Estimating Intergenerational Mobility with Incomplete Data: Coresidency and Truncation Bias in Rank-Based Relative and Absolute Mobility Measures

Emran, M. Shahe and Shilpi, Forhad

8 August 2017

Estimating Intergenerational Mobility with Incomplete Data: Coresidency and Truncation Bias in Rank-Based Relative and Absolute Mobility Measures

M. Shahe Emran [1]
IPD, Columbia University

Forhad Shilpi
DECRG, World Bank

ABSTRACT

Most of the household surveys available in developing countries suffer from sample truncation because coresidency is used to define household membership. This paper provides evidence on truncation bias in rank-based relative and absolute mobility estimates in coresident samples, and compares with the bias in intergenerational regression coefficient (IGRC) and intergenerational correlation (IGC). Using surveys from Bangladesh and India that include non-resident children, we find that the slope estimates are biased downward, while the intercept estimates are biased upward, but expected years of schooling conditional on parental education are overestimated in coresident samples. The downward bias in rank correlation is much smaller than that in IGRC, and comparable to that in IGC. The upward bias in the intercept is the largest in the regression used for IGC. Truncation bias in rank-based absolute mobility estimates is the lowest in most cases. The results add to an emerging body of evidence that the rank-based measures are more robust than the widely-used IGRC and IGC in estimating intergenerational mobility with incomplete data.

Key Words: Rank correlation, Rank-based Absolute Mobility, Coresidency, Truncation Bias, Intergenerational Mobility, Developing Countries, Intergenerational Regression Coefficient (IGRC), Intergenerational Correlation (IGC), Bangladesh, India

JEL Codes: O12, I32

Introduction

Intergenerational persistence in economic status and, more broadly, the role of family background in determining economic opportunities of a child have been the focus of a large literature in economics and sociology (for excellent surveys of economics literature, see Solon (1999), Black and Devereux (2011), Bjorklund and Salvanes (2011), and for the sociological literature, see Erikson and Goldthorpe (2002), and Fox et al. (2016)). There has been a surge in the interest in reliable estimates of intergenerational persistence in economic status, motivated by the evidence that economic inequality has increased in recent decades in many countries (World Bank (2006)). The worry is that the observed increase in inequality is symptomatic of a worsening deep-seated inequality of opportunities in education and labor market. There is growing evidence of a Great Gatsby curve, showing that cross-sectional inequality and intergenerational mobility are negatively correlated (Corak (2013), Fan et al. (2015), Neidhöfer (2016)).

A major constraint on the research on intergenerational mobility in developing countries has been data limitations, most notably, measurement error and truncated samples. That measurement error and transitory income shocks can cause strong downward bias in the estimates of intergenerational income persistence is well-established in the context of developed countries (Solon (1992), Mazumder (2005)). Since most of the household surveys available in developing countries are cross-section or panel with a limited number of years, it is difficult, if not impossible, to estimate permanent income of parents and children. In the face of such paucity of income data, the literature has focused on education and occupation as indicators of economic status. In most of the household surveys in developing countries, only the children who are coresident with parents at the time of survey are included, resulting in sample truncation. The implications of truncation of the sample as a result of coresidency restrictions for estimating intergenerational persistence have, however, not been adequately considered in the literature.

A small but growing literature analyzes how different measures of intergenerational per-

sistence used in the literature are affected by data limitations. In a widely-acclaimed paper, Chetty et al. (2014) show that the standard measure of intergenerational income persistence, called intergenerational income elasticity (IGE) and estimated as the slope of a regression of log children's income on log parental income, is fragile.[2] The lack of robustness in the log-linear model reflects the fact that the conditional expectation function of children's log income given parent's log income is highly nonlinear. They show that, in contrast, the rank correlation, estimated as the slope of a regression of children's percentile rank in income distribution on parental percentile rank, is approximately linear, and it is much more robust to measurement error and alternative treatments of zero income.[3] Dahl and Deleire (2008) and Mazumder (2015) also find that the attenuation bias in the rank-rank slope estimate is significantly smaller compared to that in intergenerational elasticity (IGE). Nybom and Stuhler (2016) use rich data from Sweden to analyze life-cycle bias and attenuation bias in IGE and rank based measures. They find that the rank-based measures are more robust to both life-cycle bias and attenuation bias when compared to the estimates of IGE.

Emran, Greene and Shilpi (forthcoming) show that intergenerational regression coefficient (IGRC), the most widely used measure of intergenerational mobility in development literature, suffers from strong downward bias due to truncation as a result of coresidency criteria used to define household membership in a survey. IGRC is estimated as the slope of a level-level regression, usually for intergenerational schooling persistence. The truncation bias, however, is much less (less than one third) in another widely used measure called intergenerational correlation (IGC) which estimates Pearson correlation between economic status of parents and children. Using British Panel Household Survey, Francesconi and Nicoletti (2006) report 12% - 39% downward bias due to coresidency in short panels when estimating persistence in Hope-Goldthorpe index of occupational prestige. To the best of our knowledge, there is no evidence in the literature

---

[2]For example, it is affected substantially by how one treats the households with zero taxable income.

[3]The rank-rank specification was introduced in economic literature by Dahl and Deleire (2008).

on how the rank-based measures of relative and absolute mobility are affected by truncation because of coresidency restrictions in a survey.

Taking advantage of two exceptionally rich household surveys from India and Bangladesh, we provide evidence on the effects of sample truncation caused by coresidency restrictions on the rank-based measures of relative and absolute mobility. Our analysis focuses on intergenerational schooling persistence, as education has been the main indicator of economic status in most of the recent research on intergenerational mobility in developing countries (Azam and Bhatt (2015), Emran and Shilpi (2015), Maitra and Sharma (2010), Jalan and Murgai (2008), Hertz et al. (2007), Nimubona, A, and D. Vecatachellum (2007), Behrman et al. (2001), Thomas (1996)). The focus on educational mobility is appropriate for at least two reasons. First, as noted above, data are not available to construct reliable estimates of permanent income, and schooling attainment is likely to be a good proxy for life-time income and economic status in most developing countries.[4] Second, there is a broad consensus among policy makers and economists that education is key to social mobility in an increasingly skill-driven economy (Stiglitz (2012), Rajan (2010), The Economist (2012), World Development Report (2006)).

The evidence reported below shows that truncation, in general, results in *downward* bias in the estimate of the *slope* (relative mobility) and *upward* bias in the estimate of the *intercept* of widely-used intergenerational persistence equations including the rank-rank regression. When the focus is on relative mobility, the bias in rank correlation is small and similar in magnitude to that in IGC, but the bias in IGRC is much larger. Truncation bias in the intercept estimate is, in general, the largest in the regression specification used to estimate IGC (i.e., Pearson correlation), while the bias in the intercept of rank-rank regression is, in most cases (26 out of 30), the smallest. Children's expected years of schooling (or expected normalized schooling in the case of IGC specification) conditional on parental schooling are overestimated in coresident

---

[4]Blanden (2013) in her review of cross-country evidence on intergenerational mobility concludes that "....the results for earnings and education tend to be fairly well correlated; this implies that information on educational mobility is a good proxy for earnings mobility in countries where earnings information is not readily available."

samples, implying that the upward biased estimate of the intercept dominates the downward bias in the slope estimate. When the focus is on absolute mobility, the rank-based measure a la Chetty et al. (2014) outperforms an alternative measure based on the slope and intercept of IGRC or IGC equation in most of the cases: the truncation bias in rank-based absolute mobility estimate is lower in 8 out of 10 cases in both Bangladesh and India.

The evidence in this paper suggests that when working with coresident samples, a researcher is, in general, better off relying on rank-based measures of relative and absolute mobility. For relative mobility, the performances of IGC and rank correlation are, however, broadly comparable, and it is advisable to report both rank correlation and IGC. The evidence presented below is especially important for developing countries where data that include all the children irrespective of residency at the time of a survey are rare. The results can be helpful for many researchers who, following the lead of Chetty et al. (2014), have adopted the rank-based measures to understand intergenerational mobility in developing countries (see, for example, Fan et al. (2015), Asher and Novosad (2017)).

The rest of this paper is organized as follows. Section 2 provides a discussion on truncation due to coresidency in household surveys in developing countries. The next section discusses the standard measures of relative and absolute mobility widely used in the literature. Section (4) is devoted to data and variables definitions. The estimates of truncation bias for different samples are reported and discussed in section (5). Section (6) provides a brief discussion on possible intuitions behind the results. The paper concludes with a summary of the findings and the implications for research on intergenerational mobility in developing countries.

