Intergenerational Occupational Mobility in a Rural Economy

M. Shahe Emran
George Washington University
and IPD, Columbia University

Forhad Shilpi
DECRG, World Bank

This Version: June 2007

ABSTRACT

This paper presents evidence on intergenerational occupational mobility from agriculture to the nonfarm sector using survey data from Nepal. In the absence of credible instruments, the degree of selection on observables is used as a guide to the degree of selection on unobservables á la Altonji et. al. (2005) to address the unobserved genetic correlations relevant for occupation choice. The results show that a moderate ability correlation can easily explain away the observed partial correlation in non-farm participation between the father and a son. In contrast, the partial correlation in occupation choice between mother and daughter is much stronger, and is unlikely to be driven by genetic correlations alone. The results suggest that mother’s nonfarm participation plays a causal role in daughter’s choice of nonfarm occupation, possibly because of “cultural inheritance” through role model and learning effects, and transfer of reputation and social capital.

Keywords: Intergenerational Occupational Correlations, Non-Farm Participation, Gender effect, Cultural Inheritance, Selection on Observables, Selection on Unobservables

JEL Codes: J62, O12

---

1We are grateful to Todd Elder and Chris Taber for very useful comments on an earlier draft, especially regarding the econometric issues. We would also like to thank Paul Carrillo, Bob Goldfarb, Don Parsons, Dilip Mookherjee, Bhashkar Mazumder, David Ribar, and Stephen Smith for useful comments, and Todd Elder for sharing STATA and MATLAB routines for implementation of the econometric estimation. The standard disclaimers apply. Emails for correspondence: shahe.emran@gmail.com (M. Shahe Emran), fshilpi@worldbank.org.
Introduction

The evolution of income distribution, inequality and occupational structure across generations has attracted increasing attention in recent economic literature\(^2\). This renewed interest reflects a widely shared view that strong intergenerational linkages in socioeconomic status may reflect inequality of opportunities and thus have profound implications for poverty, inequality and (im)mobility in a society. A large body of econometric studies focusing mainly on developed countries finds that intergenerational correlations in *earnings* are positive and statistically significant, ranging between 0.14 to 0.50 (see Blanden et. al. (2005) and Solon (1999, 2002)). There is a (relatively) small empirical literature in economics, again mostly in the context of developed countries, that indicates significant positive correlations between parents and their children in occupational choices (see, for example, Lentz and Laband (1983), and Dunn and Holtz-Eakin (2000) on U.S and Sjogren (2000) on Sweden). Economic analysis of intergenerational mobility in developing countries, however, remains a relatively unexplored terrain\(^3\), even though the importance of such analysis has been duly recognized in the recent literature.\(^4\) In this paper, we present evidence on the intergenerational occupational correlations in the non-farm sector in a developing country, Nepal.\(^5\) Although there is a substantial literature on the determinants of non-farm participation (see Lanjouw and Feder (2001) for a recent survey), the issue of inter-

\(^2\)See, for example, Arrow et al. (2000), Dearden et. al. (1997), Mulligan (1999), Solon (1999, 2002), Birdsall and Graham (1999), Fields (2001), Fields et. al. (2005), Bowles et. al. (2005), Blanden et. al. (2005), WDR (2005), Mazumder (2005), Hertz (2005), Mookherjee and Ray (2006), Bjorklund et. al. (2006).

\(^3\)This is exemplified by the fact that Solon (2002) refers to only two studies on developing countries in his survey of economic mobility (Lillard and Kilburn (1995) on Malaysia, and Hertz (2001) on South Africa). The recent analysis of economic mobility in the context of developing countries include Lam and Schoeni (1993), Behrman et. al. (2001), Fields et. al. (2005), Dunn (2004). There is, however, a substantial sociological literature that analyzes occupational mobility in both developed and developing countries (see, Ganzboom et. al. (1991), Morgan, (2005)).

\(^4\)For example, Bardhan (2005) identifies intergenerational economic mobility as one of the important but under-researched areas in development economics.

