The Journal of Economic Inequality (2023) 21:743–773
https://doi.org/10.1007/s10888-023-09599-1

Occupational dualism and intergenerational educational mobility in the rural economy: evidence from China and India

M. Shahe Emran[1] · Francisco H. G. Ferreira[2] · Yajing Jiang[3] · Yan Sun[4]

Received: 19 August 2023 / Accepted: 2 October 2023 / Published online: 30 November 2023

Abstract

We extend the Becker-Tomes model to a rural economy with farm-nonfarm occupational dualism to study intergenerational educational mobility in rural China and India. Using data free of coresidency bias, we find that fathers' nonfarm occupation and education were complementary in determining sons schooling in India, but separable in China. Sons faced lower mobility in India irrespective of fathers' occupation. Sensitivity analysis using the Altonji et al. (J. Polit. Econ. **113**(1), 151–84, 2005) approach suggests that genetic correlations alone could explain the intergenerational persistence in China, but not in India. Farm-nonfarm differences in returns to education, and geographic mobility are plausible mechanisms behind the contrasting cross-country evidence.

Keywords Educational mobility · Rural economy · Occupational dualism · Farm-nonfarm · Complementarity · Coresidency Bias · China · India

1 Introduction

Intergenerational persistence in economic status in developing countries has attracted the attention of policymakers and researchers in recent years, partly in response to growing evidence that economic liberalization increased income inequality in many countries, despite a significant reduction in poverty.[1] In the absence of reliable income data over the life-cycle

[1] Intergenerational persistence refers to the association between children's and their parents' lifetime economic outcomes. A stronger intergenerational persistence implies a lower intergenerational mobility. For recent contributions on intergenerational mobility in China, see, among others, Fan et al. (2021); Park and Zou (2019); Sato and Li (2007); Emran and Sun (2015a); Emran et al. (2020b); on India see, among others, Azam and Bhatt (2015); Emran and Shilpi (2015); Asher et al. (2018); Ahsan et al. (2022); Emran et al. (2021). For cross-country analysis, see Behrman (2019); Hertz et al. (2007), and Neidhofer et al. (2018), among others. Among the few contributions on intergenerational persistence in health, see Bhalotra and Rawlings (2013).

✉ Francisco H. G. Ferreira
f.d.ferreira@lse.ac.uk

M. Shahe Emran
shahe.emran.econ@gmail.com

1 IPD, Columbia University, New York, NY, USA

2 London School of Economics, London, UK

3 Charles River Associates, Boston, MA, USA

4 World Bank, District of Columbia, DC, USA

of parents and children, the focus of the recent literature on developing countries has been on intergenerational educational mobility.[2] However, most of the recent studies are devoted to urban households, and intergenerational mobility in rural areas remains particularly under-researched, even though the bulk of the poor live in villages in developing countries. This paper provides a comparative analysis of intergenerational educational mobility in the rural areas of the two most populous countries in the world: China and India, with more than 1.5 billion people living in villages in 2000/2001.[3]

While understanding the role of family background in the educational opportunities of 1.5 billion people is an important research agenda in itself, the policy differences between China and India make such a comparative study especially interesting.[4] Perhaps, the most important difference in the 1970s to 1990s, the period during which the children in our empirical analysis went to school, was the restrictions on rural-urban migration in China because of the Hukou registration system.[5] In contrast, there were no policy restrictions on rural-urban migration in India. Returns to education in farm vs. nonfarm occupations are also likely to be different in rural China because of policies such as the household responsibility system, rural industrialization (township and village enterprises), and a lack of a well-functioning labor market in the 1970s to 1990s.[6] Another important aspect in which rural India and rural China differed in this period is their schooling systems, with private schools playing a much more prominent role in rural India. In our analysis, we pay close attention to the implications of these cross-country differences.

The standard model of intergenerational educational mobility where parent's education is the sole indicator of family background is not suitable for our analysis because it ignores occupational dualism between farm and nonfarm sectors in the rural economy.[7] Children born into a nonfarm household may face different educational opportunities compared to the children born into a farming household even when the parents have similar educational background. At a given level of education, the parent's nonfarm occupation may affect investment in children's education through two major channels. First, higher income from nonfarm occupations may relax binding credit constraints on investment in schooling.[8] Second, the

[2] Educational mobility in this paper relates to schooling attainment. For a discussion on the broader concept of education in and outside the classroom as input to human capital production, see, for example, Behrman (2019).

[3] The rural population in India was 742.62 million in 2001 (72.16 percent of the total), according to the census of India 2001. In China, the rural population in 2000 was 807.39 million (64 percent of the total), according to the China Statistical Yearbook, 2011.

[4] China opened up its economy gradually from 1978, and the rural economy went through substantial policy changes starting with the household responsibility system that introduced market incentives at the margin. India liberalized its economy with a big bang in 1991, but the rural economy was relatively less affected as the initial trade protections were focused on the industrial sector as part of a planned industrialization strategy.

[5] The Hukou registration system was implemented in the early 1950s in China. It is like an internal passport system dividing the rural and urban population. This restricted rural-urban migration because a migrant without an urban Hokou could not access the formal jobs, and their children were not eligible for the public schools in a town or city, among other restrictions. Until recently, children inherited the Hukou status (rural vs. urban) of their mother.

[6] The household responsibility system introduced market at the margin at the early phase of liberalization after 1978 reform. A household was allowed to sell the surplus agricultural output to the market after meeting the quota requirement by the government.

[7] Although there is a long tradition in development economics following Lewis (1954) that emphasizes dualism as a central feature of underdevelopment, to the best of our knowledge, ours is the first paper to incorporate occupational dualism in an analysis of intergenerational educational mobility.

[8] However, in some countries, low-skilled nonfarm activities might be occupations of last resort, with low income, lower than the income of the farming households (Lanjouw and Lanjouw 2001).

probability of a child getting a nonfarm job may be higher when the parents themselves are employed in nonfarm occupations because of network, referral, and role model effects (Emran and Shilpi 2011), and this would increase the optimal investment when returns to education are higher in nonfarm occupations.

There is substantial evidence that structural change in favor of non-farm occupations is an important source of increasing income inequality in villages of many developing countries.[9] According to the estimates of Lanjouw et al. (2013) based on the data from Palanpur in the Indian state of Uttar Pradesh, the contribution of non-farm income to income inequality was only 4 percent in 1974/75, which increased to 67 percent in 2008/09.[10] The evidence on China also suggests that nonfarm income contributes to income inequality in rural areas (Rozelle 1994; Yang and An 2002). Such a rise in inequality associated with the expansion of the nonfarm sector is of special concern when it reflects lower intergenerational mobility.

We develop a theoretical model in the tradition of Becker and Tomes (1986) that incorporates the role played by parental farm and nonfarm occupations in shaping children's educational opportunities, and yields the almost universally used linear-in-levels estimating equation.[11] As emphasized by Mogstad (2017), in the absence of an explicit theoretical model, it is difficult to understand and interpret the economic content of the estimated intergenerational persistence. The theoretical analysis identifies a set of economic mechanisms determining the intercept and the slope of the intergenerational educational persistence equation. In particular, the household-level returns to education in the parental generation determine relative mobility as measured by the slope of the regression function (called intergenerational regression coefficient (IGRC, for short) in the literature). The intercept which shows the expected schooling attainment of the children from the most disadvantaged family background (fathers with no schooling) is determined by intergenerational persistence in occupation, among other factors.[12] A substantial literature shows that Hukou restrictions affected intergenerational persistence in nonfarm occupations in rural China (Wu and Treiman 2007).

A credible empirical analysis of the role of occupational dualism in intergenerational educational persistence needs to address two major challenges highlighted in the recent literature: (i) truncation bias due to coresidency restrictions in household surveys, and (ii) the role of genetic correlations in the observed intergenerational persistence. Most of the available household surveys, especially in developing countries, suffer from serious sample truncation as household membership is defined in terms of coresidency.[13] As a result, the

9 For an excellent survey of the literature on rural nonfarm economy in developing countries, see Lanjouw and Lanjouw (2001). On rural China see Rozelle (1994); Yang and An (2002), and on rural India see Lanjouw et al. (2013), and Ravallion and Datt (2002).

10 The Gini coefficient of income in Palanpur was 0.253 in 1974/75 and 0.427 in 2008/09.

11 The estimating equation used in the literature on intergenerational income mobility is log linear. In contrast, all of the papers on intergenerational educational mobility we are aware of use a linear-in-levels estimating equation. This is partly motivated by the fact that, in many developing countries, 20-40 percent of fathers have zero schooling. However, we are not aware of any published work on developing countries that derives the estimating equation from a theoretical model.

12 This deserves especial attention because the focus in much of the existing literature has been on the slope-based measures of mobility such as the intergenerational regression coefficient (IGRC) and intergenerational correlation (IGC).

13 For example, widely used surveys such as LSMS and DHS collect information only on the coresident children.

nonresident children and parents are not included in the survey. Sample truncation occurs as we lack information on both the dependent and independent variables in the regression for these missing household members. Recent evidence indicates that the standard measures of relative mobility, such as IGRC, suffer from substantial *downward* bias in coresident samples (Emran et al. 2018). Our empirical analysis focuses on the father-son linkage in education, and takes advantage of two exceptionally rich data sets: the rural sample from the China Family Panel Studies (2010) for China and the Rural Economic and Demographic Survey (1999) for India. Both of these surveys are unique in that they include all the children of the household head irrespective of their residency status at the time of the survey.[14] This is especially important in a comparative study such as ours, because cross-country comparisons based on coresident samples can lead to wrong conclusions (see Emran et al. 2018 on Bangladesh and India).

A longstanding concern in the literature has been whether the observed correlations are primarily mechanical, driven largely by genetic transmissions from parents to children (see, for example, the discussion by Black and Devereux (2011)). We address this issue in two ways. First, we develop a simple but plausible approach to check whether the estimated intergenerational persistence could be due solely to genetic correlations. The approach combines the recent evidence on intergenerational correlation in cognitive ability from the behavioral genetics and economics literature with the Altonji et al. (2005) biprobit sensitivity analysis (henceforth called augmented AET analysis). Second, the theoretical foundation for the empirical specification allows us to use economic mechanisms as a test for the importance of parental economic choices.

The substantive conclusions of this paper can be summarized as follows. Intergenerational educational mobility was substantially lower for sons in rural India compared to sons in rural China for the cohorts that went to school in the 1970s-1990s. The point estimates of intergenerational persistence are larger in nonfarm households in both countries. The difference between farm and nonfarm households is statistically significant in rural India, but not in rural China, and this conclusion applies to both the slope and the intercept of the intergenerational educational persistence regression. The evidence suggests that while parent's education and nonfarm occupation were complementary in determining a son's education in rural India, they were separable in rural China.[15] The long-term variance in schooling was significantly higher for the sons born into nonfarm households in rural India, and structural change from agriculture to the nonfarm sector contributed to educational inequality. The evidence from the augmented AET sensitivity analysis shows that the observed persistence in rural India cannot be accounted for by genetic correlation alone, while the persistence in China can be explained away by an ability correlation of plausible magnitude.[16] An important advantage of this approach is that the conclusions refer to the whole population of interest, rather than a subset as usually is the case with other standard approaches such as instrumental variables

[14] Although some surveys collect limited information on the non-resident parents of the household head and spouse, we are aware of only a few surveys that include all children of the household head.

[15] It is important to recognize that separability in rural China does not imply that nonfarm income did not play any role in rural income inequality; in fact, we cite a substantial literature to the contrary. The evidence in this paper suggests that intergenerational educational linkage is unlikely to be an important mechanism through which the nonfarm sector influenced the observed inequality in rural China in the 1980s and 1990s.

[16] This evidence on rural China is similar to the recent analysis on urban China based on a sample of monozygotic twins by Behrman et al. (2020). They find that the estimated impact of father's education can be explained fully by genetic correlations and unobserved family endowments.

strategies.[17] Under the null hypothesis that the observed persistence is due solely to genetics, the economic mechanisms identified in the theory are unlikely to provide a coherent explanation of the pattern of mobility across countries.