(2) Coresidency, Household Definition, and Sample Truncation in Household Surveys in Developing Countries

The availability and quality of household surveys in developing countries have improved substantially over last few decades. For example, starting from early 1980s, the Living Standards

Measurement Study has generated high quality household surveys in more than 40 countries, while the Demographic and Health Surveys (DHS) cover more than 90 countries. Although there is no uniformity in the definitions of 'household' across different surveys, almost all of the surveys use coresidency criteria to define household membership such as 'living together', 'eating together', and sometimes 'pooling of funds' (United Nations (1989), Deaton (1997)). For example, according to Tanzania DHS, "a household is defined as a person or group of persons, related or unrelated who live together and share a common source of food". Most LSMS surveys include an individual as household member only if she/he lived in the household for more than 3 months in the last 12 months (Glewwe (2000)), which would exclude the children in colleges away from home, as the college students do not get 3 months break to go back home.[5] Most household surveys done by national statistical agencies also use similar coresidency restrictions to define a household.[6]

The coresidency restrictions used in the surveys result in truncation of the sample as there is no information on the nonresident children, and truncation is likely to be nonrandom.[7] In developing countries such as India and Bangladesh,, many girls move out of parental household when they drop out of school because of marriage, and the sons leave the household for jobs. In addition, the rural households are likely to miss more educated children systematically because most of the villages do not have a college (or even a high school), and children have to migrate to pursue higher education.

Figure 1 provides estimates of conditional probability of non-residence of children at the time of the survey at various schooling levels of children. Both in India and Bangladesh the probability

---

[5]There are exceptions such as infants less than 3 months old are included as household member. For an excellent discussion on the issues involved in defining a household, please see Glewwe (2000).

[6]Some panel surveys carefully track the households over the years, and thus take care of the sample selection arising from attrition. However, most of them still use coresidency criteria to define household membership at the baseline. Thus even with no attrition, the data still suffer from truncation bias.

[7]As noted by Emran, Greene and Shilpi (forthcoming), this implies that it is not possible to implement selection correction using Heckman procedure, as it is not possible to estimate a selection equation without knowing which households are missing children from the survey. The maximum likelihood procedure developed by Bloom and Killingsworth (1985) can be used, but it relies on strong distributional assumption.

of non-residence is much lower at the middle of the schooling distribution. Truncation thus is more prominent at the tails of the schooling distribution. There are also significant differences across countries and gender. The incidence of non-residency at the time of the survey is lower in India in general compared to Bangladesh, and the non-residency rates are higher for girls (daughters) in both countries. The gender gap in non-residency rate is much larger in the case of India, and in Bangladesh the gender gap is higher at the bottom of the schooling distribution and effectively vanishes at the top of children's schooling distribution.

(3) Measures of Intergenerational Mobility: Relative and Absolute

The most widely used measure of intergenerational mobility is regression-based relative mobility which is estimated as the slope of a log-log (for income) or level-level (for schooling) OLS regression. The standard regression model for intergenerational persistence in schooling is:

$$S_i^c = \beta_0 + \beta_1 S_i^p + \varepsilon_i \tag{1}$$

where S_i^c and S_i^p are indicators of educational attainment for child i and his/her parents respectively.[8] We do not include additional controls such as age of the child and parents to ensure comparability to the recent work on relative and absolute mobility based on rank-rank regression such as Chetty et al. (2014). The focus is on estimating the parameter β_1 which is called Intergenerational Regression Coefficient (IGRC).[9] We report summary of a sample of recent papers on intergenerational mobility on developing countries in Table 1; most of the papers use IGRC as the measure of intergenerational mobility, and the most common indicator of economic status is education.

---

[8]Most of economic literature in the context of developed countries use income as the indicator, while education is the most widely used indicator in studies on developing countries. Economic analysis of intergenerational persistence in occupation has been relatively neglected; for recent contributions, see Emran and Shilpi (2011) on Nepal and Vietnam, Azam (2015), Ahsan and Chatterjee (2016), and Iversen et al. (2016) on India, and Emran and Sun (2013) on China.

[9]The literature on intergenerational persistence in income uses a log-linear model, and thus the slope provides an estimate of intergenerational income elasticity (IGE).

A second widely used measure of relative mobility is estimated from the following OLS regression:

$$\frac{S_i^c}{\sigma_c} = \rho_0 + \rho_1 \left(\frac{S_i^p}{\sigma_p} \right) + \epsilon_i \tag{2}$$

where the indicators of educational attainment such as years of schooling are normalized by their standard deviation, i.e., σ_c and σ_p are standard deviations of children's and parent's schooling respectively. The focus in most of the literature is on estimating the parameter ρ_1 which provides an estimate of Pearson correlation between educational attainment of two generations. It is called intergenerational correlation (IGC) in the literature.

An increasingly popular measure of relative mobility is rank correlation, used originally by Dahl and DeLeire (2008), and made salient by the recent work of Chetty et al. (2014). The rank correlation is estimated from the following OLS regression:

$$R_i = \delta_0 + \delta_1 P_i + \zeta_i \tag{3}$$

where R_i is the percentile rank of child i in the distribution of children's schooling, and P_i is the percentile rank of parent of child i in the distribution of parental schooling. The parameter of interest is δ_1 which provides an estimate of Spearman rank correlation in schooling across generations.

We compare the bias in the estimates of the slope and intercept terms in the rank-rank regression in equation (3) with the bias estimates for the slope and intercept from equations (1) and (2). As noted before, Emran et al. (forthcoming) provide estimates of coresidency bias in IGC and IGRC using the same data sets on India and Bangladesh used here. However, the regression specification used in that paper includes quadratic age controls for both children and parents, following Solon (1992). Thus their estimates are not comparable to the rank-rank estimates without any controls in this paper; we report estimates of $\hat{\beta}_1$ and $\hat{\rho}_1$ without any

controls for comparability.[10]

The different measures of relative mobility discussed above have been the preferred ones in economics literature for analyzing intergenerational mobility in both developed and developing countries. However, as emphasized by Chetty et al . (2014), relative mobility measures have an important limitation: an improvement in relative mobility may be driven by a worsening of outcomes for children at the upper tail of the distribution rather than an improvement for the children at the lower tail of the distribution. To address this issue they propose measures of absolute mobility based on the estimates of δ_0 and δ_1 in equation (3) above. Chetty et al. (2014) define absolute mobility at percentile p as the expected percentile rank of the children whose parents belong to percentile p in the distribution of parental schooling, i.e., $P_i = p$. Denote the OLS estimates of the parameters by a hat $\hat{}$, then the absolute mobility at the percentile p is calculated as below:

$$\bar{r}_p = \hat{\delta}_0 + \hat{\delta}_1 p \tag{4}$$

Following Chetty et al. (2014), we focus on absolute mobility at $P_i = p = 25$. With linear conditional expectation function, $\bar{r}_{25} = E\left(R_i \mid P_i \leq 50\right)$, i.e., the expected rank of children born to parents who fall in the lower half of parental schooling distribution, and thus is a measure of upward mobility. In addition, we also report absolute mobility at the $75th$ percentile of parental schooling distribution which shows the expected rank of children born to parents in the upper half of parental schooling distribution.

One can also derive expected years of schooling (normalized schooling) for children using the intercept and slope estimates from equations (1) and (2) above. Denote the expected years of schooling for child i as $\hat{S}_i^c$, then we have (corresponding to equation (1) above):

$$\hat{S}_i^c = \hat{\beta}_0 + \hat{\beta}_1 S_i^p \tag{5}$$

---

[10]These estimates also provide additional robustness checks for their central conclusion that IGC is preferable to IGRC as a measure of relative mobility in coresident samples.

Using expected years of schooling $\hat{S}_i^c$ from equation (5) above, we calculate expected schooling attainment of the children from lower half of parental schooling distribution as:

$$\bar{S}_{25}^c = E\left(\hat{S}_i^c \mid S_i^p \ni P_i \leq 50\right) \tag{6}$$

To get a measure comparable to the absolute mobility measure defined by Chetty et al. (2014) at 25th percentile, we calculate the percentile rank of $\bar{S}_{25}^c$ in the schooling distribution of children. This provides the expected rank of the children from lower half of parental schooling distribution when expected schooling attainment is determined by equation (6) above. We denote this rank as R_{25}^β to convey that $\hat{\beta}_0$ and $\hat{\beta}_1$ were used to derive this rank estimate. Note that the expected normalized schooling of children ((i.e., normalized by the standard deviation) and expected years of schooling of children are monotonically related to each other, and thus the rank remains same. This implies that if we calculate rank based on the expected normalized schooling by using OLS estimates of parameters of equation (2), i.e., $\hat{\rho}_0$ and $\hat{\rho}_1$, it will be identical to R_{25}^β, i.e., $R_{25}^\beta = R_{25}^p$.