\(^5\)As emphasized by Goldberger (1989), there are some important advantages to focusing on occupational mobility rather than income mobility. For example, the intergenerational linkages might be stronger for occupation choice (relative to income), and focusing on income correlations “could lead an economist to understate the influence of family background on inequality” (P.513).
tergenerational linkages has so far not received any attention. Understanding occupational mobility in a rural economy is, however, important for poverty alleviation, as mobility from agriculture to nonfarm is often an avenue to escape poverty trap (WDR (2005); Lanjouw and Feder (2001)). In the presence of strong intergenerational linkage in occupational choice, the standard cost-benefit analysis is likely to under-estimate the long-run social returns to policy interventions that encourage non-farm participation, as the intergenerational multiplier effect is ignored. Non-farm participation often leads to ‘visible’ income contribution by women and thus positively affects their bargaining power.\(^6\) A different but powerful argument derives from the role of non-farm entrepreneurship in the long-run structural transformation of an economy. A dynamic non-farm sector can be the seedbed for experimentation and development of an entrepreneurial class that eventually graduates to industrial activities, as was the case in Japan’s rise to a modern industrial state from late Tokugawa to Meiji period (See Smith (1988)).

Our focus in this paper is on two issues: (i) intergenerational occupational persistence beyond the observed determinants like education and assets, and the unobserved genetic correlations across generations, and (ii) gender effects in occupational mobility. The literature on the intergenerational economic mobility has been fraught with econometric challenges that arise from the unobservability of the genetic characteristics (ability and preference transmissions across generations), and the partial correlations observed in the data (from multivariate regressions) might be driven largely by such unobserved genetic correlations between parents and children.\(^7\) The distinction between genetic transmissions and other sources of intergenerational linkages can be important from policy perspective. If the observed intergenerational occupational correlation is primarily due to ability correlations, then the role public policy can play to influence economic mobility is relatively

\(^6\)Women’s work in agriculture is usually unpaid and remains invisible in developing countries.

\(^7\)Genetic transmissions relevant for occupational choice include both ability and preference (especially the degree of risk aversion). However, the focus of the literature has been on ability correlations. In what follows, we couch the discussion primarily in terms of ability correlations, following the literature.
limited.\textsuperscript{8} In contrast, when other tangible (like education and wealth) or intangible (like role model effects\textsuperscript{9}, learning and reputation externalities and social capital) environmental factors are important in intergenerational linkages, it provides additional arguments for policy interventions.

In the absence of experimental data\textsuperscript{10}, the standard approach to identification when facing unobserved heterogeneity like ability correlations is to look for credible instrumental variables (IV). In the specific context of occupational mobility, the econometric challenge is to find exogenous variation that affects parental occupation choice but does not have any independent effect on children's occupation choice. However, most of the potential candidates for IV such as family background variables that affect parent's occupation choice tend to affect children's choice also. Thus it is difficult to defend the exclusion restriction. Moreover, the common practice of using parental characteristics (like parental education) as IVs is also suspect, as they are likely to be correlated with the unobserved common ability, and thus likely to violate the exogeneity criterion. There is a small literature in economics that uses adoption as a quasi-experimental design to isolate the effects of environmental factors in intergenerational economic mobility (see, for example, Bjorklund et. al. (2006); Plug (2004); Plug and Vijverberg (2003), Scaerdote (2002)). A third strategy is to use twin samples to try to isolate the effects of nurture from that of nature (see, for example, Behrman and Rosenzweig (2002)). However, these studies using adoption or twin samples are confined mostly in the developed countries where such data of reliable quality are available. In the absence of quasi-experimental data on adoptions and twins

\textsuperscript{8}We, however, note that one should not take the distinction between genetic transmissions and other environmental factors too far. The evidence from Behavioral Genetics shows that there may be significant dynamic interactions between nature and nurture in determining human behavior (see, for example, Plomin et. al. (2001)).

\textsuperscript{9}The definition of role model adopted so far in economic literature is not uniform. While Durlauf (2000) defines role model as the influence of “characteristics of older members” on the “preferences of younger members”, Manski (1993) and Streufert (2000) define it as observations on older members whose choices reveal information relevant for the choice of younger members. In this paper, we adopt a broad view that accommodates both of these definitions.