We use household income data from Chinese Household Income Project (CHIP 2002, CHIP 1995) and National Sample Survey (NSS 1993) of India to explore the mechanisms behind the pattern of relative mobility (the IGRCs) across farm and nonfarm households. The evidence suggests an important role for the economic forces (returns to education and occupational persistence) in explaining the pattern of educational persistence across farm and nonfarm households, both in rural India and rural China. The theory also suggests that we should observe changes in the intergenerational educational persistence in rural China for the younger generation because of the effects of reform on the relevant mechanisms such as higher returns to education in nonfarm occupations. In contrast to the separability observed for the older cohorts, we indeed find evidence of emerging complementarity between father's education and nonfarm occupation in rural China for the educational attainment of the younger generation (18-28 years old in 2010).

The rest of the paper is organized as follows. Section 2 develops a model of intergenerational persistence with credit constraint that incorporates the salient features of farm vs. nonfarm households relevant for educational attainment. The next section describes the estimating equations derived from the theory and the empirical issues in understanding potential complementarity between parent's education and nonfarm occupation in determining children's schooling. Section 4 discusses the data and Section 5 reports the main empirical estimates. The following section explores the evidence on the mechanisms identified by the theory underlying the observed pattern of slope (IGRC) and intercept estimates for the cohorts who went to school in the 1990s or earlier. Section 7 provides evidence on possible changes in the pattern of educational mobility in rural China for the younger generation (18-28 years old in 2010). The paper concludes with a summary of the main results from the theoretical and empirical analysis.

2 A theory of intergenerational educational persistence in a dualistic rural economy

We develop an extension of the Becker-Tomes model with credit constraint (Becker and Tomes 1986) to understand the role played by nonfarm occupations of parents in intergenerational educational mobility of children in a rural economy. The differences in expected returns to education in farm vs. nonfarm households play an important role in this set-up. Expected returns to education in our context depends on two factors: the probability of getting a non-farm employment, and the difference between returns to education in farm vs. nonfarm occupations. The goal in this section is to derive an estimating equation that incorporates these differences in farm and nonfarm households.

The basic set-up

The economy consists of households with a father and a son. We couch the discussion in terms of father and son given that our empirical analysis focuses on the father-son linkages in schooling attainment. The father of child i is described by a pair $\left(S_i^P, O_i^P\right)$ where S_i^P is the education (years of schooling) and $O_i^P \in \{f, n\}$ with f denoting farming occupation and n

[17] It is well-understood that the instrumental variables strategy provides estimates of local average treatment effect (LATE), and refers only to the subset of the population whose treatment status is affected by the instrument.

denoting nonfarm occupation of the father. Given his education and occupation, the father's income is determined as follows:[18]

$$Y_i^p = Y_0^{pj} + R^{pj} S_i^p \ ; \ j = f, n \tag{1}$$

The income determination equation assumes that the fathers with zero years of schooling working in occupation j earns $Y_0^{pj} > 0$, and the returns to education in occupation j is R^{pj} for the parental generation. The assumption that $Y_0^{pj} > 0$ is motivated by our empirical context where a substantial proportion of fathers has zero years of schooling, but positive household income. It is important to underscore that the focus is on how a household's ability to invest in education changes with the education of the father. The "returns to education" relevant here thus relate to permanent household income, not an individual's labor market earnings in a given year which has been the focus of much of the literature on the Mincerian returns to education.[19] In general, the intercepts are likely to be different, but whether Y_0^{pn} is larger or smaller than Y_0^{pf} will depend on the quality of the nonfarm activities at low education level $\left(S_i^p = 0\right)$ which is likely to vary across countries. If low-end nonfarm activities have low-productivity, then it is possible that $Y_0^{pn} < Y_0^{pf}$.

The father allocates Y_i^p to own consumption C_i^p and investment in child's education I_i; thus the budget constraint is

$$Y_i^p \geqq C_i^p + I_i \tag{2}$$

The educational investment is made from a father's own income, as there is little or no financing available from the credit market for such investments in developing countries. The education production function assumes that years of schooling depends on father's investment (I_i) and a child's cognitive ability (ϕ_i):

$$S_i^c = F(I_i) = \theta_0 + \theta_1 \phi_i + \theta_2 I_i \tag{3}$$

The a priori sign restrictions are: $\theta_0 \geq 0$, $\theta_1, \theta_2 > 0$. We would expect θ_0 to be higher when government policies such as free primary schooling (including free books and midday meals etc) are in place so that a child can get a certain level of education, for example, primary schooling without any significant investment by the parents. A higher ϕ_i implies that child i has higher innate cognitive ability, and a higher ability produces more schooling, *ceteris paribus*. We normalize the ability measure such that $E(\phi_i) = 0$.

The productivity of parental financial investment is represented by the parameter θ_2 which captures, among other things, the quality of schools available (and affordable) to a family. For example, there is evidence that children have better learning outcomes in rural India when they attend private schools (Kingdon (2017)). The parents need to pay fees for private school while the public schools do not charge any fees, but the quality of schooling is, in general, low in the public schools. The productivity of financial investment is likely to be higher when the private education market is well-developed and parents can buy better quality by paying higher tuition and/or donations for admission. In contrast, when schooling is primarily provided by the government free of charge (including free books and midday meals) and the private market is thin or nonexistent, the role played by parent's investment in children's

[18] This specification is similar to that in Solon (2004).

[19] This point may be especially important in rural China during 1980s and early 1990s when the labor market was still not functioning very well, and the labor market earnings would be a poor measure of a household's economic status.

education is expected to be rather limited, making θ_2 small.[20] Note that the access to better quality schools does not depend on a parent's occupation in the formulation in Eq. 3 (i.e., θ is not indexed by j).

Parent's optimization

The consumption sub-utility function of the parent is given by:

$$U\left(C^p\right) = \alpha_1 C^p - \alpha_2 \left(C^p\right)^2 \tag{4}$$

Denote the expected income of a child i with education S_i^c at the time of the parental investment choice by $E\left(Y_i^c \mid S_i^c\right)$. The parent's optimization problem is (denoting the Lagrange multiplier on the budget constraint by λ):

$$Max_{C_p, I} V^p = U\left(C^p\right) + \sigma E\left(Y_i^c \mid S_i^c\right) + \lambda\left[Y_i^p - C_i^p - I_i\right] \tag{5}$$

subject to Eqs. 1 and 3. In this formulation, the parameter σ is the degree of parental altruism. The expected income of the child $E\left(Y_i^c \mid S_i^c\right)$ depends on the probability of getting a nonfarm job, and is given as follows:

$$E\left(Y_i^c \mid S_i^c\right) = \left[\pi_i^{nj} R^{cn} + \left(1 - \pi_i^{nj}\right) R^{cf}\right] S_i^c \tag{6}$$

where $\pi_i^{nj} \geq 0$ is the probability that child i gets a nonfarm job when the father is employed in occupation $j = n, f$. To simplify notation, we write $\pi_i^n\left(O_i^p = j\right) \equiv \pi_i^{nj};\ \ j = n, f$. If there is intergenerational persistence in non-farm occupations then the probability that a child gets a nonfarm job is higher when the parent is also in the nonfarm occupations, i.e. $\pi_i^{nn} > \pi_i^{nf}$. This may reflect learning by doing at the parent's workplace through informal apprenticeship, referral and network effects in the labor market, and role model effects (for a discussion, see Emran and Shilpi (2011)). Government policies also affect the strength of occupational persistence. As noted earlier, a major policy difference between rural India and rural China during our study period is the restrictions on rural-urban migration in China. These restrictions implied that many more children had to stay back in the rural areas in China compared to the counterfactual of no such restrictions (the India case). A substantial literature on occupational mobility in rural China shows that the Hukou restrictions reduced the persistence in non-farm occupations, as many children of the nonfarm parents had to take up farming activities because they could not migrate to the urban labor market. This implies that we should expect $\pi_i^{nn}(India) > \pi_i^{nn}(China)$.

The first order conditions for parent's optimization are:

$$\begin{gathered} \alpha_1 - 2\alpha_2 C^p - \lambda = 0 \\ \sigma\theta_2\left\{\pi_i^{nj} R^{cn} + \left(1 - \pi_i^{nj}\right) R^{cf}\right\} - \lambda = 0 \end{gathered} \tag{7}$$

[20] However, the "free schooling" offered by the governments may not be free, especially for the poor households, because of corruption when the enforcement system is not unbiased and impersonal. Emran et al. (2020a) show that, in Bangladesh, the poor parents are more likely to pay bribes for admission into "free" public schools.

The first order conditions and the budget constraint together yield the following solution for the optimal investment in a son's education:[21]

$$I_i^* = \chi_0^j + R^{pj} S_i^p \tag{8}$$

where

$$\chi_0^j = Y_0^{pj} + \frac{1}{2\alpha_2}\left[\sigma\theta_2\left\{\pi_i^{nj} R^{cn} + \left(1-\pi_i^{nj}\right) R^{cf}\right\} - \alpha_1\right] \tag{9}$$

Intergenerational persistence equation

Combining Eqs. 3 and 8 above, we get the following relationship between the education of the father and that of a son, determined by the optimal investment decision:

$$S_i^{cj*} = \psi_0^j + \psi_1^j S_i^p + \tilde{\varepsilon}_i \tag{10}$$

where

$$\begin{aligned} \psi_{0i}^j &= \theta_0 + \theta_2 \chi_0^j \\ \psi_1^j &= \theta_2 R^{pj} \quad ; \tilde{\varepsilon}_i = \theta_1 \phi_i \end{aligned} \tag{11}$$

Equation 10 is consistent with the almost universally used specification for intergenerational schooling persistence in the literature, but it allows for possible differences in educational opportunities across the farm and nonfarm households. Some examples of studies that use this linear-in-levels specification are: Neidhofer et al. (2018); Narayan et al. (2018) and Hertz et al. (2007) on cross-country analysis, Azam and Bhatt (2015) and Emran and Shilpi (2015) on India, and Emran and Sun (2015a) on China. However, we are not aware of any studies on intergenerational educational mobility in developing countries that derive the estimating equation from a theoretical model.

It is standard to assume homothetic functional forms in the analysis of intergenerational income mobility, as the estimating equation is log-linear (see, for example, Solon (1999, 2004)). The reliance on the linear in levels (years of schooling) specification in the literature on intergenerational educational persistence partly reflects the fact that a substantial proportion of fathers have no schooling. This is especially relevant in the rural areas of developing countries such as India where about 40 percent of fathers have no schooling (estimate based on REDS 1999 data). The estimating Eq. 10 is also consistent with the common assumption in the literature that the omitted ability is captured in the error term of the intergenerational persistence regression, and can lead to ability bias in the OLS estimates of the parameters.[22]

An important implication of Eq. 10 is that both the slope and the intercept of the persistence equation capture the differences in educational opportunities faced by the children in farm vs. nonfarm households, although the existing literature focuses largely on the slope (IGRC).[23]

[21] Note that the missing credit market is important for an interior solution given that the education production function has constant returns to parental financial investment. If there is a credit market where parents can borrow at a fixed interest rate to finance investments in children's schooling, the solutions to the optimization problem takes the bang-bang form (either zero or infinite). For a model where such a credit market exists, and there is diminishing returns to investment in education production function, see Becker et al. (2018).

[22] Although most of the existing studies on intergenerational mobility adopt this additively separable specification for the impact of ability, there is little evidence on the validity of this assumption. In a recent paper, Ahsan et al. (2020) use measures of ability based on Raven's test and two memory tests in Indonesia and find evidence in favor of this assumption.

[23] For example, none of the 13 studies on educational mobility in developing countries summarized in Emran et al. (2018) report estimates of intercepts. Some of the more recent works report measures of absolute mobility that combines both the slope and the intercept effects (following Chetty et al. (2014)), but do not report the intercept estimates separately.

An interpretation of the intercept term in our context is that it provides an estimate of the expected education of the children from the subset of households where the fathers have zero schooling. Thus, the intercept estimate may be especially important in developing countries where a significant proportion of the households have parents with zero schooling in the data. In the empirical analysis, we thus pay close attention to the estimated intercepts across farm and nonfarm households in addition to the standard relative mobility measures based on slopes.

