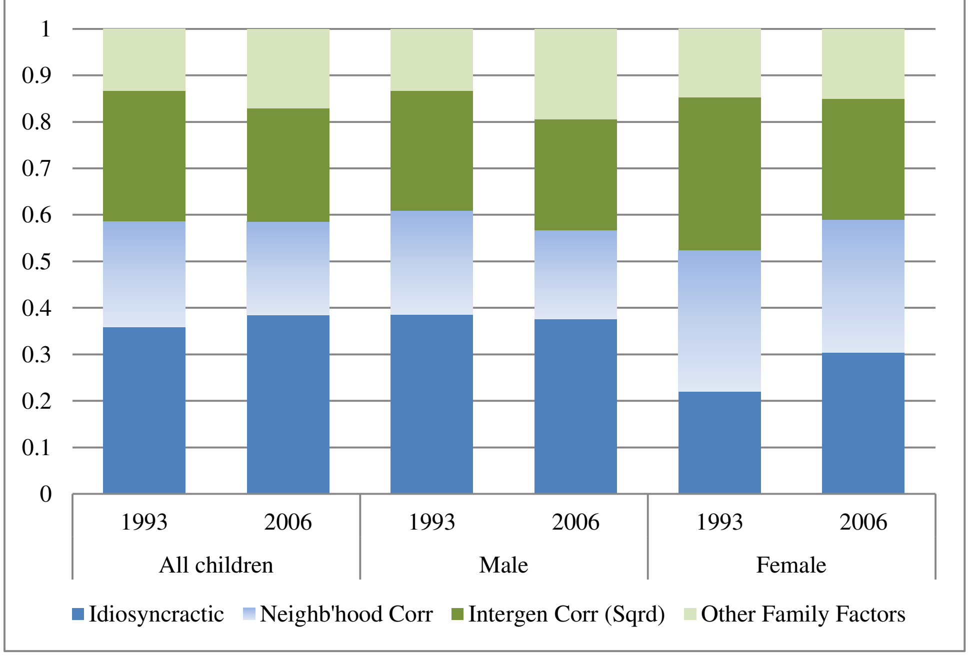

Figure 2: Variance Decomposition of Siblings Years of Schooling: Rural Sample

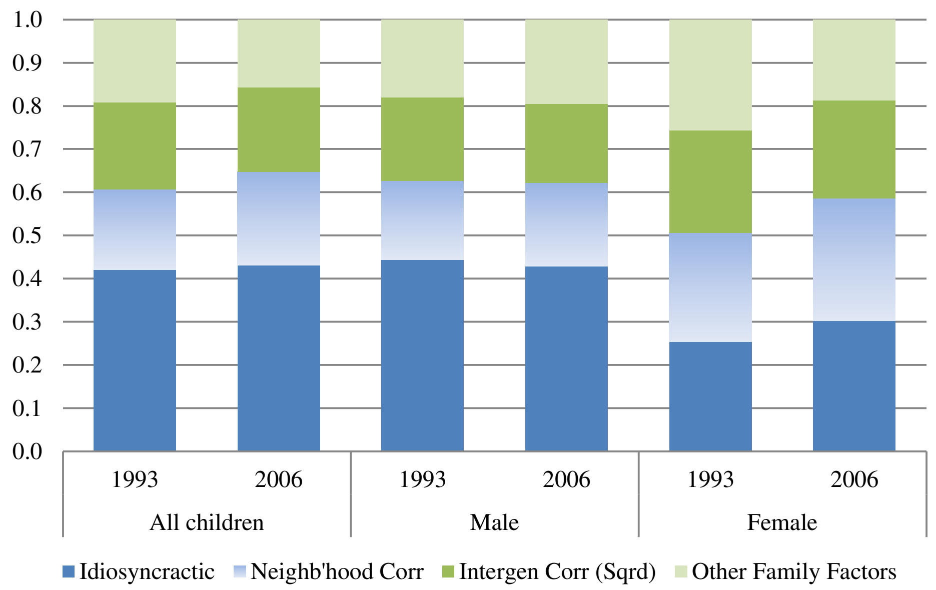

Figure 3: Variance Decomposition of Sibling's Years of Schooling: Urban Sample