(4) Data and Variables: India and Bangladesh

To estimate the bias caused by truncation arising from the coresidency rules in surveys, we need data sets that include all of the children of household head (and spouse) irrespective of their residency status, and also need to identify which members of the household were coresident at the time of the survey implementation. We take advantage of two exceptionally rich data sets particularly suited for our analysis. The data on India come from the 1999 Rural Economic and Demographic Survey done by the National Council for Applied Economic Research, and the data on Bangladesh are from the 1996 Matlab Health and Socioeconomic Survey (MHSS).[11]

The Bangladesh survey collected information on all children of household head and spouse (including from past marriages) irrespective of their residency status from 4538 households in

---

[11]The MHSS 1996 is a collaborative effort of RAND, the Harvard School of Public Health, the University of Pennsylvania, the University of Colorado at Boulder, Brown University, Mitra and Associates and the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B).

Matlab thana of Chandpur district. The India survey also collected information on all of household head's children from current marriage, but did not gather information on non-coresident mothers of children from earlier marriage(s). We use these information to create data sets containing education and other individual characteristics of parents and children. Both of these surveys focus on rural areas in respective countries. The bias from censoring due to possible non-completion of younger children may not be as important in rural areas, because the proportion of children who go on to have more than middle school (or high school) education is not likely to be large. The children who go for more than high school education (10 years of schooling in Bangladesh and India) are also the children who leave the village household, because the "colleges" (for grades 11 and 12) and universities (for three-four year undergraduate, and graduate study) are located in the cities and large towns.

Our estimation sample consists of household head and spouse, and their children, including those from other marriages in the case of Bangladesh. Our main empirical analysis is based on a sample of children aged 13-60 years. To test the sensitivity of our conclusions with respect to the specific age cutoffs, we estimate the rank correlation and absolute mobility at 25th percentile for a number of alternative age ranges; 16-60, 20-69 and 13-50 years.

Table 2 reports the summary statistics for both the Bangladesh and India data for our main estimation sample (children in the age range 13-60 years). Several interesting observations and patterns emerge. The average schooling attainment is low in rural areas of both of the countries at the time of the survey years. The mean and median years of schooling for children are 4.97 and 5.00 respectively for Bangladesh, and 6.23 and 7 for India. The relatively lower education attainments in Bangladesh compared with India are also observed for parent's generation: median years of father's schooling was 2 years in Bangladesh compared with 2.50 years in India. The average number of children per household in Bangladesh is about 5.74 compared with 3.53 in India. This difference probably partly reflects the fact that Bangladesh data include informa-

10

tion on children from other marriages while India data do not. The age distribution of children also differs: the median age for Bangladesh data is 30 years compared with 33 years for India. The gender gap in education between boys and girls is about 1 year in Bangladesh in contrast with 2 years in India.

In a study of intergenerational schooling persistence, one can define parental schooling in a number of different ways: some researchers use average schooling of father and mother as the relevant indicator, while others use maximum of parents education, which in most cases amounts to father's schooling in many developing countries, as historically the husband is older and with more schooling. Also, many existing studies focus on the sample of sons, and use father's schooling as the relevant indicator of parental human capital. We use three indicators of parental education: father's years of schooling, mother's years of schooling, and average years of schooling for father and mother. The percentile rank for father is estimated from the distribution of father's schooling, and similarly for mothers, while the percentile rank for average parental education is calculated using the distribution of average schooling.

(5) Estimates of Truncation Bias in Rank-Based Measures of Intergenerational Mobility

We begin with the estimates of relative mobility, i.e., the estimates of slope parameters, and then report estimates of the intercept terms, as it is necessary to estimate absolute mobility. Next we discuss the the expected years of schooling (and normalized schooling) given parental schooling. Finally, we report and discuss truncation bias in the measures of absolute mobility.

The estimates from three different samples are provided: all children sample (includes both sons and daughters), the father-sons sample, and the mother-daughter's sample. As noted before, our main estimates are based on the age range 13-60 years for children. We provide estimates for different age ranges later to check robustness of the conclusions. In all of the rank-rank regressions, the dependent variable (the indicator of educational attainment) is children's

percentile rank in the distribution of years of schooling for all children.

Define the bias in the estimate of parameter $\omega \in \left[\hat{\delta}_0, \hat{\delta}_1, \hat{\beta}_0, \hat{\beta}_1, \hat{\rho}_0, \hat{\rho}_1 \right]$ as PB_ω (short for percentage bias in the estimate of ω):

$$PB_\omega = \left[\frac{Estimate\ of\ \omega\ from\ Coresident\ Sample - Estimate\ of\ \omega\ from\ Full\ Sample}{Estimate\ of\ \omega\ from\ Coresident\ Sample} \right] \times 100 \tag{7}$$

Thus when truncation causes downward biased estimate, then the estimated bias in definition (7) above is negative, and it is positive when the estimate of a parameter from the coresident sample is biased upward.[12] The advantage of using the estimate from coresident sample as the base in equation (7) above is that it is directly observable to most of the researchers facing data constraints with access to only coresident sample.

(5.1) Truncation Bias in Estimates of Relative Mobility

(5.1.1) Estimates from the All Children Sample

The estimates of rank correlation $\hat{\delta}_1$ are reported in Table 3; the estimates of IGC $\hat{\rho}_1$ and IGRC $\hat{\beta}_1$ are also included for comparison. The top panel contains the estimates for Bangladesh and the lower panel for India. The odd columns show the estimate from full sample (including nonresident children) and the even columns the estimates from coresident sample. The first row reports the estimates for the case when father's schooling is the indicator of parental education, the third row for mother's schooling, and the fifth row for average of father's and mother's years of schooling. The estimated percentage biases are reported in the even numbered rows.

The point estimate of rank correlation $\hat{\delta}_1$ from the coresident sample is *smaller* than that from the full sample (including nonresident children) across all three different indicators of parental education for both Bangladesh and India. This is consistent with the *a priori* expectation that

---

[12]This definition of bias is different from that in Emran et al. (forthcoming). We adopted this definition because unlike the analysis in Emran et al. (forthcoming), we have to deal with both positive and negative biases in this paper.

truncation causes downward bias in the estimate of the slope parameter(s) of a OLS regression (Hausman and Wise (1977), Cohen (1991)). More striking is the fact that the differences between the two estimates are very small in magnitude across the board. For example, the estimates for father's schooling as the indicator of parental education in Bangladesh are 0.523 (full sample) and 0.483 (coresident sample). This implies that the downward bias in rank correlation estimate due to coresidency is only -0.04, which amounts to $PB_{\hat{\delta}_1} = -8.31\%$ according to formula (7) (see the bias estimates in row 2).[13] To appreciate the order of magnitudes involved, it is instructive to compare this estimate to the corresponding estimate of truncation bias in the most widely used measure of intergenerational persistence in development literature, i.e., IGRC; the bias in IGRC is almost four times as large at $PB_{\hat{\beta}_1} = -29.40\%$ (see row 6). The extent of coresidency bias in the rank correlation estimate varies somewhat with different indicators of parental education, with the bias being highest when percentile rank of mother's schooling is the indicator (-11.36%). The pattern of bias in rank correlation estimates across different indicators of parental education is similar in India. The average percentage bias across three indicators of parental education in Bangladesh is -8.31% and larger in India at -13.19%. The average bias in IGRC is -29.39% in Bangladesh and -20.42% in India.