Equally important, Eqs. 10 and 11 help improve our understanding of the economic mechanisms behind the observed pattern of mobility. For example, consider the factors that determine the intercepts across farm and nonfarm households. The intercept is, *ceteris paribus*, higher (lower) for the nonfarm children when the income of the parents with zero schooling is higher (lower) in the nonfarm occupations. These nonfarm occupations are, however, likely to be unskilled as they do not require any schooling. In some countries, the low-skilled nonfarm occupations may yield very low income, lower than the income of the farmers (Lanjouw and Lanjouw 2001, World Bank (2011)), making the intercept for the nonfarm households smaller. Another important implication, noted before, is that intergenerational persistence in occupational choices is likely to affect the relative magnitudes of the intercept terms. When $\hat{\pi}_i^{nn} > \hat{\pi}_i^{nf}$, it is more likely to have $\hat{\psi}_0^n > \hat{\psi}_0^f$, *ceteris paribus*, assuming that $R^{cn} > R^{cf}$. Conversely, if $\hat{\pi}_i^{nn} > \hat{\pi}_i^{nf}$ but $R^{cn} < R^{cf}$, then it is more likely to have $\hat{\psi}_0^n < \hat{\psi}_0^f$, *ceteris paribus*. This implies that the evidence on intergenerational occupational persistence (farm vs. nonfarm) accumulated independently in a sub-strand of the literature is necessary to understand the pattern of intergenerational educational mobility in a rural economy. As emphasized by Emran and Shilpi (2019), the interactions between occupational and educational mobility are not considered in the existing literature on developing countries; there are two sub-strands of the literature that grew independently: one focusing solely on education and the other focusing solely on occupation.

The relative magnitudes of the slope parameters (IGRC) across farm and nonfarm households depend on the household returns to education in the parental generation $\left(R^{pj}\right)$ according to Eq. 11. Thus, $R^{pn} > R^{pf}$ generates complementarity between the parent's education and occupation in determining children's schooling.[24] This provides testable implications to check the importance of economic forces in the observed differences in relative mobility across farm and nonfarm households in China vs. India.[25]

3 Empirical approach

Equation 10 above suggests the following estimating equation for the combined farm and nonfarm sample which we take as a benchmark:

$$S_i^c = \psi_0 + \psi_1 S_i^p + \varepsilon_i \tag{12}$$

[24] The theoretical analysis shows that different roles are played by the parental returns to education and the *expected returns* to education in children's generation, a point not adequately recognized in the current literature. It is important to appreciate that the children's *expected returns* to education at the time of the investment decision do not affect the slope (IGRC), their effects are mediated only through the intercept of the persistence regression.

[25] Returns to education are identified as a major factor in changes in intergenerational income persistence (see Becker and Tomes (1979, 1986), Solon (1999, 2004).

where $\varepsilon_i = \tilde{\epsilon}_i + \eta_i = \theta_1\phi_i + \eta_i$, and η_i captures exogenous idiosyncratic shocks to children's schooling. We normalize so that $E(\eta_i) = 0$. The corresponding estimating equation allowing for different intercepts and slopes for the farm and nonfarm households is:

$$S_i^c = \psi_0^f + \psi_1^f S_i^p + \lambda_0 D_i^{np} + \lambda_1 \left(S_i^p * D_i^{np}\right) + \varepsilon_i \tag{13}$$

where D_i^{np} is a dummy variable that takes on the value of 1 when the father of child i is employed in nonfarm occupations, and zero otherwise. In this formulation, the measure of relative mobility is IGRC: ψ_1^f for the farm households and $\psi_1^f + \lambda_1$ for the nonfarm households; we denote $\psi_1^f + \lambda_1 \equiv \psi_1^n$. Similarly, the intercepts are given by ψ_0^f (farm) and $\psi_0^n \equiv \psi_0^f + \lambda_0$ (nonfarm).

Note that we do not include any controls in Eqs. 12 and 13. As emphasized by Emran and Shilpi (2019), including household or individual characteristics may bias the estimate in our case because father's education is a summary statistic for all family background factors that determine children's schooling. Following the seminal contribution of Solon (1992), it is standard in the literature on intergenerational income mobility to include quadratic age controls for both the parents and the children. This helps reduce the biases that arise from life-cycle effects in estimating permanent income.[26] The life-cycle bias is not likely to be a concern in our application, as we chose the age cut-off to ensure that most of the children completed schooling by the time the survey was done. Our main estimates thus do not include any age controls; but, as a robustness check, we estimate Eq. 13 including age controls.

It is important to appreciate that a comparison of farming and nonfarming households based solely on the most widely used measure of mobility, i.e., IGRC (ψ_1^f and ψ_1^n), may be misleading. The caveat that IGRC or other measures of relative mobility such as intergenerational correlation (IGC) may be misleading in comparing mobility across groups has been emphasized by Hertz (2005), Mazumder (2014), and Bhattacharya and Majumder (2011) in their analysis of racial (black-white) differences in intergenerational income mobility in United States of America. But it has not been adequately appreciated in the literature on intergenerational educational mobility, both in economics and sociology.[27] This is especially so in developing countries, as is evident from the fact that most of the available studies on China and India we are aware of focus exclusively on relative mobility measures such as IGRC, and IGC (intergenerational correlation).

To see the pitfalls in relying on IGRC alone in our context, it is instructive to consider the case where $\psi_1^n > \psi_1^f$ so that intergenerational persistence is higher in nonfarm households. However, whether this higher persistence leads to convergence or divergence in schooling attainment of children born into farm and nonfarm households depends on the relative magnitudes of the intercepts. When the intercepts are $\psi_0^n > \psi_0^f$, the expected schooling is higher for children born into nonfarm households across the distribution of parental schooling, and the gap between the two groups widens as parental education increases (please see Fig. 1). On the other hand, we can have two sub-cases when the intercepts are: $\psi_0^n < \psi_0^f$ (please see Fig. 2). If the IGRC for the nonfarm group is high enough, the children born to lower educated nonfarm households are disadvantaged compared to the children of low-educated farmer parents, but at the higher end of parental education distribution they are relatively advantaged (see nonfarm(a) line in Fig. 2). When the difference between IGRC estimates is

[26] When data on many years spanning the appropriate phases of life-cycle are available, the age controls are not necessary. Some recent contributions do not include any age controls, as it may wipe out the inter-cohort differences in income mobility.

[27] See the discussion on this point by Torche (2015) in the context of Sociological literature on mobility.

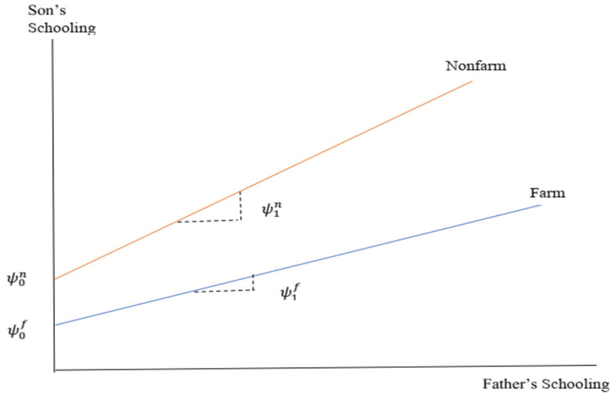

Fig. 1 Intergenerational schooling relationships when both the intercept and the IGRC are greater for nonfarm households

small enough, the farmer's children are better off in educational attainment over the entire distribution, and only in this special case, the conclusion based on IGRC that nonfarm children face lower relative mobility is consistent with the idea that they are at a disadvantage in educational attainment (please see the nonfarm(b) line in Fig. 2).

Rank-based measures of intergenerational mobility

While most of the existing studies on intergenerational educational mobility in developing countries rely on years of schooling as the indicator of educational attainment, following the influential contribution of Chetty et al. (2014), the recent literature is increasingly adopting the rank-based measures where the indicator of educational status is the percentile rank in the relevant distribution. A growing literature suggests that the rank-based measures of mobility are significantly more robust to data limitations compared to the measures based on years of schooling (Nybom and Stuhler 2017; Emran and Shilpi 2018).

Denote r_i^c as the percentile rank of child i in the over-all (including both farm and nonfarm) schooling distribution of children, and r_i^p the percentile rank of the father of i in the over-all schooling distribution in fathers generation.[28] For the rank-based estimates, the estimating equations are as follows:

$$r_i^c = \delta_0 + \delta_1 r_i^p + \xi_i \;\; \forall i \tag{14}$$

$$r_i^c = \delta_0^f + \delta_1^f r_i^p + \lambda_2 D_i^{np} + \lambda_3 \left(r_i^p * D_i^{np} \right) + \xi_i \tag{15}$$

The slope parameters of regression Eqs. 14-15 represent intergenerational rank-rank slope (IRRS, for short) which is a measure of relative mobility similar to IGRC. The IRRS is given by δ_1^f for farm and $\delta_1^n \equiv \delta_1^f + \lambda_3$ for nonfarm. Similar relations hold for the intercepts.

However, there are important differences between IGRC and IRRS as measures of mobility, especially for education. A common argument is that IRRS is not affected by the growth at the top of the distribution, or the changes in cross-sectional inequality across generations. As noted recently by Ahsan et al. (2022), this argument is valid for a continuous variable with large support such as income, but not for education (years of schooling) which is a discrete variable with limited support. They report evidence that when ranks are calculated for years of schooling using the standard mid-rank method, the cross-sectional schooling inequality

[28] Following the statistics and economics literature, we calculate the ranks using the mid-rank method.

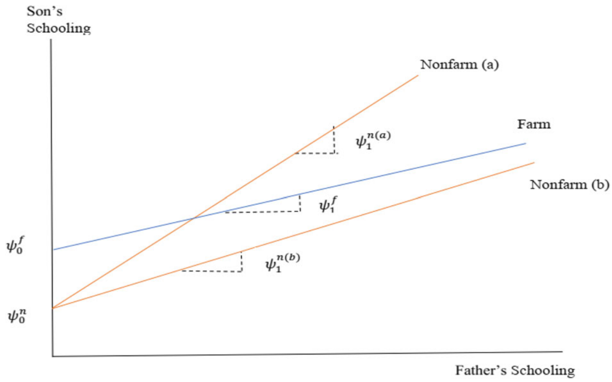

Fig. 2 Intergenerational schooling relationships when the IGRC is greater but the intercept is lower for nonfarm households

across generations can be substantially different.[29] The economic interpretation of these two measures are also very different. IGRC captures the effects of changes in economic policies such as school construction and trade liberalization as they primarily affect the marginal distributions. IRRS, in contrast, captures primarily the effects of deep-seated institutional structure and social norms, for example, racial bias, and caste discrimination. It is important to appreciate these differences when comparing estimates of IGRC and IRRS.[30]

Interaction between parent's education and occupation: complementary, substitutes or separable?

An important advantage of the empirical models discussed above is that they provide a straight-forward way to test the nature of interaction between parent's occupation and education in determining intergenerational persistence in schooling. Consider, for example, the estimating Eq. 13 above; from the theoretical analysis in Section (2), it is easy to derive the conditions under which parental education and non-farm occupation can be complementary, i.e., $\lambda_1 > 0$ implying$\left(\psi_1^n > \psi_1^f\right)$, substitutes, i.e., $\lambda_1 < 0$ implying $\left(\psi_1^n < \psi_1^f\right)$, or separable, i.e., $\lambda_1 = 0$ implying $\left(\psi_1^n = \psi_1^f\right)$. The prevailing view among many observers is that nonfarm occupation and education are likely to be complementary in determining children's education, leading to cumulative forces of inequality in educational attainment and income in villages in developing countries (see, for example, Rama et al. (2015)). Yet, to the best of our knowledge, there is no evidence in the literature on the existence and the nature of the interaction between parent's education and occupation in determining children's educational attainment. Also, without a formal model, the economic mechanisms behind the hypothesized complementarity cannot be assessed. According to the theoretical model above, such complementarity requires that returns to education for the parents is higher in non-farm

29 As a results, the rank-rank slope is different from the rank correlation in this case. For a continuous variable, the rank-rank slope is equal to the rank correlation.

30 Recent evidence suggests that these two measures can lead to conflicting conclusions regarding the causal effects of government policies. See, for example, the analysis of the INPRES primary schools (60,000 primary schools constructed in the early 1970s in Indonesia) on intergenerational educational mobility by Ahsan et al. (2023).

occupations, i.e., $R^{pn} > R^{pf}$.[31] This is one of the predictions that we take to the data as a test of the importance of economic mechanisms underlying the observed pattern of intergenerational educational persistence.

4 Data

For our main empirical analysis, we use two exceptionally rich surveys that collected data on children irrespective of their residency status at the time of the survey. The data for rural India come from the Rural Economic and Demographic Survey (REDS) carried out by the National Council for Applied Economic Research, and the source of the data for rural China is the China Family Panel Studies (CFPS) implemented by the Institute of Social Science Survey unit of Peking University.[32]

This is an important advantage for the empirical analysis, as most of the evidence on intergenerational educational mobility in India and China currently available are based on data that suffer from truncation due to coresidency restrictions used to define household membership. Emran et al. (2018) summarize 13 studies on intergenerational educational mobility in developing countries, only two of which use data not affected by coresidency bias.[33] While Emran et al. (2018) provide evidence of substantial downward bias (average 18 percent) in the IGRC estimate from the coresident sample in rural India, we are not aware of any similar estimate for rural China. In online appendix A, we provide evidence on the extent of coresidency bias in rural China in a widely used household survey: the Chinese Household Income Project (CHIP). In particular, we compare the estimates from the CFPS (without any sample truncation) with those from the CHIP 2002 for the overlapping age cohorts. The evidence shows that the IGRC estimate from the CHIP 2002 is 25 percent smaller because of truncation of the sample arising from coresidency restrictions (see Table A.2 in the online appendix). A comparison of CHIP 2002 with the CFPS is also of independent interest, because CHIP 1995 and 2002 have been used by many researchers to study intergenerational mobility in China.[34] For a more complete discussion, please see online Appendix A.

We use the 1999 round of the REDS and the first round of the CFPS in 2010. From the REDS data, we obtain the relevant information for our analysis on all father-son pairs irrespective of residency status at the time of the survey. For the CFPS data, we restrict to rural communities subsample, given our focus on intergenerational mobility in rural areas, and use the family roster to obtain a complete list of father-son pairs that includes all sons of the household head irrespective of their residency status at the time of the survey.

31 If private school locations are motivated by higher income associated with nonfarm activities, then school quality may also play a role in generating complementarity. In this case, the productivity of parental investment θ_2 will be correlated with occupation, i.e., $\theta_2^n > \theta_2^f$.

32 One might wonder why we chose not to use the IHDS 2012 round survey for India which would provide a survey year close to the survey year of CFPS in China. The CFPS and REDS are the most comparable in that they provide *a random sample of parents* with information on all their children irrespective of the residency status of a child at the time of the survey. The IHDS, in contrast, contains *a random sample of children* with information on their parents irrespective of their residency status at the time of the survey.

33 The exceptions are Fan et al. (2021) and Azam and Bhatt (2015).

34 Two of the authors of this paper, M. Shahe Emran and Yan Sun, used CHIP 2002 data to analyze the effects of farm and nonfarm occupations on intergenerational educational mobility in rural China (see Emran and Sun (2015b)). We decided not to publish that paper because of the worry about the biases due to sample truncation arising from coresidency. This paper replaces (Emran and Sun 2015b) and the conclusions on rural China here supersede those in Emran and Sun (2015b).

The main samples for our analysis consist of children aged 18 - 54 in the 1999 REDS survey, and 29 - 65 in the 2010 CFPS survey. This ensures that we focus on the same age cohorts of children who went to school mostly during the 1980s and 1990s. It is important to recognize that such an analysis for the overlapping age cohorts is meaningful for education, as most of the children under focus (29-65 years old in 2010) in China have completed their schooling by 1999, even though the information was gathered later in 2010. The observations with fathers aged over 100 years or missing, or sons aged over 65 years are excluded from the samples used in the empirical analysis.

In each data set, we observe the education level and an indicator of whether the main occupation is agriculture or nonfarm activity for both the father and the son. Our main analysis of educational mobility is based on years of schooling as the measure of educational attainment. Father's schooling is used as the indicator of parental education to avoid complications from many missing observations on mother's schooling. In our data sets, the maximum of parental education coincides with the education level of the father in most of the cases. We define the parental occupation dummy $D_i^{np} = 0$ when the father of child i reports agriculture as the main occupation (corresponding to $O_i^p = f$ in the theoretical model), $D_i^{np} = 1$ otherwise. This means that the households who are primarily engaged in farming with some nonagricultural sources of income are classified as agricultural occupation.

Online appendix Table A.1 shows the descriptive statistics of our main data samples from the REDS and the CFPS. In the REDS sample, we have 6887 observations, and the children's are 29 years old on average in the survey year 1999. Fathers are 60 years old on average. About half of the children's main occupation is agriculture, while 60% of the fathers also reported agriculture as their main occupation. The children attain significantly higher levels of education than the fathers, when comparing their average years of schooling (6.26 vs. 4.13.)

In CFPS sample, a similar pattern is observed. We have 3,305 father-son pairs, and children's age is about 40 years in the survey year, 2010 (29 years in 1999, same as that for India in 1999 REDS data). Fathers are aged 68 years on average in 2010. About half of the fathers work in the agricultural sector. Children receive 6.31 years of schooling on average, significantly higher than their fathers (less than 3.81 years).

While our main empirical analysis is based on the CFPS 2010 and REDS 1999, we take advantage of a number of additional data sets for exploring the economic mechanisms identified by the theoretical analysis. To understand how the relation between a father's education and household income varies by farm and nonfarm occupation in rural China we utilize the data from the Chinese Household Income Project (CHIP) 1995 and 2002. To estimate the relation between father's education and household income in rural India, we use the data on household total expenditure from the National Sample Survey 1993.

5 Empirical results

5.1 Evidence on relative mobility and test of complementarity

Table 1 reports the estimates of relative mobility using two measures: intergenerational regression coefficient (IGRC) and intergenerational rank-rank slope (IRRS). In addition to the separate estimates for the farm and nonfarm households, we report the estimates from the combined farm and nonfarm sample as a benchmark.

Table 1 Relative Mobility and Test of Complementarity: Rural China and Rural India

	IGRC $\left(\psi_1^j\right)$		IRRS $\left(\psi_1^j\right)$	
	CHINA	INDIA	CHINA	INDIA
Combined Sample	**0.313**	**0.518**	**0.337**	**0.456**
	(0.025)	(0.020)	(0.017)	(0.014)
Farm	**0.311**	**0.488**	**0.331**	**0.422**
	(0.033)	(0.027)	(0.032)	(0.024)
Nonfarm	**0.316**	**0.555**	**0.344**	**0.500**
	(0.029)	(0.027)	(0.031)	(0.024)
Test of Complementarity (Farm/Nonfarm)				
H_0: Farm and Nonfarm Coefficients are Equal				
	IGRC $\left(\psi_1^n = \psi_1^f\right)$		IRRS $\left(\delta_1^n = \delta_1^f\right)$	
	CHINA	INDIA	CHINA	INDIA
F Statistic	**0.014**	**3.803**	**0.134**	**6.421**
P-Value	**0.906**	**0.052**	**0.715**	**0.012**

Notes: (1) IGRC stands for Intergenerational Regression Coefficient, and IRRS stands for Intergenerational Rank-Rank Slope. (2) The numbers in parenthesis are robust standard errors clustered at the Primary Sampling Unit level. (3) The number of observations for China: Combined (3305), Farm (1662), Nonfarm (1643), and for India: Combined (6952), Farm (4035), Nonfarm (2917)

Table 2 Estimates of intercepts and test of equality

	IGRC Intercept $\left(\psi_0^j\right)$		IRRS Intercept $\left(\delta_0^j\right)$	
	CHINA	INDIA	CHINA	INDIA
Combined Sample	**5.862**	**5.339**	**0.381**	**0.340**
	(0.221)	(0.165)	(0.010)	(0.009)
Farm	**5.713**	**5.612**	**0.371**	**0.364**
	(0.282)	(0.219)	(0.025)	(0.017)
Nonfarm	**6.012**	**4.980**	**0.391**	**0.307**
	(0.237)	(0.193)	(0.022)	(0.015)
Test of Equality (Farm/Nonfarm)				
H_0: Farm and Nonfarm Intercepts are Equal				
	IGRC Intercepts $\left(\psi_0^n = \psi_0^f\right)$		IRRS Intercepts $\left(\delta_0^n = \delta_0^f\right)$	
	CHINA	INDIA	CHINA	INDIA
F Statistic	**1.181**	**6.233**	**0.626**	**7.784**
P-Value	**0.279**	**0.013**	**0.430**	**0.006**

Notes: (1) The numbers in bold in the upper panel are estimates of the intercepts from intergenerational persistence regression using years of schooling (Called IGRC intercepts), and from rank-rank regressions (called IRRS intercepts). (2) the numbers in parenthesis are robust standard errors clustered at the Primary Sampling Unit level

Table 3 AET (2005) sensitivity analysis for ability bias

The effects of father's higher education on the probability of higher education of sons

	CHINA		INDIA	
	Farm	Nonfarm	Farm	Nonfarm
ρ = 0	**1.72**	**2.83**	**8.68**	**9.99**
	(0.38)	(0.46)	(0.44)	(0.45)
ρ = 0.10	**1.17**	**2.14**	**7.64**	**9.46**
	(0.44)	(0.51)	(0.47)	(0.48)
ρ = 0.20	**0.38**	**1.2**	**6.27**	**8.65**
	(0.50)	(0.57)	(0.51)	(0.53)
ρ = 0.30	**−0.66**	**−0.02**	**4.53**	**7.48**
	(0.57)	(0.63)	(0.55)	(0.58)
ρ = 0.40	**−1.99**	**−1.52**	**2.39**	**5.88**
	(0.64)	(0.68)	(0.58)	(0.63)

Notes: (1) AET (2005) stands for Altonji, Elder and Taber (2005, Journal of Political Economy) Biprobit sensitivity analysis. (2) **ρ** stands for correlation in cognitive ability of father and son. Estimates in the first row are the univariate Probit estimates. The upper bound **ρ = 0.40** is based on economics and behavioral genetics literature. (3) The numbers in bold are percentage points increase in the probability of higher education of sons when the father has higher education. (3) Higher education for parents implies more than primary, and for sons in India more than 10 years of schooling, for sons in China more than 9 years of schooling. (4) The numbers in parenthesis are standard errors clustered at the PSU level

The point estimates of IGRC show that, both in rural India and rural China, intergenerational persistence in schooling is higher for the sons born into nonfarm households, but the estimates for farm and nonfarm households are similar in magnitude in China. A son of a father with 1 year more schooling in India is expected to gain 0.49 year of schooling on average if the father is a farmer, while the expected gain increases to 0.56 year of schooling when the father is employed in a nonfarm occupation (column 2 of Table 1). The corresponding estimates for rural China are 0.31 year (farm) and 0.32 year (nonfarm) of additional schooling for the sons born to a father with 1 year of more schooling (column 1 of Table 1). Another important conclusion from the evidence in Table 1 is that all of the IGRC estimates in rural China are smaller compared to the corresponding estimates in rural India, providing strong evidence that the sons in rural China who went to school in the 1980s and 1990s enjoyed substantially more relative mobility in schooling. The conclusions above remain valid when we include age controls in the specifications (see Table A.3 in the online appendix).[35]

The estimates of intergenerational rank-rank slope (IRRS) reported in columns 3 and 4 of Table 1 also tell a similar story: the point estimates of the influence of a father's schooling rank on the son's schooling rank are higher for the nonfarm households, both in China and India. Again, the influence of parental education does not vary substantially between farm and nonfarm household in rural China, but there is substantial difference in rural India. The magnitudes of the IRRSs are consistently smaller in rural China compared to those in India, reinforcing the conclusion from the IGRC estimates that the sons in rural India faced lower educational mobility. These conclusions from the IRRS estimates remain intact when we include age controls in the specification (see Table A.4 in the online appendix).