The results in columns 3-6 show that the IGC estimates suffer less truncation bias compared to IGRC, a conclusion established before in Emran et al. (forthcoming) using a different regression specification that includes quadratic controls of parents and children's age. The results presented here provide a robustness check that the conclusion does not depend on the inclusion of additional controls in the specification. The more interesting evidence in Table 3 is that the extent of truncation bias in the rank correlation is, in general, close to that in IGC. When compared to the bias in IGC $(\hat{\rho}_1)$, the estimated bias is lower in rank correlation $\left(\hat{\delta}_1\right)$ in three

---

[13]The estimates from full and coresident samples are statistically different from each other as the estimated standard errors are very small; the intersection of the 95 percent confidence intervals is a null set. However, the statistical precision and formal rejection of equality of the estimates are not informative in our context, as the differences in magnitudes are very small.

out of six cases in Table 3. The average bias across six estimates for India and Bangladesh is -10.75% in $\left(\hat{\delta}_1\right)$, -9.69% in $(\hat{\rho}_1)$, and -24.91% in $\left(\hat{\beta}_1\right)$. The estimates thus suggest that rank correlation and IGC (Pearson correlation) are practically in a tie in coresident samples, but the most widely used measure IGRC is clearly the most biased by the truncation due to coresidency criteria.

(5.1.2) Estimates of Father-Sons and Mother-Daughters Intergenerational Persistence

Many studies focus on the intergenerational persistence in economic status along gender lines; there is evidence that the intergenerational linkages between father-sons and mother-daughters may be stronger than the cross-gender effects. In this subsection, we discuss the truncation bias in the estimates of intergenerational persistence between schooling of fathers and sons, and between mothers and daughters. Table 4 presents the estimates for Bangladesh (top panel) and India (lower panel) using our main estimation sample of 13-60 years of age range for the children.

Consistent with the evidence from all children sample in Table 3, the estimates in table 4 show that the bias in $\hat{\beta}_1$ (IGRC) is significantly higher compared to the bias in $(\hat{\rho}_1)$ and $\left(\hat{\delta}_1\right)$. A comparison of the estimates of $(\hat{\rho}_1)$ and $\left(\hat{\delta}_1\right)$, however, shows that the conclusion depends on the gender: for the father-sons intergenerational link, the estimated bias in rank correlation is somewhat smaller than that in Pearson correlation. In contrast, for the mother-daughter link, the bias in estimated rank correlation is larger in magnitude compared to the bias in Pearson correlation coefficient. The results in Table 4 also show that the estimates of mother-daughter schooling persistence in general suffer more severe downward bias compared to the estimates for father-sons; the difference in the bias estimates between father-sons and mother-daughters is substantial even for the rank correlation estimates. The larger bias in daughters sample probably reflects the fact that coresidency rates are much lower for them both in Bangladesh and India. In the father-son sample the coresidency rate is 79 percent in India, while the corresponding

14

rate is 52 percent in Bangladesh. In the mother-daughter samples, the coresidency rates are much lower: 39 percent in India and 26 percent in Bangladesh, reflecting the fact that women leave the natal family following marriage in both countries. The evidence thus suggests that it is important to report estimates of both $(\hat{\rho}_1)$ and $\left(\hat{\delta}_1\right)$ to understand the father-sons and mother-daughters links in schooling attainment.

(5.1.3) Estimates from Alternative Age Ranges

The empirical results discussed so far are based on our "main sample" composed of children in the age range 13-60 years. To check if the conclusions above are specific to this sample, we report estimated bias in the slope parameters of equations (1)-(3) from two alternative samples: 13-50 years, and 20-60 years. Table 5A reports the estimated bias from 13-50 years sample, and Table 5B from the 16-60 years sample. We omit the underlying estimates from full and coresident sub-samples which allows us to put the estimated bias from all children sample, and father-sons, and mother-daughters sub-samples in a single Table. The estimates for Bangladesh are in the first three columns, and those for India are in the last three columns in each table.

Consider the bias estimates from 13-50 years age sample in Table 5A. The estimates for the slope confirm the conclusion that the bias in the estimates of $\left(\hat{\delta}_1\right)$ and $(\hat{\rho}_1)$ are consistently smaller that that in $\left(\hat{\beta}_1\right)$ and this is true in both Bangladesh and India. A comparison of the bias estimates for the rank correlation with those for IGC shows that the bias in IGC is smaller in six out of ten cases. However, judged in terms of average bias, IGC and rank correlation are close to each other: the average bias across India and Bangladesh over different indicators of parental educations is -9.4% for $(\hat{\rho}_1)$ and -10.01% for $\left(\hat{\delta}_1\right)$. The estimates form 16-60 years sample also lead to the same set of conclusions. Averaging over the 20 estimates for 13-50 years and 16-60 years age ranges, the bias estimates are 8.48% for rank correlation, 8.38% for IGC, and 21.74% for IGRC. Thus the evidence cannot discriminate between IGC and rank correlation, but both are clearly much better than the IGRC as a measure of relative mobility. In the light

15

of the evidence, it is advisable to report both rank correlation and IGC when the focus is on relative mobility.

(5.2) Truncation Bias in Absolute Mobility Estimates

(5.2.1) Bias in The Estimates of the Intercept

All of the studies on intergenerational mobility in developing countries listed in Table 1 rely exclusively on some measure(s) of relative mobility estimated as the slope of a OLS regression, and do not consider the intercepts in equations (1), (2), and (3) above. However, as noted earlier, it is important to understand the biases in the estimated intercepts, as absolute mobility depends on both the slope and intercept of the regression equations. Table 6 reports the estimates of the intercept of equations (1)- (3), i.e., $\hat{\delta}_0$, $\hat{\beta}_0$ and $\hat{\rho}_0$ for the age range 13-60 years (similar to Table 3). The corresponding estimates for 13-50 years age range are in Table 7A, and those for 16-60 years age range are in Table 7B. Again, the estimates for Bangladesh are in the top panel and those for India in the bottom panel, and the odd columns show the estimates from full sample and the even columns the estimates from coresident sample. The estimates from all children sample are in the top panel, and those for father-sons, and mother-daughters are in the lower panel.

Consider the estimates for all children sample in Table 6. The pattern of the estimates is exactly the opposite of that found in the estimates of the slope parameter in first two columns of Table 3 above; the estimate from coresident sample is consistently *higher* when compared to the corresponding estimate from the full sample across all three indicators of parental education, and this holds for both Bangladesh and India. Again, more important for our analysis is the evidence that the estimates from full and coresident samples differ by small magnitudes in the case of the intercept of rank-rank regression $\hat{\delta}_0$.

For comparison, we turn to the estimates of the bias in the intercept terms from the equations (1) and (2) above $\hat{\beta}_0$ and $\hat{\rho}_0$, i.e., $PB_{\hat{\beta}_0}$ and $PB_{\hat{\rho}_0}$. The first thing to notice is that the bias

estimates for the intercept term is positive across the board for both $\hat{\beta}_0$ and $\hat{\rho}_0$, reinforcing the conclusion from the estimates of $\hat{\delta}_0$ that truncation leads to *upward* bias in the estimated intercept term. This is consistent with the classic analysis of sample truncation by Hausman and Wise (1977): while truncation leads to downward bias in the estimate of the slope in an OLS regression, it also, in general, results in upward bias in the estimate of the intercept. The second important point that comes across clearly from the bias estimates in Table 6 is that the upward bias is significantly higher across the board in $\hat{\beta}_0$ and $\hat{\rho}_0$ when compared to the bias in $\hat{\delta}_0$, with the extent of bias largest in $\hat{\rho}_0$. In Bangladesh, the average bias estimated across three indicators of parental education are 6.54% for $\hat{\delta}_0$, 15.91% for $\hat{\beta}_0$, and 25.08% for $\hat{\rho}_0$. The corresponding average bias estimates for India are 8.78% for $\hat{\delta}_0$, 16.37% for $\hat{\beta}_0$, and 18.05% for $\hat{\rho}_0$.

The results on father-sons and mother-daughters samples in the lower panel of Table 6 are, however, somewhat different, as the conclusion depends on the gender. For mother-daughter the twin conclusions that the estimate is upward biased in coresident sample and the degree of bias is lowest for the intercept of rank-rank regression hold. However, these conclusions are not valid for father-sons estimates. The estimates for alternative age ranges show that the different results for father-sons is sample-specific; the twin conclusions regarding direction and magnitude of bias noted above remain valid for father-sons estimates in 13-50 years age range, but are violated in the 16-60 years age range. The important take away from the results on the bias in intercept estimates in Tables 6, 7A, and 7B is that, in most cases (26 out of 30), the bias is lowest in the intercept of rank specification.

(5.2.2) Truncation Bias in Absolute Mobility: Expected Schooling and Expected Schooling Rank

Following Chetty et al. (2014) we combine the estimates of the slope and intercept of the rank-rank regression and report estimates of absolute mobility using equation (4) above. Since

truncation due to coresidency restrictions causes *downward* bias in the slope estimate, but, at the same time, leads to an *upward* biased estimate of the intercept in general, one might conjecture that the bias in the absolute mobility is likely to be smaller than that in the estimates of relative mobility because of offsetting effects. It is, however, important to appreciate that the bias in the intercept may dominate the estimates at the lower tail, while the bias in slope is likely to be more consequential for the estimates at the upper tail, and in general, it is not possible to know the direction of net bias at a given percentile of parental schooling. Following Chetty et al. (2014), many recent studies on developing countries focus on absolute mobility at 25th percentile. In addition to 25th percentile, we also provide estimates of absolute mobility at 75th percentile. For equations (1) and (2), we begin with the estimates of expected years of schooling (normalized years of schooling in case of equation (2)). We discuss the results from the main sample (13-60 years of age) below. The conclusions from other age ranges are similar, and the estimates are omitted for the sake of brevity.

Expected Years of Schooling (and Expected Normalized Schooling)

Tables 8A and 8B present the results for 25th and 75th percentile of parental schooling distribution; the estimates provide the average expected years of schooling or average expected normalized schooling for the subset of children whose parental schooling belongs to a certain percentile. Table 8A reports the estimated expected years of schooling (using $\hat{\beta}_0$, $\hat{\beta}_1$) and percentage bias in coresident samples, while Table 8B reports the corresponding estimates for normalized schooling (using $\hat{\rho}_0$, $\hat{\rho}_1$).

The first point to notice is that the estimates of expected schooling for 25th percentile are same as the estimated intercepts $\hat{\beta}_0$, $\hat{\rho}_0$, which reflects the fact that almost 50 percent parents have zero schooling in our data sets. Thus the results on truncation bias in the intercepts discussed above in subsection (5.2.1) imply that the expected years of schooling at 25th percentile are likely to be substantially overestimated in the coresident samples, for both $\hat{\beta}_0$ and $\hat{\rho}_0$, and this

is true in both Bangladesh and India. Also, the extent of upward bias is the largest if one relies on $\hat{\rho}_0$. This evidence suggests that even though IGC ($\hat{\rho}_1$) is a robust measure of relative mobility when working with coresident samples (Emran, Greene, and Shilpi (forthcoming)), estimates of expected schooling attainment using equation (2) may be the least reliable at the lower tail of the distribution.

The estimates of expected years of schooling for children show that the estimates from coresident samples are consistently larger than those from the full sample at the 25th percentile, and at the 75th percentile, it is true in 9 out of 10 cases.[14] The evidence thus suggests that the upward bias in the intercept dominates the downward bias in the slope estimate. However, the magnitude of bias does not exhibit any consistent pattern across the 25th and 75th percentiles of parental distribution. In Bangladesh, the truncation bias (absolute value of the bias) at 75th percentile is smaller than that st 25th percentile in four out of five cases: the average bias is 7.94 percent for 75th percentile, while it is 15.58 percent for 25th percentile. In India, in 2 out of 5 cases, the bias is higher at the 75th percentile : the average bias estimates are 11.57 percent for 75th and 13.36 percent for 25th percentile.

The estimates of expected normalized schooling (i.e., years of schooling normalized by standard deviation) also tells a similar story. The estimates are consistently upward biased in coresident samples, but the magnitude of bias does not follow a consistent pattern across 25th and 75th percentiles. A comparison of the biases in the expected years of schooling to that in normalized schooling shows that the average bias is substantially higher in the estimates of normalized schooling when compared to expected years of schooling in Bangladesh. In the case of India the pattern holds, i.e., the average bias is lower in expected years of schooling, but the magnitudes are much closer, at 25th percentile: 13.36% (years of schooling) and 14.58% (normalized schooling), and at 75th percentile: 11.57% (years of schooling) and 12.9% (normalized

---

[14]The only exception at the 75th percentile is the expected years of schooling for sons conditional on father's schooling. However, the estimates from coresident sample (6.666) and full sample (6.767) are very close to each other.

schooling).

Expected Rank of a Child

Tables 9 presents the estimates of absolute mobility using the estimated $\hat{\delta}_0$, $\hat{\delta}_1$ at 25th percentile of parental schooling distribution, following the definition of Chetty et al. (2014). We also calculate the rank of expected schooling of children whose parental schooling belongs to 25th percentile to have a comparable measure of absolute mobility (using equation (5)). As noted before, the ranking of a child's expected schooling in the distribution of children's schooling does not vary between equations (1) and (2) by construction. Thus we focus on the rank of the expected years of schooling from equation (5) in the distribution of children's schooling. In addition to absolute mobility at 25th percentile, we also report absolute mobility at 75th percentile. The bias estimates for Bangladesh are in the top panel, and those for India in the lower panel.

The results show that the bias is smaller in the rank-based absolute mobility estimate in 8 out of 10 cases both in Bangladesh and India. At the 25th percentile of parental schooling, the average bias (absolute magnitudes ignoring the signs) in rank-based absolute mobility ($\bar{r}_{25}$) estimates from coresident sample over different specifications in Bangladesh is 5.67% , and is 13.32% for the alternative measure based on the predicted schooling using equation (5) above, i.e., $\left(R_{25}^{\beta}\right)$. The corresponding estimates at 75th percentile of parental distribution are 2.44 percent for ($\bar{r}_{25}$) and 6.66 percent for $\left(R_{25}^{\beta}\right)$. The evidence from India is similar: the average bias at 25th percentile is 3.44% in rank-based absolute mobility estimates, and 10.34% in the IGRC-based absolute mobility estimates. The corresponding average bias estimates at the 75th percentile for India are 2.54% (rank-based) and 8.1% (IGRC-based). The evidence thus is strong that absolute mobility estimates based on rank-rank specification are much more robust to the truncation bias arising in coresident samples.

20

(6) Discussion

The evidence reported and discussed above suggests that the rank-rank specification originally used by Dahl and Deleire (2008) and made popular in the recent literature by the influential work of Chetty et al. (2014) performs better in estimating intergenerational persistence in economic status in coresident samples. When the focus is on relative mobility, the evidence in this paper along with the evidence presented earlier in Emran, Greene and Shilpi (forthcoming) leads to two conclusions: (i) IGRC is affected the most by truncation due to coresidency among the three measures of mobility, (ii) the biases in rank correlation and Pearson correlation are relatively small and similar in magnitude. Our results also show that the estimate of the intercept term is, in general, biased upward in the coresident sample (Hausman and Wise, 1977), and this bias is the largest in $\hat{\rho}_0$. When estimating absolute mobility with coresident data, the rank-rank specification clearly outperforms the alternative based on regression specification used to estimate IGRC or IGC.

Emran, Greene, Shilpi (forthcoming) discuss a rationale for the observed lower bias in IGC $(\hat{\rho}_1)$ in coresident samples when compared to the bias in IGRC $\left(\hat{\beta}_1\right)$. They point out that IGC cancels out some of the downward truncation bias in IGRC by normalizing children's schooling by a downward biased estimate of standard deviation. The rank correlation takes an additional step to purge the effects of changing marginal distributions across generations by focusing on the copula of the bivariate distribution of parents' and children's schooling. An explanation for the less sensitivity of rank correlation estimates to truncation can be developed in terms of the fact that truncation tends to exclude observations from the tails of a distribution. It has long been understood that rank correlation is less sensitive to outliers, because rank-ordering pulls the outlying observations more towards the center of the distribution (Lehmann (1975), Shevlyakov and Oja (2016), Bishara and Hittner (2012)). Since observations from the tails of the distribution are lost because of truncation, the effect on standard OLS is likely to be strong,

but rank correlation is more robust to inclusion (exclusion) of these tail observations in (from) the sample.

The pattern of bias found in the estimates across countries, samples, and gender suggests that researchers (economists and sociologists) working with coresident samples should utilize both rank correlation and Pearson correlation (IGC) for understanding relative mobility. The conclusion in Emran et al. (forthcoming) that IGC is preferable as a measure of relative mobility to the most widely used measure IGRC remains valid, but the results presented in this paper also show that the truncation bias is much smaller in rank correlation when compared to IGRC.