[35] Since life-cycle bias is not likely to be a major issue in our context, our preferred estimates are from the specification without age controls.

Table 4 Father's education and household income

	Panel A: Estimates for Rural China			
	Intercept		Slope (Returns to Education)	
	Farm	Nonfarm	Farm	Nonfarm
CHIP 2002	**8344.64**	**8582.80**	**275.32**	**335.42**
	(670.34)	(832.97)	(91.38)	(108.72)
CHIP 1995	**4608.86**	**4457.28**	**34.21**	**107.45**
	(666.28)	(383.14)	(71.14)	(43.19)
	Test of Equality Between Farm and Nonfarm			
	H_0: Intercepts are Equal		H_0: Slopes are Equal	
CHIP 2002				
F Statistic	**0.08**		**0.30**	
P-value	**0.78**		**0.58**	
CHIP 1995				
F Statistic	**0.06**		**0.96**	
P-value	**0.81**		**0.33**	
	Panel B: Estimates for Rural India			
	Intercept		Slope (Returns to Education)	
	Farm	Nonfarm	Farm	Nonfarm
NSS 1993	**1199.27**	**1172.17**	**68.17**	**77.92**
	(9.67)	(13.57)	(2.52)	(3.28)
	Test of Equality Between Farm and Nonfarm			
	H_0: Intercepts are Equal		H_0: Slopes are Equal	
F Statistic	**2.99**		**5.71**	
P-value	**0.08**		**0.02**	

Notes: (1) The dependent variable for Rural China is the average household income (total). CHIP 2002 is the average of the last 5 years of total household income, and CHIP 1995 is the average of the last 3 years of household income. The dependent variable for India is total household expenditure. (2) The numbers in parenthesis are standard errors. (3) H_0 stands for Null Hypothesis. (4) The number of observations for CHIP 1995: Farm (1709), Nonfarm (3893), and for CHIP 2002: Farm (4457), Nonfarm (4087). For NSS (1993), the number of observations are Nonfarm (20535), and Farm (48196)

The contrasting evidence in China vs. India suggests that father's education and nonfarm occupation are likely to be complementary in India, but separable in China. We formally test the null hypothesis of separability $H_0 : \psi_1^f = \psi_1^n$. The results are reported in the lower panel of Table 1, with standard errors clustered at the primary sampling unit (village in REDS data, and county in CFPS data). The evidence from both IGRC and IRRS estimates shows that, in rural China, the null hypothesis of separability cannot be rejected at the 10 percent significance level; the F statistic for IGRC estimates is 0.014 with a P-value of 0.90, and the corresponding numbers for IRRS are 0.13 (F statistic) and 0.72 (P-value). In contrast, in rural India, the null hypothesis of separability is rejected at the 10 percent level for IGRC (F=3.80, P-value=0.052), and at the 5 percent level for IRRS (F=6.42, P-value=0.012). Since the estimated influence of parental schooling is larger in the nonfarm households in rural

Table 5 Intergenerational persistence in education (18-28 Age Cohort, CFPS)

	Panel A: Estimates Based on Years of Schooling			
	Intercept (ψ_0^j)		IGRC $\left(\psi_1^j\right)$	
	Farm	Nonfarm	Farm	Nonfarm
	6.997	**7.173**	**0.274**	**0.337**
	(0.341)	(0.308)	(0.035)	(0.032)
	Test of Equality Between Farm and Nonfarm			
	H_0: Intercepts are Equal ($\psi_0^n = \psi_0^f$)		H_0: IGRCs are Equal ($\psi_1^n = \psi_1^f$)	
F Statistic	**0.244**		**2.632**	
P-Value	**0.622**		**0.107**	
	Panel B: Estimates Based on Schooling Ranks			
	Intercept (δ_0^j)		IRRS (δ_1^j)	
	Farm	Nonfarm	Farm	Nonfarm
	0.399	**0.404**	**0.304**	**0.373**
	(0.027)	(0.027)	(0.034)	(0.035)
	Test of Equality Between Farm and Nonfarm			
	H_0: Intercepts are Equal ($\delta_0^n = \delta_0^f$)		H_0: Slopes are Equal ($\delta_1^n = \delta_1^f$)	
F Statistic	**0.028**		**2.792**	
P-Value	**0.868**		**0.097**	

India, the evidence suggests complementarity between nonfarm occupation and a father's education in determining a son's schooling.

Relative mobility and long-term variance in schooling

When interpreted as a dynastic model of the evolution of schooling across generations, a higher IGRC implies a higher long-term variance in schooling.[36] To see this, note that for the IGRC Eq. 12, we can write the long-term variance of education as:

$$\sigma_s^2 = \frac{1}{\left(1 - \psi_1^2\right)}\sigma_\varepsilon^2 \tag{16}$$

where σ_s^2 is the long-term variance of education and σ_ε^2 is the long-term variance of the error term capturing all other factors unrelated to father's schooling such as market luck, and macro and trade shocks. $\frac{1}{\left(1 - \psi_1^2\right)}$ is called the 'family background multiplier' by Emran and Shilpi (2019), which amplifies the impact of the shocks to education. Using Eq. 16 and the estimates of ψ_1^f and ψ_1^n reported in Table 1, we have the following estimates for sons in farm and nonfarm households in rural India:

$$\sigma_{s,If}^2 = 1.31\sigma_\varepsilon^2 \quad (farm)$$
$$\sigma_{s,In}^2 = 1.45\sigma_\varepsilon^2 \ (nonfarm)$$

[36] For a discussion on the dynastic interpretation of the model and the implications for long-term variance, see Acemoglu and Autor (undated).

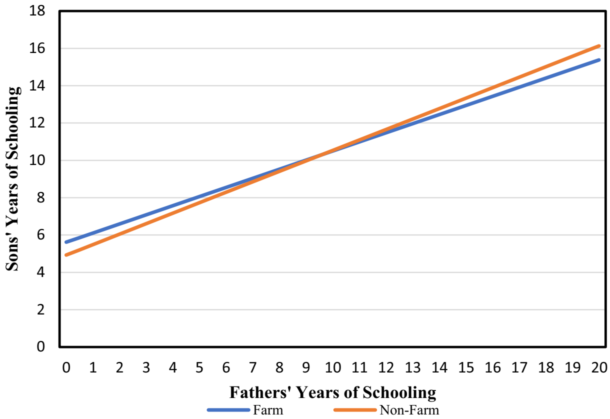

Fig. 3 Regression of fathers' years of schooling against sons' years of schooling in rural India

The subscripts I and s denote India and schooling, respectively, and as before, n=nonfarm and f=farm. The long-term variance of education of sons in the farming sample is 31 percent higher than the variance due to idiosyncratic factors alone (i.e., σ_{ε}^2), and is 45 percent higher in the nonfarm sample. Thus the contribution of family factors to the long-term variance is 14 percentage points higher in the nonfarm households.

The long-term variances in schooling for the farm and nonfarm households in China are:

$$\begin{aligned} \sigma_{s,cf}^2 &= 1.107\sigma_{\varepsilon}^2 \quad (farm) \\ \sigma_{s,cn}^2 &= 1.111\sigma_{\varepsilon}^2 \ (nonfarm) \end{aligned}$$

The multiplier effect of family background is much smaller in the case of China; the long-term variance in schooling is only about 10 percent higher than the variance of idiosyncratic shocks, and the estimates are virtually identical across the farm and nonfarm samples.

5.2 Intercepts and steady states

As noted earlier, measures of relative mobility give us an incomplete, and sometimes misleading, picture of intergenerational mobility across groups such as farm and nonfarm households. A simple but important reason is that different groups may be converging to different steady states due to different intercepts in the intergenerational persistence equations. Perhaps more importantly, the theory in Section (2) suggests that factors such as persistence in occupation choices, and expected returns to investment in schooling for children work through the intercept, leaving relative mobility as measured by IGRC and IRRS largely unaffected.

The estimated intercepts of Eqs. 12-13 and 14-15 above are reported in Table 2. The point estimates show that the intercept of the IGRC equation in India is significantly *higher* for the *farm* households (p-value 0.013). The evidence for the intercept of the IRRS equation is similar (p-value 0.006). When considered along with the evidence that the slope estimates (IGRC and IRRS) are smaller for the farm households in India, the evidence implies a set of interesting conclusions. First, whether the sons born to fathers in farm or nonfarm occupation enjoy educational advantage depends on the level of their fathers' education with a switching threshold of 9-10 years of schooling. Please see Fig. 3. An interpretation of the evidence is that the national public examination administered at 10th grade (known as Matriculation

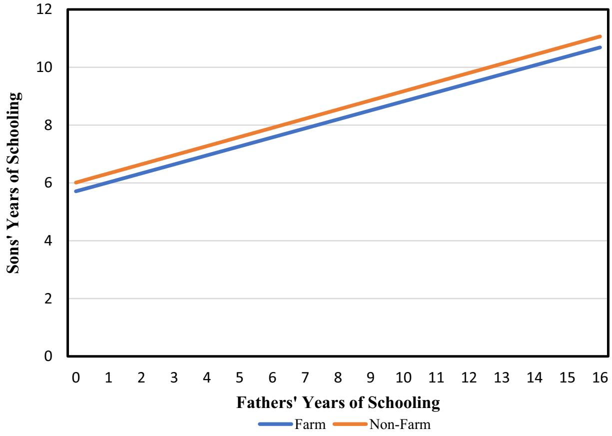

Fig. 4 Regression of fathers' years of schooling against sons' years of schooling in rural China

examination, or all India Secondary School Examination (SSC)) represents a bifurcation point. The children of non-farm fathers with Matriculation or more schooling are expected to achieve better schooling attainment when compared to the children of farmer fathers with similar educational credential, but the children of nonfarm fathers with lower education (and probably employed in unskilled nonfarm jobs) are likely to be worse-off when compared to the children of low educated farmer fathers (who likely own land). Second, the steady state level of education is not substantially different across farm and nonfarm households: 10.81 years of schooling (farm) and 11.20 years of schooling (nonfarm). This reflects the fact that the sons born into nonfarm households gain more from the higher schooling of a father, although they start from a lower intercept.

The picture for rural China is different (please see Fig. 4). The evidence in Table 2 shows that there is no statistically significant difference across the farm and nonfarm households in the intercepts of the intergenerational persistence regressions (p-values are 0.279 (IGRC intercept) and 0.43 (IRRS intercept)). When combined with the evidence on IGRC and IRRS in Table 1, this implies that the schooling attainment of the sons in rural China converges to virtually the same steady state (8.53 years of schooling) irrespective of whether the father is a farmer or is engaged in a nonfarm occupation.[37]

5.3 Structural change and cross-sectional schooling inequality

To understand the implications of the higher variance in the nonfarm households in India for the cross-sectional variance in rural schooling, it is important to consider the structure of the rural economy (i.e., proportion of fathers employed in the nonfarm sector) and both the within group and between groups variances. Denote the proportion of nonfarm households by ω, then we can write the long-term variance as:

$$Var\,(S) = \omega\sigma_{s,n}^{2} + (1-\omega)\,\sigma_{s,f}^{2} + \omega\,(1-\omega)\left(\mu_f - \mu_n\right)^2 \tag{17}$$

[37] The estimate of the steady state is based on the combined farm and nonfarm sample. Although they are not statistically different, the point estimates differ numerically across farm and nonfarm subsamples.

where μ_n and μ_f are the long-term means (the steady state) of a father's education in farm and nonfarm households, respectively. The effects of a marginal increase in the proportion of nonfarm sector is given by:

$$\frac{dVar(S)}{d\omega} = \left(\sigma_{s,n}^2 - \sigma_{s,f}^2\right) + (1 - 2\omega)\left(\mu_f - \mu_n\right)^2 \tag{18}$$

As discussed in Sections (5.1) and (5.2) above, there are no significant differences in the long-term means or long-term variances across the farm and nonfarm households in rural China. This implies that both terms in Eq. 18 are zero, indicating that structural change in favor of the non-farm sector during the decades of 1970s-1990s is unlikely to contribute to the cross-sectional variance of schooling. As noted earlier, this, however, does not imply that the nonfarm sector did not play any role in the increasing income inequality in rural China during this period, only that the nonfarm sector's effect is not mediated through intergenerational educational persistence.