The finding that the intercept estimate from the IGC specification (equation (2) above), in general, contains the largest upward bias can be explained by observing that $\rho_0 = \dfrac{\beta_0}{\sigma_c}$. It is widely known that, in general, truncation biases the estimate of $\hat{\beta}_0$ upward in the IGRC regression equation (1) (Hausman and Wise (1977), Cohen (1991)). However, as noted by Greene (2012), truncation also biases the estimate of variance of children's schooling downward, i.e., the estimate of $\hat{\sigma}_c$ is biased downward in coresident samples. This implies that the upward bias in the estimate of $\hat{\rho}_0$ is higher that that in the estimate of $\hat{\beta}_0$.

The magnitude of truncation bias in the estimates of absolute mobility based on a rank-rank regression $\bar{r}_{25}$ and $\bar{r}_{75}$ is small, and substantially smaller than that in the alternative based on equation (5), i.e., R^{β}_{25} and R^{β}_{75}. The evidence suggests that it is unlikely that the estimates of $\bar{r}_{25}$ and $\bar{r}_{75}$ in coresident samples will result in incorrect conclusions regarding the trend and pattern of absolute mobility. This implies that the large number of household surveys available through LSMS and DHS where non-resident children are not included can be profitably used for understanding intergenerational mobility if researchers focus on rank correlation and IGC as measures of relative mobility and rely on rank-based absolute mobility measure.

The growing evidence on the robustness of rank-based measures of mobility, however, raises the question what we are missing when we ignore the measures based on level-level or log-linear

specifications of intergenerational persistence equation. As discussed above, compared to IGRC (or IGE in the case of income persistence), IGC and rank correlation progressively purge the effects of changes in marginal distributions across generations. Many authors argue that this is desirable as this allows a researcher to focus on the fundamental structure of dependence between parents and children. For example, Bjorklund and Jantti (2009) point out that IGC provides a measure of mobility that is not affected mechanically by changes in inequality across generations. But others such as Mazumder (2015) and Mittnik et al. (2014) argue that IGRC or IGE are valuable precisely because they contain information about the marginal distributions and relate to cross sectional inequality directly. Thus it is important to appreciate that a focus on rank correlation may limit our ability to understand the interactions between intergenerational mobility and cross-sectional inequality.

Conclusions

Following the influential contribution of Chetty et al. (2014), many researchers working on intergenerational mobility in developing countries have adopted the rank-based measures. There is an emerging body of evidence that rank-based measures are more reliable as they are less sensitive to measurement error and life-cycle biases. We focus on the implications of a common data limitation faced by researchers working on developing countries: most of the available data sets suffer from truncation as household membership is defined in terms of a set of coresidency criteria. Are rank-based measures also more robust to truncation bias due to coresidency when compared to the more standard measures of intergenerational mobility such as IGRC and IGC? This paper provides an empirical analysis of the effects of truncation on the estimates of rank-based relative and absolute mobility.

We utilize two exceptionally rich household surveys from Bangladesh and India where information on non-resident children was collected, and the subset of children coresident at the time of survey implementation was identified. The evidence shows that truncation results in down-

ward bias in the slope estimate and upward bias the intercept estimate both in Bangladesh and India. The truncation bias in rank correlation is relatively low and similar in magnitude to that in intergenerational correlation (IGC), and the downward bias in coresident samples is substantially higher in the estimates of IGRC. The evidence thus suggests that a researcher working with coresident sample should report both rank correlation and IGC to understand relative mobility. Our results also show that the magnitude of bias in rank-based absolute mobility proposed by Chetty et al (2014) is usually small, and in most cases, suffer significantly less from truncation bias compared to alternative measures based on years of schooling. This paper thus strengthens the case for rank-based measures for analyzing intergenerational mobility in developing countries where most of the available data sets, including LSMS and DHS, suffer from sample truncation due to coresidency restrictions used to define household membership.

REFERENCES

Ahsan, Reshad & Arpita Chatterjee, (2015). Trade Liberalization and Intergenerational Occupational Mobility in Urban India, Discussion Papers 2015-23, School of Economics, The University of New South Wales.

Asher, S, and P. Novosad (2017), Estimating Intergenerational Mobility with Coarse Data: A Nonparametric Approach, Working Paper, World Bank.

Azam, Mehtabul (2015). "Intergenerational Occupational Mobility among Men in India," Journal of Development Studies, 51:10, 1389-1408.

Azam, M and V. Bhatt, 2015. "Like Father, Like Son? Intergenerational Educational Mobility in India," Demography, Springer;Population Association of America (PAA), vol. 52(6), pages 1929-1959, December.

Behrman, J., A. Gaviria and M. Szekely (2001), "Intergenerational Mobility in Latin America," Economia, Vol. 2 (1): 1 44.

Blanden, Jo (2013). "Cross-Country Rankings In Intergenerational Mobility: A Comparison

Of Approaches From Economics And Sociology," Journal of Economic Surveys, Wiley Blackwell, vol. 27(1), pages 38-73

Bishara, A. J., & Hittner, J. B. (2012). "Testing the significance of a correlation with non-normal data: Comparison of Pearson, Spearman, transformation, and resampling approaches." Psychological Methods, 17, 399-417.

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

Björklund, A. and M. Jäntti (2009) "Intergenerational mobility and the role of family background" in W. Salverda, B. Nolan and T. Smeeding (eds) Oxford Handbook of Economic Inequality, Oxford University Press.

Black, S. E. and P. Devereux (2011). Recent Developments in Intergenerational Mobility, Handbook of Labor Economics, Amsterdam, North-Holland.

Black, Sandra E., Paul J. Devereux and Kjell G. Salvanes (2005), Why the apple does not fall far: Understanding intergenerational transmission of human capital," American Economic Review 95: 437-449.

Bloom, David E., and Killingsworth, Mark R. (1985), "Correcting for Truncation Bias caused by a Latent Truncation Variable," Journal of Econometrics, 1985, pp. 131-135.

Chetty, R, and N. Hendren, P. Kline, and E. Saez, (2014). "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States," The Quarterly Journal of Economics, Oxford University Press, vol. 129(4), pages 1553-1623.

Cohen, A (1991), Truncated and Censored Samples: Theory and Applications, CRC Press.

Corak, M (2013). "Income Inequality, Equality of Opportunity, and Intergenerational Mobility," Journal of Economic Perspectives, American Economic Association, vol. 27(3), pages 79-102, Summer.

Corak, M and A. Heisz (1999), The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data, Journal of Human Resources. Volume 34, Number 3 (Summer), pages 504-533.

Dahl, Molly W., and Thomas DeLeire, (2008), The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data, Institute for Research on Poverty, University of Wisconsin–Madison, 2008.

Deaton, A (1997), The analysis of household surveys: A microeconometric approach to development policy. Oxford University Press.

Emran, M Shahe, William Greene and F. Shilpi (forthcoming), When Measure Matters: Coresidency, Truncation Bias, and Intergenerational Mobility in Developing Countries, Forthcoming in Journal of Human Resources.

Emran, M. Shahe and F. Shilpi (2011). "Intergenerational Occupational Mobility in Rural Economy: Evidence from Nepal and Vietnam", Journal of Human Resources, issue 2, 2011.

Emran, M. Shahe and F. Shilpi (2015). Gender, Geography and Generations : Intergenerational Educational Mobility in Post-reform India", World Development, Vol. 72, 362-380.

Emran, M. Shahe and Yan Sun (2015), Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China: 1988-2002, World Bank Policy Research Working Paper 7316.

Fan, Y, J. Yi and J. Zhang (2015), The Great Gatsby Curve in China: Cross-Sectional Inequality and Intergenerational Mobility, Working Paper, CUHK, Hongkong.

Fox, L, Florencia Torche, and Jane Waldfogel (2016), Intergenerational Mobility, Oxford Handbook of the Social Science of Poverty.

Glewwe, P (2000), Household Roster, Chapter 6 in Grosh, M and P. Glewwe ed. Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study, vol 1.

Greene, W (2012), Limited Dependent Variables - Truncation, Censoring, and Sample Selection, Chapter 19, Econometric Analysis, Pearson.

Hausman, J and D. Wise (1977), "Social Experimentation, Truncated Distribution, and Efficient Estimation," Econometrica, May 1977.

Hertz Tom, T. Jayasundera, P. Piraino, S. Selcuk, N. Smith and A. 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.

Iversen, V, and A. Krishna (2016), Rags to riches? Intergenerational occupational mobility in India, Working paper, Manchester University.