In India, the evidence in Section (5.1) shows that the long-term variance is substantially higher in the nonfarm households because of a large family background multiplier. The estimates of the long-term (steady state) means also show a higher mean for the nonfarm households. When we plug in the estimates from Sections (5.1) and (5.2) for rural India in Eq. 18 above, we get (using $\omega = 0.40$ from the summary statistics table in online appendix):

$$\frac{dVar(S)}{d\omega}\,|_{India} = 0.078 + 0.13\sigma_\varepsilon^2 > 0$$

Thus, the evidence suggests that structural change in favor of the nonfarm sector contributed to higher cross-sectional variance in rural India during our study period.

6 Economic mechanisms: towards an explanation of the differences between rural China and rural India

A major concern in the literature has been whether the observed pattern of intergenerational linkages is primarily driven by omitted variables bias due to unobserved genetic correlations between parents and children. An obvious approach to this question is to try to correct the estimates for possible positive bias due to genetic correlations in cognitive ability. We develop a simple but plausible approach by taking advantage of the recent evidence on intergenerational correlation in cognitive ability from economics and behavioral genetics. There is substantial evidence that intergenerational correlation in cognitive ability (denoted as ρ) falls in a narrow interval, $\rho \in [0.20, 0.40]$; see, for example, Black et al. (2009); Bjorklund et al. (2010) on economic literature, and Plomin and Spinath (2004) on behavioral genetics literature. We use this information in a biprobit sensitivity analysis as developed by Altonji et al. (2005) to check if the estimates of intergenerational persistence in schooling remain positive and statistically significant for plausible values of intergenerational correlation in ability. We call this approach augmented AET (AAET) sensitivity analysis, and it requires binary indicators of educational attainment instead of years of schooling. We use a dummy for higher than primary schooling for fathers. For sons in rural China a dummy for higher than 9 years of schooling (higher middle school), and for sons in rural India, a dummy for more than 10 years of schooling (called SSC or matriculation) are used. The details of this approach are provided in online Appendix B, and the estimates are reported in Table 3.

The evidence from the AAET sensitivity analysis suggests that, the estimated intergenerational schooling persistence in India is very strong, and the estimates remain statistically significant and numerically substantial even when we impose $\rho = 0.40$ in the biprobit model. In contrast, the estimates turn negative in the case of rural China when $\rho = 0.30$, suggesting that the observed positive effect of a father's education could be explained away by ability correlation between parents and children.[38] This evidence strengthens substantially the conclusions that educational mobility was much lower in India in the 1970s-1990s, and that economic forces are likely to be important in explaining the differences between India and China. The advantage of this approach is that it is easily implementable, and thus could be used fruitfully by other researchers. However, it is also important to appreciate the limitations of such an a-theoretical approach. For example, the evidence that the persistence in rural China could be explained by genetic correlations alone does not necessarily imply that economic forces were not at play. The theoretical analysis in Section (2) provides us a way to explore the question by focusing on the economic mechanisms behind the pattern of the slope and intercept estimates across farm and nonfarm households. We turn to this exercise next.

Under the null hypothesis that genetic transmission is the main force at work, we should not expect the economic mechanisms identified in the model to offer a consistent explanation of the observed pattern of intergenerational persistence across India and China. If economic forces are important, the theory provides us with testable implications even in the case of rural China; the equality of the slopes (IGRCs) across the farm and nonfarm households in this case implies equality of the returns to education for the farm and nonfarm parents.

6.1 Differences in the IGRCs

The estimates of IGRC in Table 1 imply the following (denoting an estimate by a hat):

$$\begin{array}{ll} (China) & \hat{\psi}_1^f = \hat{\psi}_1^n \Rightarrow \theta_2 \hat{R}^{pf} = \theta_2 \hat{R}^{pn} \\ (India) & \hat{\psi}_1^f < \hat{\psi}_1^n \Rightarrow \theta_2 \hat{R}^{pf} < \theta_2 \hat{R}^{pn} \end{array}$$

The theoretical analysis thus highlights the importance of household-level returns to schooling in the father's generation (R^{pj}) across occupations for understanding the pattern of relative mobility. We have $\hat{\psi}_1^f - \hat{\psi}_1^n = \theta_2 \left(\hat{R}^{pf} - \hat{R}^{pn}\right)$, and the effects of a widening gap in returns to education for household income between farm and nonfarm households would be low if θ_2, the productivity of financial investment in children's education, is low. We would expect θ_2 to be low when the private market for education is not well-developed in a country.[39] Since the expansion of private schooling has been much larger in India compared to that in China during the study period, the value of θ_2 is likely to be higher in India.

It is important to recognize that the "returns to education" for the parents (i.e., R^{pj}) differ from most of the available estimates of returns to education for three reasons. First, we are interested in the total income of all household members rather than the individual income (i.e., not only a father's income). Second, the focus of the existing literature has been on labor market returns, while our analysis requires both labor and non-labor income. Third, a father's education in our analysis is not only a measure of human capital, but a summary statistic for a family's socio-economic status and captures the effects of other correlated factors, for

[38] These conclusions remain robust when we define the schooling cut-off to be the same in the two countries, and also when age controls are included in the specification. The details are available from the authors.

[39] To see this clearly, consider the polar case where schooling is provided only by the government free of charge and there is no private schools (or private tutoring). In this case, the scope for parental financial investment to improve a child's educational attainment is effectively nonexistent, making $\theta_2 \approx 0$.

example, a mother's education due to assortative matching in the marriage market.[40] Thus, we need a measure of permanent household income.

The Chinese Household Income Project (CHIP) provides us with high quality household income data for rural households for multiple years (5 years in CHIP 2002, and 3 years in CHIP 1995).[41] Unlike China, the data on household income in rural India are, however, more limited; we are not aware of any household survey data set that has good quality income information for consecutive multiple years, similar to the CHIP data on China. We thus take household expenditure reported in the National Sample Survey as our measure of household permanent income.

Table 4 (panel A) provides estimates of household-level returns to education, R^{pn} and R^{pf}, in rural China and tests the null hypothesis that $R^{pn} = R^{pf}$ using data from two rounds (1995 and 2002) of the Chinese Household Income Project (CHIP) survey. The standard errors are clustered at the primary sampling unit (county). The estimates based on the 5-year average income of a household in CHIP 2002 data in the last two columns of Table 4 show that the null hypothesis cannot be rejected with a P-value equal to 0.58.[42] The evidence from the 1995 data (three year average income) also delivers a similar conclusion: the null hypothesis cannot be rejected with a P-value of 0.33.[43] The conclusion that returns to education measured in permanent household income do not differ significantly across farm and nonfarm households in rural China during the study period is robust to inclusion of number of children in the household as a control (see online appendix Table A.5). *The evidence on returns to education when put together with the evidence on complementarity discussed earlier in Table 1 provide a theoretically consistent explanation: a lack of difference in returns to education across farm and non-farm occupations leads to separability between father's education and nonfarm occupation in determining son's schooling in rural China.*

Also, the magnitude of θ_2 is likely to be low in rural China during the relevant period (the children who went to school during 1970s-1990s) which would reduce the impact of any emerging advantage in favor of nonfarm households in returns to education. Recall that θ_2 is the efficiency of parental investment, determined primarily by the supply side of the education market such as availability of high-quality private schools. In China, the availability of private schools was limited; in 1996, only 4 percent of the schools in China were private (Kwong 1996). Most of the private schools in rural areas in the 1990s were primary schools with limited facilities and equipment, and they catered to children from the low-income households. At the secondary level, the private schools primarily met the demand by the students who were unsuccessful in the admission test given after grade 9 to screen for the senior secondary public schools (Lin 1999). This implies that, in contrast to many other countries, any quality advantage in education in rural China is associated with better quality public schools. While local financing and various types of fees increasingly played a role in public schools after the fiscal decentralization, it is unlikely to create a significant impact on the magnitude of θ_2 for the following reason: the share of private expenditure remained small compared to the public expenditure during the 1980s and 1990s; for example, tuition and

[40] The available estimates on Mincerian labor market returns to education at the individual level in China show low returns in the early years after the reform, but there is evidence of increasing returns in the later years, as one would expect with the deepening of the labor market. The evidence also suggests higher labor market returns in nonfarm occupations (de Brauw and Rozelle 2008).

[41] Note that the estimates of the effects of father's education on household permanent income using CHIP data do not suffer from truncation bias, unlike the estimates of intergenerational persistence; whether some of the children were nonresident at the time of the survey is not relevant for this analysis.

[42] The 5 year income data cover from 1998 to 2002 in the CHIP 2002.

[43] The 3 years income data in CHIP 1995 cover 1991, 1993, and 1995.

other fees paid by the parents amount to only 4.42 percent of total educational expenditure in 1991 and 10.72 percent in 1995 (see Table 7.2 in Hannum et al. 2008).

Rural India

For the estimates of IGRCs across farm and nonfarm households in India to be consistent with the extended Becker-Tomes model of Section (2), the returns to schooling in nonfarm households need to be *higher* than that in the farming households in the parental generation. Table 4 (panel B) reports the estimates of household-level returns to education in rural India using household expenditure data from the NSS 1993 survey (the employment and unemployment round). The returns to education are, in fact, *higher* in nonfarm occupations and the difference is significant at the 5 percent level (P-value= 0.02). The conclusion that returns to education at the household level are higher for the nonfarm activities is robust; for example, the null hypothesis of equality is rejected with a P-value=0.002 when we control for number of children in the household (see the online appendix Table A.5).

The available evidence also suggests that the magnitude of θ_2 is likely to be much higher in rural India when compared to that in rural China. A higher value of θ_2 would act as a multiplier for higher returns to schooling in nonfarm activities for the parents, and amplify the difference between farm and nonfarm slopes (IGRCs). This can lead to the complementarity we found earlier in Table 1 above. In India, private schools have historically been more important than in rural China, and they have become more important over time, especially after the liberalization in 1991. Muralidharan and Kremer (2008) report that, in 2003, 28 percent of rural households had access to fee-charging private schools. They also provide evidence that private schools are more likely to be established in places where public school quality is low, and the students in private schools perform better academically.[44] Thus, the relative quality of private and public schools in rural India is opposite to that in rural China. This suggests that the higher income (and better educated) households can take advantage of the high-quality private schools making θ_2 higher. Since the private schools are more likely to locate in villages where the public school quality is low, the differential effects of school quality are likely to be strong in rural India, as the better educated nonfarm parents with high income send children to private schools, and the other children (including the children of low-educated and low-skilled nonfarm parents) go to low quality public schools.

6.2 Differences in the intercepts

According to the theory, the estimated intercepts in Table 2 discussed above imply the following relations (using a hat to denote an estimate):

$$\begin{aligned}
(China)\quad \hat{\psi}_0^f \approx \hat{\psi}_0^n \Longrightarrow Y_0^{pf} + \frac{1}{2\alpha_2}\left[\sigma\theta_2 E\left(RI^{cf}\right) - \alpha_1\right] \approx Y_0^{pn} + \frac{1}{2\alpha_2}\left[\sigma\theta_2 E\left(RI^{cn}\right) - \alpha_1\right]\\
(India)\quad \hat{\psi}_0^f > \hat{\psi}_0^n \Longrightarrow Y_0^{pf} + \frac{1}{2\alpha_2}\left[\sigma\theta_2 E\left(RI^{cf}\right) - \alpha_1\right] > Y_0^{pn} + \frac{1}{2\alpha_2}\left[\sigma\theta_2 E\left(RI^{cn}\right) - \alpha_1\right]
\end{aligned} \tag{19}$$

where $E\left(RI^{cj}\right) = \left\{\pi_i^{nj} R^{cn} + \left(1 - \pi_i^{nj}\right) R^{cf}\right\}$ is the expected return to financial investment in son's education when the father is employed in occupation $j = n, f$.