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

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

Mazumder, Bhashkar (2005), "Fortunate Sons: New Estimates of Intergenerational Mobility in U.S. Using Social Security Earnings Data," Review of Economics and Statistics, May, 2005.

Mazumder, Bhashkar, (2015). "Estimating the Intergenerational Elasticity and Rank Association in the U.S.: Overcoming the Current Limitations of Tax Data,", Working Paper Series WP-2015-4, Federal Reserve Bank of Chicago.

Neidhofer, G (2016), Intergenerational Mobility and the Rise and Fall of Inequality: Lessons from Latin America, Working Paper.

Nimubona, A, and D. Vecatachellum (2007), "Intergenerational education mobility of black and white South Africans,", Journal of Population Economics.

Nybom, Martin, and J. Stuhler, (2017). "Biases in standard measures of intergenerational income dependence," Journal of Human Resources, Summer, 2017.

Rajan, R (2010), Fault Lines: How Hidden Fractures Still Threaten the World Economy,

Princeton University Press.

Shevlyakov, G, and Hannu Oja (2016), Robust Correlation: Theory and Applications, Wiley Series in Probability.

Solon, Gary, 1992. "Intergenerational Income Mobility in the United States," American Economic Review, American Economic Association, vol. 82(3), pages 393-408, June.

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.

Stiglitz, J (2012), The Price of Inequality, W. W. Norton and Company.

The Economist (2012), "For Richer, For Poorer", Special Report on Inequality by Zanny Minton Beddoes, October 13th 2012.

Thomas, D (1996). "Education across Generations in South Africa", American Economic Review, American Economic Association, vol. 86(2), pages 330-34, May.

United Nations. 1989. "Household Income and Expenditure Surveys: A Technical Study.", National Household Survey Capability Programme, New York.

World Development Report (2006), Equity and Development, Oxford University Press.

Table 1: Intergenerational Mobility in Developing Countries: Data, Status Indicator and Measures

STUDY	COUNTRY Status Indicator	DATA SAMPLE	MEASURES
Thomas (1996)	South Africa Education	Cross-section 1991/93 Coresident Children	IGC
Behrman and Wolfe (1987)	Nicaragua Education	Single Cross-section: 1977/78 Coresident Children of Sisters	IGRC
Jalan and Murgai (2008)	India Education	Cross-sections: 92/93, and 98/99 Coresident Children	IGRC
Maitra and Sharma (2010)	India Education	Single Cross-Section: 2005 Coresident Children	IGRC
Azam and Bhatt (2015)	India Education	Single Cross-section: 2005 Coresident Sons +Nonresident Sons in College '+ Head's Parents But does not include nonresident sons due to work migration, and household split	IGRC and IGC
Emran and Shilpi (2015)	India Education	Cross-sections: 1993, 2006 Coresident Children	IGC and Sibling Corr. Transition Matrices
Emran and Sun (2011)	China Education Occupation	Cross-sections: 1995, 2002 Coresident Children ` + Head's and Spouses Parents ` + nonresident members with financial links to the household	IGRC
Fan, Yi, Zhang (2015)	China Income, Education	Cross-section, 2010 Coresident + Nonresident Children	IGRC, IGC and Rank-Rank
Hertz et al. (2007)	21 Developing Countries Education	Cross-section, various years Coresident Children + Head's and Spouse's Parents	IGRC and IGC
Nimubona and Vencatachellum (2007)	South Africa Education	Repeated Cross-section (Pseudo-Panel) Coresident Children	IGRC
Lillard and Willis (1995)	Malaysia Education	Cross-Section, 1988 One coresident and up to 2 nonresident children randomly selected	IGRC
Lam and Schoeni (1993)	Brazil Earnings	Cross-section Household Head's and Spouse's Parents	IGRC
Behrman et al (2001)	Brazil, Colombia, Mexico, Peru Education	Cross-section (various years) Coresident Children ' + Head's and Spouse's Parents	IGRC, Transition Matrices

Notes: IGRC Stands for Intergenerational Regression Coefficient, IGC for Intergenerational Correlation, Sibling Corr. for Sibling Correlation, and Rank-Rank for Spearman Rank Correlation.

Table 2: SUMMARY STATISTICS

	ALL CHILDREN			CO-RESIDENT CHILDREN		
	Mean	Median	N	Mean	Median	N
	(1)	(2)	(3)	(4)	(5)	(6)
			BANGLADESH			
Years of Education of			**Both Sons and Daughters Sample**			
Children	4.97	5.00	18587	5.52	5.00	5852
Father	3.39	2.00	14017	3.74	3.00	5599
Mother	1.46	0.00	14527	1.81	0.00	5523
Average of Parents	2.33	1.00	18505	2.78	2.00	5806
			Sons Sample			
Children	5.84	5.00	9056	5.56	5.00	3873
Father	3.38	2.00	7126	3.53	2.00	3713
Mother	1.45	0.00	7261	1.64	0.00	3648
Average of Parents	2.34	1.00	9010	2.59	1.50	3844
			Daughters Sample			
Children	4.14	4.00	9531	5.44	5.00	1979
Father	3.41	2.00	6891	4.16	3.00	1886
Mother	1.47	0.00	7266	2.14	0.00	1875
Average of Parents	2.33	0.50	9495	3.14	2.50	1962
			INDIA			
Years of Education of			**Both Sons and Daughters Sample**			
Children	6.23	7.00	14877	6.97	8.00	9132
Father	4.37	2.50	14877	4.74	5.00	9132
Mother	1.83	0.00	14877	2.12	0.00	9132
Average of Parents	3.10	2.50	14877	3.43	2.50	9132
			Sons Sample			
Children	7.29	8.00	8341	7.54	8.00	6561
Father	4.31	2.50	8341	4.59	5.00	6561
Mother	1.82	0.00	8341	1.99	0.00	6561
Average of Parents	3.06	2.50	8341	3.29	2.50	6561
			Daughters Sample			
Children	4.87	5.00	6536	5.54	6.00	2571
Father	4.46	2.50	6536	5.14	5.00	2571
Mother	1.84	0.00	6536	2.45	0.00	2571
Average of Parents	3.15	2.50	6536	3.79	3.25	2571

Notes: Data Sources: India: Rural Economic and Demographic Survey (REDS) 1999; Bangladesh: Matlab Health and Socioeconomic Survey 1996.

Table 3: Truncation Bias in Relative Mobility: All Children Sample

	Rank Correlation		IGC		IGRC	
	Full	**Coresident**	**Full**	**Coresident**	**Full**	**Coresident**
BANGLADESH						
Father's Schooling	0.523	0.483	0.506	0.459	0.546	0.422
Bias	-8.31%		-10.4%		-29.40%	
Mother's Schooling	0.548	0.492	0.465	0.421	0.842	0.616
Bias	-11.36%		-10.3%		-36.67%	
Average Schooling	0.531	0.504	0.50	0.47	0.71	0.58
Bias	-5.26%		-5.15%		-22.09%	
INDIA						
Father's Schooling	0.448	0.409	0.439	0.397	0.483	0.419
Bias	-9.42%		-10.8%		-15.36%	
Mother's Schooling	0.453	0.382	0.367	0.318	0.569	0.454
Bias	-18.77%		-15.5%		-25.52%	
Average Schooling	0.467	0.420	0.456	0.403	0.654	0.543
Bias	-11.39%		-13.1%		-20.37%	

Notes: (1) The sample consists of daughters and sons of 13-60 years age. (2) IGC stands for Intergenerational correlation, IGRC stands for intergenerational regression coefficient. (3) Average schooling is the average of father's and mother's schooling. (4) Bias is defined as Percentage Bias=
 [(Coresident sample estimate - Full sample estimate)/Coresident sample estimate] * 100

Table 4: Bias in Relative Mobility for Father-Sons and Mother-Daughters

		Rank Correlation		IGC		IGRC	
		Full	Coresident	Full	Coresident	Full	Coresident
BANGLADESH							
Father-Sons		0.516	0.499	0.488	0.449	0.565	0.444
	Bias	**-5.19%**		**-8.61%**		**-27.03%**	
Mother-Daughter		0.577	0.505	0.529	0.490	0.845	0.597
	Bias	**-14.27%**		**-8.10%**		**-41.60%**	
INDIA							
Father-Sons		0.426	0.421	0.420	0.409	0.457	0.427
	Bias	**-1.26%**		**-2.65%**		**-7.22%**	
Mother-Daughters		0.549	0.481	0.466	0.428	0.688	0.571
	Bias	**-14.18%**		**-8.70%**		**-20.44%**	

Notes: (1) The samples consist of daughters only (mother-daughters) and sons only (father-sons) for the 13-60 years age range. (2) IGC stands for Intergenerational correlation, IGRC stands for intergenerational regression coefficient. (3) Bias is percentage bias defined in Table 3.