Rural China

[44] Private schools have more teachers with college degree and teacher absenteeism is less of a problem compared to the public schools. Azam et al. (2016) find that the students in private secondary schools in rural Rajasthan scored about 1.3 standard deviation (SD) higher than their counterparts in the public schools in a comprehensive standardized math test.

The following observations are important for understanding the role played by occupational persistence in educational mobility in rural China. First, when $\pi_i^{nn} \simeq \pi_i^{nf}$, we have $E\left(RI^{cn}\right) \simeq E\left(RI^{cf}\right)$, irrespective of whether $R^{cn} > R^{cf}$ or $R^{cn} \leq R^{cf}$. Since expected returns to education for the children do not vary significantly across farm and nonfarm households in this case, we would expect parental investment in education and thus educational mobility to be similar also. The second observation is that when there is low or no intergenerational persistence in nonfarm (or farm) occupations, we have $\pi_i^{nn} \simeq \pi_i^{nf}$.

A substantial body of independent evidence, in fact, suggests that, for the relevant cohorts, there was no significant intergenerational persistence in nonfarm occupation choices ($\pi_i^{nn} \simeq \pi_i^{nf}$) in rural China. Wu and Treiman (2007) use the 1996 national probability sample of Chinese men and show that there is high degree of *mobility into agriculture*; the sons of nonfarm parents also face a substantial probability of becoming a farmer. They identify the geographic restrictions on mobility of rural people because of the Hukou registration system as the primary factor behind this weak intergenerational persistence in nonfarm occupations.[45] Using CHIP 2002 data, Emran and Sun (2015a) report evidence supporting (Wu and Treiman 2007) finding.

The evidence that $E\left(RI^{cn}\right) = E\left(RI^{cf}\right)$, along with Eq. 19, above implies that a sufficient condition for the equality of the intercepts of the intergenerational persistence equations is that $Y_0^{pn} = Y_0^{pf}$, i.e., the intercepts of the returns to education function in the parent's generation are the same across farm and nonfarm households. We would expect $Y_0^{pn} = Y_0^{pf}$ when the fathers with zero schooling have similar income (permanent income) and face similar credit constraint, irrespective of their occupation.

The estimates in panel A of Table 4 show that the null hypothesis $Y_0^{pn} = Y_0^{pf}$ cannot be rejected at the 10 percent level with a p-value of 0.78 for the CHIP 2002 data on five-year average income. The evidence from 1995 data is also similar (p-value is 0.81). Again, these conclusions from CHIP 2002 and CHIP 1995 remain intact when we include the number of children in the household as a control (please see online appendix Table A.5).

The evidence above is also consistent with other available studies on the nonfarm sector and rural industries (TVEs) in rural China. The income gap between the farm and nonfarm households was mitigated in the early years of reform by two factors: the household responsibility reform increased the farmer's income, and, in many cases, people employed on the farm were paid wages similar to the wages paid to workers in the township village enterprises (TVEs), the growing TVE sector in effect subsidizing the agricultural employment (Peng, 1998). This also reflects in part the lingering effects of policies during the cultural revolution that were successful in eliminating any significant differences between the peasants and non-peasants in rural China (Hannum et al. (2008)).

Rural India

In contrast to China, there were no restrictions on rural-urban migration in India during the study period. A substantial body of independent evidence on occupational mobility in rural India suggests strong intergenerational persistence in farm/nonfarm occupations (Reddy 2015; Motiram and Singh 2012; Azam and Bhatt 2015; Hnatkovska et al. 2013). Hnatkovska et al. (2013) show that there is strong persistence in rural occupations both in 1983 and 2004-

[45] The link between restrictions on geographic mobility of rural people and a lack of intergenerational occupational persistence (farm/nonfarm) is, however, not unique to China, similar evidence is available on Vietnam where the Ho Khau registration system has been in place since 1964; see the evidence and the analysis in Emran and Shilpi (2011). This enhances the credibility of Wu and Treiman (2007) analysis that the Hukou restrictions played an important role in the low occupational persistence in rural China.

2005; the son of a farmer is highly likely to be a farmer himself. Using the IHDS (2005) survey, Motiram and Singh (2012) also provide similar evidence.

The fact that there was significant persistence in farm/nonfarm occupations implies that the expected returns to investing in children's education are likely to differ across farm and nonfarm households. But whether the intercept in the nonfarm households would be higher or lower depends partly on the expected relative returns to education in the children's generation, i.e., whether $R^{cn} > R^{cf}$ or $R^{cn} < R^{cf}$. It is, however, much more difficult to estimate *expected returns* to education for children. One can argue that a parent's expectation would depend on his/her information set, a salient element in which is his/her own returns to education. In other words, the evidence of higher returns to education in nonfarm occupations in the parental generation suggests that the parents would expect similarly higher returns for children in nonfarm occupations. Note that even when $R^{cn} > R^{cf}$, the intercept for the nonfarm households can be smaller as we find in the empirical analysis above (Table 2), if the households with zero (or very low) parental schooling have sufficiently lower income in nonfarm occupations, i.e., $Y_0^{pn} < Y_0^{pf}$.

The estimates of Y_0^{pn} and Y_0^{pf}, i.e., the intercepts of the income equation for parents, using data from NSS 1993 round (employment and unemployment round), are reported in panel B of Table 4. The estimated intercept is *larger* for the farm households and the difference is statistically significant at the 10 percent level (standard errors clustered at the PSU level). The evidence in favor of a larger intercept in the farm households is stronger when we add controls to the regression; for example, the difference is significant at the 1 percent level when number of children is added to the specification (see online appendix Table A.5). This conclusion is also supported by other available evidence on India; for example, the All India Debt and Investment Survey 1991 (NSS 48th round) shows that the assets of farming households ("cultivators") are higher than those of the nonfarming households (see P. ii, NSSO report No. 491, 1998).

The evidence in panel B of Table 4 also accords well with a substantial body of related evidence available on the nonfarm activities in India. Note that it is likely to have $Y_0^{pn} < Y_0^{pf}$ if the low-end nonfarm occupations are primarily low productivity residual activities and provide the last resort for the poorest households. Lanjouw and Murgai (2009) use three rounds of NSS data (1983, 1993/94, and 2004/2005) and show that nonfarm employment is positively associated with rural poverty in India, consistent with the observation that nonfarm employment involves primarily low productivity economic activities (see also World Bank (2011)).

This evidence on the intercepts of the income equation provides an explanation for the higher intercept for farm households in the intergenerational mobility equation as discussed earlier.

7 Evolution of mobility: evidence on the younger generation in rural China

The rural economy and educational policies in China went through significant changes in recent decades. The changes include gradual relaxation of Hukou restrictions, increasing returns to education as the labor market matured, accelerated structural change in favor of the nonfarm sector, and a more prominent role for private educational expenditure after the fiscal decentralization (for an excellent discussion, see Fan et al. (2020)).[46] There is credible evidence that the economic changes have adversely affected intergenerational mobility of the younger generation in China. Fan et al. (2020) find that intergenerational *income persistence* has increased over time; they report an IGE estimate of 0.390 for the 1970-1980 birth cohorts, which increased to 0.442 for the 1981-1988 birth cohorts. The extended Becker-Tomes model suggests that the pattern of educational persistence across farm and nonfarm households should also change for the younger generation because of the changes in the economic mechanisms.

To check if the pattern of educational mobility across farm and nonfarm households has changed in the younger generation, we estimate the intergenerational educational mobility equation for 18-28 years age cohorts who were excluded from our main estimation sample.[47] The estimates are reported in Table 5, with the upper panel containing the results for years of schooling, and the lower for rank-based estimates. There is, in fact, evidence of emerging divergence between farm and nonfarm households in relative mobility as measured by IGRC and IRRS. For example, the IGRC estimate for the farm households has declined a bit from 0.31 (main sample) to 0.27 (younger sample), while the IGRC estimate has increased for the nonfarm households from 0.316 (older) to 0.34 (younger). Similarly, the IRRS estimate for farm households declined from 0.331 (main sample) to 0.304 (younger sample), and it has increased from 0.337 (main sample) to 0.373 (younger sample) for the nonfarm households. The difference between farm and nonfarm households is significant at the 10 percent level in the case of IRRS estimates (P-value=0.097). While the difference in IGRC estimates is not significant at the 10 percent level, the p-value in the younger sample is much smaller: 0.107 (younger sample) vs. 0.90 (main sample), providing evidence in favor of emerging complementarity between a father's education and nonfarm occupation.

8 Conclusions

This paper develops a model of intergenerational educational persistence in a rural economy taking into account the role of parental farm vs. nonfarm occupations, and derives a theoretically-grounded estimating equation which we take to the data from rural India and rural China. We use two unique data sets that include the required information for all the children of the household head irrespective of their residency status at the time of the survey; thus eliminating the truncation bias common in the existing studies based on the standard surveys that use coresidency criteria to define household membership.

The empirical analysis delivers the following conclusions for the sons who went to school in the 1990s or earlier: (i) the intergenerational educational mobility in rural China was significantly higher compared to that in rural India, (ii) the farm/nonfarm occupations did not play any significant role in the intergenerational schooling linkage in rural China, and this is true for both the intercept and the slope of the intergenerational persistence regressions, (iii) both the slopes and intercepts were significantly different across farm and nonfarm households in rural India. The estimates suggest that a father's education and nonfarm occupation were complementary in determining son's schooling in rural India, but separable in rural China.

[46] The share of tuition and miscellaneous fees in educational expenditure rose from 4.42 percent in 1991 to 18.59 percent in 2004 (Hannum et al. (2008)). The spread of better quality public schools to the rural areas has accelerated. All these changes would increase the magnitude of θ_2 in the extended Becker-Tomes model for the younger generation.

[47] Our main estimation sample consists of 29-65 age cohorts in 2010 to ensure that the sample of children in China refers to the same age groups as in the data for India.

Structural change in favor of the nonfarm sector contributed to higher educational inequality in rural India during the study period, in part due to the complementarity. In contrast, such structural change was not an important factor in schooling inequality in rural China. Evidence from an approach that combines the biprobit sensitivity analysis of Altonji et al. (2005) with recent evidence on intergenerational correlation in cognitive ability suggests that the observed educational persistence in rural India is unlikely to be due solely to mechanical transmission of ability across generations, while the persistence in rural China could be explained by genetic correlations alone.

We analyze whether the economic forces identified in the extended Becker-Tomes model provide a coherent explanation of the observed pattern of educational mobility across countries (rural China vs. rural India) and over time (older vs. younger cohorts in rural China). A lack of intergenerational persistence in nonfarm occupations in rural China because of the Hukou restrictions seems to have played an important role in making the intercepts similar in rural China, but strong intergenerational occupational persistence in rural India resulted in significant differences between farm and nonfarm households. In rural India, the observed complementarity can be explained by higher returns to education in nonfarm occupations in the parental generation. The separability between a father's education and occupation in rural China was driven by the absence of any significant differences in the household-level returns to education across farm and nonfarm occupations. However, because of economic forces unleashed by the policy reform in China, the returns to education in nonfarm occupations for parents have increased and Hukou restrictions have been relaxed progressively. According to the theory, this should tighten the link between father's education and son's schooling in the nonfarm sector. Indeed, we find evidence that the separability between a father's education and occupation broke down for the 18-28 years old sons, implying that structural change in favor of the nonfarm sector is increasingly contributing to educational inequality in rural China.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10888-023-09599-1.

Acknowledgements We are grateful to Forhad Shilpi for help with REDS data and insightful comments throughout this project, and to Yang Huang for help with the CFPS data. We would like to thank the participants in Equal Chances conference 2018, and LACEA conference 2019, and Matthew Lindquist, Guido Neidhofer, Reshad Ahsan, Koen Decancq, and Hanchen Jiang for helpful comments on an earlier draft and Rakesh Gupta Nichanametla Ramasubbaiah for excellent research assistance. An earlier version of the paper was titled "Intergenerational Educational Mobility in the Rural Economy: Evidence from China and India". This project was partially funded by World Bank RSB. The standard disclaimers apply.