Table 5A: Bias in Relative Mobility: Estimates from 13-50 Years Age Sample

	BANGLADESH			INDIA		
	Rank Corr.	IGC	IGRC	Rank Corr.	IGC	IGRC
All Children Sample						
Father's Schooling	-8.27%	-10.30%	-29.22%	-9.76%	-11.10%	-15.73%
Mother's Schooling	-11.16%	-10.20%	-36.27%	-18.96%	-15.70%	-25.60%
Average Schooling	-5.14%	-5.09%	-21.86%	-11.65%	-13.37%	-20.58%
Father-Sons	-5.05%	-8.46%	-26.72%	-1.63%	-3.04%	-7.54%
Mother-Daughters	-14.16%	-8.05%	-41.44%	-14.28%	-8.68%	-20.37%

Table 5B: Bias in Relative Mobility: Estimates from 16-60 Years Age Sample

	BANGLADESH			INDIA		
	Rank Corr.	IGC	IGRC	Rank Corr.	IGC	IGRC
All Children Sample						
Father's Schooling	-6.14%	-7.7%	-21.23%	-6.72%	-9.8%	-13.47%
Mother's Schooling	-7.63%	-7.5%	-27.67%	-17.33%	-15.2%	-24.55%
Average Schooling	-1.4%	-1.26%	-11.13%	-9.09%	-12.54%	-19.09%
Father-Sons	-5.1%	-7.26%	-21.47%	-0.58%	-2.41%	-6.88%
Mother-Daughters	-8.14%	-4.54%	-30.84%	-7.3%	-5.34%	-13.17%

NOTES: (3) The numbers in the table are the percentage bias as defined in Table 3. (2) IGC provides estimate of Pearson correlation, IGRC is intergenerational Regression Coefficient. (3) Average schooling is the average of mother's and father's schooling.

Table 6: Truncation Bias in the Intercepts: 13-60 Years Age Sample

	(δ_0)		(ρ_0)		(β_0)	
	Full	**Coresident**	**Full**	**Coresident**	**Full**	**Coresident**
BANGLADESH						
All Children Sample						
Father's Schooling	0.247	0.26	0.741	1.016	3.225	3.955
Bias		**4.79%**		**27.05%**		**18%**
Mother's Schooling	0.232	0.256	0.869	1.134	3.813	4.435
Bias		**9.52%**		**23%**		**14.04%**
Average Schooling	0.234	0.247	0.751	0.999	3.303	3.896
Bias		**5.32%**		**25%**		**15.22%**
Father-Sons	0.297	0.262	0.847	0.980	3.994	3.999
Bias		**-13.38%**		**14%**		**1.37%**
Mother-Daughters	0.170	0.237	0.770	1.200	3.002	4.218
Bias		**28.39%**		**35.82%**		**28.83%**
INDIA						
All Children Sample						
Father's Schooling	0.276	0.295	0.838	1.037	4.113	4.987
Bias		**6.53%**		**19.18%**		**17.53%**
Mother's Schooling	0.273	0.309	1.056	1.25	5.185	6.012
Bias		**11.58%**		**15.49%**		**13.76%**
Average Schooling	0.266	0.290	0.855	1.062	4.199	5.110
Bias		**8.23%**		**19.48%**		**17.83%**
Father-Sons	0.350	0.326	1.097	1.179	5.321	5.581
Bias		**-7.32%**		**6.95%**		**4.67%**
Mother-Daughters	0.146	0.164	0.776	0.878	3.603	4.137
Bias		**11.06%**		**11.68%**		**12.90%**

NOTES: (1) (δ_0) is the intercept of rank-rank, (ρ_0) is the intercept of IGC regression, and (β_0) is the intercept of IGRC regression.

Table 7A: Truncation Bias in the Intercept: Estimates from 13-50 Years Age Sample

	BANGLADESH			INDIA		
	(δ_0)	(ρ_0)	(β_0)	(δ_0)	(ρ_0)	(β_0)
All Children Sample						
Father's Schooling	4.93%	26.97%	18.44%	6.76%	19.44%	17.78%
Mother's Schooling	9.27%	23.14%	13.91%	11.71%	15.63%	13.90%
Average Schooling	5.19%	24.67%	15.11%	8.43%	19.69%	18.03%
Father-Sons	-13.28%	13.69%	1.40%	-7.01%	7.24%	4.97%
Mother-Daughters	28.62%	35.64%	28.62%	10.96%	11.76%	12.92%

Table 7B: Truncation Bias in the Intercept: Estimates from 16-60 Years Age Sample

	BANGLADESH			INDIA		
	(δ_0)	(ρ_0)	(β_0)	(δ_0)	(ρ_0)	(β_0)
All Children Sample						
Father's Schooling	2.66%	28.31%	23.40%	5.11%	20.98%	20.40%
Mother's Schooling	6.81%	24.45%	19.29%	11.61%	17.07%	16.47%
Average Schooling	1.59%	24.63%	19.53%	7.15%	21.40%	20.83%
Father-Sons	-18.5%	13.37%	4.02%	-11.01%	7.38%	5.61%
Mother-Daughters	27.42%	37.30%	35.71%	0.41%	6.80%	12.94%

NOTES: (1) (δ_0) is the intercept of rank-rank, (ρ_0) is the intercept of IGC regression, and (β_0) is the intercept of IGRC regression. (2) The reported numbers are the percentage bias as defined in Table 3.

Table 8A: Bias in Expected Years of Schooling Conditional on Parental Schooling Rank

	BANGLADESH		INDIA	
	P25	P75	P25	P75
All Children Sample				
Father's Schooling	18.46%	8.19%	17.53%	4.31%
Mother's Schooling	14.04%	8.13%	13.76%	20.19%
Average Schooling	15.22%	5.67%	17.83%	9.30%
Father-Sons	1.37%	-1.52%	4.67%	0.16%
Mother-Daughters	28.83%	16.17%	12.90%	23.88%

Table 8B: Bias in Expected Normalized Schooling Conditional on Parental Schooling Rank

	BANGLADESH		INDIA	
	P25	P75	P25	P75
All Children Sample				
Father's Schooling	27.1%	17.9%	19.2%	6.2%
Mother's Schooling	23.3%	18.1%	15.5%	21.8%
Average Schooling	24.8%	16.4%	19.5%	11.1%
Father-Sons	13.7%	11.2%	7%	2.6%
Mother-Daughters	35.8%	24.4%	11.7%	22.8%

NOTES: (1) P25 is the 25th percentile of parental education rank defined in terms of a given indicator such as father's schooling. P75 is similarly defined. (2) Normalized schooling is years of schooling divided by its standard deviation. (3) The numbers reported are the percentage bias as defined in Table 3.

Table 9: Truncation Bias in Absolute Mobility: 13-60 Years Age Range

	P25		P75	
	Rank-based	**IGRC-based**	**Rank-based**	**IGRC-based**
		BANGLADESH		
All Children Sample				
Father's Schooling	-0.7%	-22%	0.2%	1.90%
Mother's Schooling	-2.2%	4.7%	-2.4%	-8.1%
Average Schooling	-6%	-25.9%	-1.4%	-8.7%
Father-Sons	-12.9%	-3.9%	-5.7%	4.2%
Mother-Daughters	7.2%	-10.1%	2.5%	-10.4%
		INDIA		
All Children Sample				
Father's Schooling	1.60%	-25.5%	-3.8%	5.70%
Mother's Schooling	0.9%	-10.3%	1.2%	16.1%
Average Schooling	2.40%	-0.1%	-1%	5.70%
Father-Sons	-7.8%	-8.8%	-6.4%	-3.7%
Mother-Daughters	-4.5%	-7%	-0.3%	-9.3%

NOTES: (1) Rank-based absolute mobility is defined following Chetty et al. (2014). The IGRC-based absolute mobility is the expected rank of the predicted years of schooling using equation (5) in the actual schooling distribution of children. (2) The reported numbers are the percentage bias estimates as defined in Table 3.

Figure 1a: Probability of non-residency in Bangladesh

Figure 1b: Probability of non-residency in India