References

Acemoglu, D., Autor, D.: Lectures in Labor Economics. Book Manuscript, MIT (undated)

Ahsan, Md.N., Emran, M.S., Shilpi, F.J.: Public primary school expansion, gender-based crowding out, and intergenerational educational mobility. Policy Research Working Paper Series 10418, The World Bank (2023)

Ahsan, N.M., Emran, M.S., Jiang, H., Murphy, O., Shilpi, F.J.: When measures conflict: Towards a better understanding of intergenerational educational mobility. Working paper, The World Bank (2022)

Ahsan, N., Emran, M.S., Shilpi, F.: The rural-urban divide and intergenerational educational mobility in a developing country: Theory and evidence from Indonesia. Working Paper, World Bank (2020)

Altonji, J.G., Elder, T.E., Taber, C.R.: Selection on observed and unobserved variables: Assessing the effectivenessof catholic schools. J. Polit. Econ. **113**(1), 151–84 (2005). https://doi.org/10.1086/426036

Asher, S., Novosad, P., Rafkin, C.: Intergenerational mobility in India: Estimates from new methods and administrative data. Working Paper, Dartmouth College (2018)

Azam, M., Bhatt, V.: Like father, like son? Intergenerational educational mobility in India. Demography **52**(6), 1929–1959 (2015)

Azam, M., Kingdon, G., Wu, K.B.: Impact of private secondary schooling on cognitive skills: evidence from India. Educ. Econ. Taylor & Francis J. **24**(5), 465–480 (2016)

Bardhan, P.: The State of Indian Economic Statistics: Data Quantity and Quality Issues. Unpublished (2014)

Becker, G.S.,Tomes, N.: Human capital and the rise and fall of families. J. Labor. Econ. **4**(3), 1–39 (1986). University of Chicago Press

Becker, G. S., Kominers, S. D., Murphy, K. M., Spenkuch, J. L.: A theory of intergenerational mobility. J. Polit. Econ. **126**, S7–S25 (2018)

Becker, G., Tomes, N.: An equilibrium theory of the distribution of income and intergenerational mobility. J. Polit. Econ. **87**(6), 1153–1189 (1979)

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

Behrman, J., Hu, Y., Zhang, J.: The causal effects of parents' schooling on children's schooling in urban China. PIER Working Paper Archive 20-005, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania (2020)

Behrman, J.: Human capital and social mobility in low-and middle-income countries. WIDER Working Paper Series wp-2019-85, World Institute for Development Economic Research (UNU-WIDER) (2019)

Behrman, J., Gaviria, A., Szekely, M.: Intergenerational mobility in latin America. Economia **2**(1), 1–44 (2001)

Bhalotra, S., Rawlings, S.: Gradients of intergenerational transmission of health in developing countries. Rev. Econ. Stat. **95**(2), 660–672 (2013)

Bhattacharya, D, Mazumder B.: A nonparametric analysis of black-white differences in intergenerational income mobility in the United States. Quant. Econ. **2**, 335–379 (2011)

Bjorklund, A., Salvanes, K.: Education and family background: Mechanisms and policies. In: Hanushek, E.A., Machin, S., Woessmann, L. (eds.) Handbook in the Economics of Education vol 3 pp. 201–247 (2011)

Bjorklund, A., Lindahl, L., Lindquist, M. J.: What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents. The B.E. Journal of Economic Analysis and Policy **10**(1), 42 (2010)

Black, S.E., Devereux, P.: Recent developments in intergenerational mobility. Handbook of Labor Economics (2011)

Black, S.E., Devereux, P.J., Salvanes, K.G.: Like father, like son? A note on the intergenerational transmission of IQ scores. Econ. Lett. **105**(1), 138–140 (2009)

Chetty, R., Hendren, N., Kline, P., Saez, E.: Where is the land of opportunity? The geography of intergenerational mobility in the United States. Q. J. Econ. **129**(4), 1553–1623 (2014)

de Brauw, A., Rozelle, S.: Reconciling the returns to education in rural China. Rev. Dev. Econ. **12**(1), 57–71 (2008)

Emran, M.S., Greene, W., Shilpi, F.: When measure matters: Coresidency, truncation bias, and intergenerational mobility in developing countries. J. Hum. Resour. **53**(3), 589–607 (2018). University of Wisconsin Press

Emran, M.S., Islam, A., Shilpi, F.: Distributional effects of corruption when enforcement is biased: Theory and evidence from bribery in schools in Bangladesh. Economica **87**(348), 985–1015 (2020a). London School of Economics and Political Science

Emran, M.S., Jiang, H., Shilpi, F.: Gender bias and intergenerational educational mobility: Theory and evidence from China and India. Policy Research Working Paper Series 9250, The World Bank (2020b)

Emran, M.S., Jiang, H., Shilpi, F.: Is gender destiny? Gender bias and intergenerational educational mobility in India. GLO Discussion Paper Series 807, Global Labor Organization (GLO) (2021)

Emran, M.S., Shilpi, F.: Economic approach to intergenerational mobility: Measures, methods, and challenges in developing countries. WIDER Working Paper Series wp-2019–98, World Institute for Development Economic Research (UNU-WIDER) (2019)

Emran, M.S., Shilpi, F.J.: Estimating intergenerational mobility with incomplete data: Coresidency and truncation bias in rank-based relative and absolute mobility measures. Policy Research Working Paper Series 8414, The World Bank (2018)
Emran, M.S., Sun, Y.: Are the children of uneducated farmers doubly disadvantaged? Farm, nonfarm and intergenerational educational mobility in rural China. Policy Research Working Paper Series 7459 (2015b)
Emran, M.S., Sun, Y.: Magical transition? Intergenerational educational and occupational mobility in rural China: 1988–2002. World Bank Policy Research, Working Paper 7316 (2015a)
Emran, M.S., Shilpi, F.: Intergenerational occupational mobility in rural economy: Evidence from Nepal and Vietnam. J. Hum. Resour. **46**(2), 427–458 (2011)
Emran, M.S., Shilpi, F.: Gender, geography and generations: Intergenerational educational mobility in post-reform India. World Dev. **72**, 362–380 (2015)
Fan, Y., Yi, J., Zhang, J.: Rising intergenerational income persistence in China. Am. Econ. J. Econ. Policy **13**(1), 202–30 (2021)
Hannum, E., Behrman, J., Wang, M., Liu, J.: Education in the reform era. Chapter 7 in Brandt and Rawski (eds.) China's Great Economic Transformation. Cambridge University Press (2008)
Hertz, T.: Rags, riches, and race: the intergenerational economic mobility of black and white families in the United States. In: Bowles, S., Gintis, H, and Osborne, G.M. (eds.) Unequal chances: Family background and economic success. Princeton, Princeton University Press (2005)
Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N., Veraschagina, A.: The inheritance of educational inequality: International comparisons and fifty-year trends. B.E. J. Econ. Anal. Policy Adv. **7**(2) (2007)
Hnatkovska, V., Lahiri, A., Paul, S.B.: Breaking the caste barrier: Intergenerational mobility in India. J. Hum. Resour. **48**(2), 435–473 (2013). University of Wisconsin Press
Knight, J., Sicular, T., Yue, X.: Educational inequality in China: The intergenerational dimension. In: Shi, L., Chuliang, L., Sicular, T. (eds.) Rising Inequality in China: Challenge to the Harmonious Society, ch.4. Cambridge and New York: Cambridge University Press. (2013)
Kwong, J.: The new educational mandate in China: running schools, running businesses. Int. J. Educ. Dev. **16**(2): 185–194 (1996)
Lanjouw, J.O., Lanjouw, P.: The rural non-farm sector: issues and evidence from developing countries. Agric. Econ. Blackwell **26**(1), 1–23 (2001)
Lanjouw, P., Murgai, R.: Poverty decline, agricultural wages, and nonfarm employment in rural India:1983–2004. Agric. Econ. **40**(2), 243–263 (2009)
Lanjouw, P., Murgai, R., Stern, N.: Nonfarm diversification, poverty, economic mobility, and income inequality: a case study in village India. Agric. Econ. **44**(4–5), 461–473 (2013)
Lefgren, L., Lindquist, M.J., Sims, D.: Rich dad, smart dad: Decomposing the intergenerational transmission of income. J. Polit. Econ. **120**(2), 268–303 (2012). University of Chicago Press
Li, H., Loyalka, P., Rozelle, S., Wu, B.: Human capital and China's future growth. J. Econ. Perspect. **31**(1), 25–48 (2017)
Lin, J.: Social transformation and private education in China. Praeger, CT, USA (1999)
Mazumder, B.: Black white differences in intergenerational economic mobility in the U.S. Economic perspectives, federal reserve Bank of Chicago, issue Q I, pages 1-18 (2014)
Mogstad, M.: The human capital approach to intergenerational mobility. J. Polit. Econ. **125**(6), 1862–1868 (2017). University of Chicago Press
Motiram, S., Singh, A.: How close does the apple fall to the tree? Some evidence on intergenerational occupational mobility from India. WIDER Working Paper Series 101, World Institute for Development Economic Research (UNU-WIDER) (2012)
Muralidharan, K., Kremer, M.: Public-private schools in rural India. In: Chakrabarty, R., Peterson, P. (eds.) School Choice International: Exploring Public-Private Partnerships, MIT Press (2008)
Narayan, A., Van der Weide, R., Cojocaru, A., Lakner, C., Redaelli, S., Mahler, D.G., Ramasubbaiah, R.G.N., Thewissen, S.: Fair progress? Economic mobility across generations around the world. World Bank Publications, The World Bank, number 28428, (2018)
Neidhofer, G., Serrano, J., Gasparini, L.: Educational inequality and intergenerational mobility in latin America: A new database. J. Dev. Econ. **134**, 329–349 (2018)
Neslehova, J.: On rank correlation measures for non-continuous random variables. J. Multivar. Anal. **98**(3), 544–567 (2007)
Nybom, M., Stuhler, J.: Biases in standard measures of intergenerational income dependence. J. Hum. Resour. **52**(3): 800–825 (2017)
Park, A., Zou, X.: Intergenerational mobility pathways: Evidence from a long panel from rural China, Working paper, HKUST (2019)
Peng, Y.: Agriculture, rural industries, and peasant income in China. Development and Society (1998)

Plomin, R., Spinath, F.: Intelligence: genetics, genes, and genomics. J. Pers. Soc. Psychol. **86**(1): 112–129 (2004)
Rama, M., Béteille, T., Li, Y., Mitra, P.K., Newman, J.L.: Addressing inequality in south Asia. World Bank Publications, The World Bank, number 20395, (2015)
Ravallion, M. and Datt, G.: Why has economic growth been more pro-poor in some states of India than others? J. Dev. Econ. **68**(2): 381–400 (2002)
Reddy, A.B.: Changes in intergenerational occupational mobility in India: Evidence from national sample surveys, 1983–2012. World Development **76**(C), 329–343 (2015)
Rozelle, S.: Rural industrialization and increasing inequality: Emerging patterns in China's reforming economy. J. Comp. Econ. **19**(3), 362–391 (1994)
Sato, H., Li, S.: Class origin, family culture, and intergenerational correlation of education in rural China. IZA Discussion Paper series No. 2642 (2007)
Solon, G.: Intergenerational mobility in the labor market. In: Ashenfelter, O., Card, D. (eds.) Handbook of Labor Economics, Vol. 3, Elsevier: Amsterdam, North Holland (1999)
Solon, G.: Intergenerational Income Mobility in the United States. Am. Econ. Rev. **82**(3), 393–408 (1992)
Solon, G.: A model of intergenerational mobility variation overtu time and space. In: Corak, M. (ed.) Generational Income Mobility. Cambridege University Press (2004)
Torche, F.: Intergenerational mobility and equality of opportunity. Eur. J. Sociol. **56**(3), 343–371 (2015)
Wu, X., Treiman, D.J.: Inequality and equality under chinese socialism: The hukou system and intergenerational occupational mobility. Am. J. Sociol. **113**(2), 415–45 (2007)
Yang, D.T., An, M.Y.: Human capital, entrepreneurship, and farm household earnings. J. Dev. Econ. Elsevier **68**(1), 65–88 (2002)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.