\textsuperscript{10}Designing and implementing a randomized experiment that can generate the data required for understanding the intergenerational occupational persistence can be challenging on both ethical and feasibility grounds.
or any credible identifying instruments, we exploit the econometric methodology recently developed by Altonji, Elder and Taber (2005, 2000) (henceforth AET (2005, 2000)) which provides a way to gauge the importance of common (across generations) unobserved heterogeneity in explaining an observed partial correlation and thus helps to determine if at least part of the observed partial correlation is causal (due to environmental factors). We note that genetic transmissions (ability and preference) influence both parents’ and children’s nonfarm participation decisions and hence can be treated as an unobserved (and correlated) determinant of nonfarm employment choices of both generations. This allows us to utilize a battery of recently developed econometric tests to ascertain whether the observed intergenerational occupational correlations can be attributed solely to the unobserved ability correlations between children and their parents. \footnote{Note that although the unobserved ability correlations have been the focus of much of the recent literature on mobility (and returns to education), the unobserved common factors in our context include other common determinants of occupation choice like social norms.}

The results from the econometric analysis are as follows. The univariate probit estimation indicates the presence of strong and positive intergenerational occupational correlations along gender lines (mother-daughter and father-son) even after controlling for a rich set of regressors including education (own, parents’ and spouse’s), assets (inherited land), age and ethnicity (caste/tribe). The estimated occupational linkages from univariate probit can, however, at least in principle, be due entirely to genetic transmissions across generations. The evidence from the econometric analysis using the AET (2005) methodology shows that this might actually be the case for the observed occupational correlation between the father and son; even a moderate correlation of unobserved ability can explain away the estimated occupational linkage completely. The intergenerational occupational correlation between mother and daughter, on the other hand, is much stronger and unlikely to be driven solely by unobserved ability correlations. The evidence thus suggests that at least part of the correlation between mother and daughter is likely to be causal due to “cultural inheritance” (Bjorklund, Jantti, and Solon (2007)) that includes role model effects, learning externalities.
and transfer of reputation and social capital from parents to children, among other things.

The substantive conclusions above are very robust, confirmed by alternative econometric techniques as developed by AET (2005, 2000): (i) sensitivity analysis using a bivariate probit model, and (ii) estimates of lower bounds on intergenerational occupational correlations. In case of daughters, the lower bound estimate of intergenerational occupational correlation with mother is 0.66 with an implied marginal effect of 0.14 and t-value of 4.8. The 95 percent confidence interval for the marginal effect is [0.07 0.24], which does not include zero. These results suggest that the genetic correlations account for about half of the partial correlation between the mother and a daughter given that the marginal effect in the univariate probit model is 0.30. The other half of the intergenerational correlation can be attributed to cultural inheritance by a daughter from her mother in the form of role model effects, learning externalities and transfer of reputation and social capital, among other things. In the case of sons, the lower bound estimate is negative and statistically insignificant which implies that the observed partial correlation may be driven entirely by the unobserved factors common to both generations. The results from the sensitivity analysis yield the same conclusions as above.

The rest of the paper is organized as follows. Section 2 provides a conceptual framework that underpins the empirical work presented in the subsequent sections. The section 3

\[^{12}\text{The empirical methodology proposed by AET (2005, 2000) can be used to provide a lower bound on intergenerational occupational correlation under the assumption that the ‘selection on observables’ is at least as large as the ‘selection on unobservables’. As we discuss in more detail later in the text, the assumption that the selection on observables dominates the selection on unobservables is a natural one in the context of intergenerational occupational mobility analyzed in this paper. The univariate probit model assumes “no selection on unobservables” and thus can be thought of as the upper bound estimate of the intergenerational correlation.}\]

\[^{13}\text{Note that the environmental factors like role model effects and learning externalities only affect the occupational choice of children and thus are NOT subsumed under the common intergenerational correlation.}\]

\[^{14}\text{Following AET (2005, 2000) we also estimate the bias in the partial correlation estimates from univariate probit. The estimates of the bias might be useful as robustness check as they are not dependent on distributional assumptions. However, as noted by AET (2005, 2000), the bias estimates are based on the strong assumption that the bias in the linear projection is similar to the bias in the probit equation. Since this assumption is difficult to justify, we chose not to present the bias estimates (available upon request). We, however, note that the bias estimates also lend strong support to the central conclusions discussed above.}\]

5
discusses the data and variables, while the next section presents some preliminary evidence. Section 4, arranged in a number of subsections, presents the main empirical results that focus on gauging the role played by unobserved common determinants of occupational choice across generations following the approach due to AET (2005, 2000). Section 5 concludes the paper with a summary of the main findings.

(2) The Conceptual Framework

In this section, we outline a simple model of participation in nonfarm sector highlighting different channels through which intergenerational linkages may operate, especially the role cultural inheritance might play through factors like role model and learning effects and transfer of reputation and social capital from parents to children. The model is based on the standard occupational choice model but is augmented to capture the essentials of the intergenerational linkages.15

There are two sectors in the economy: agriculture ($A$) and non-farm sector ($N$). At the beginning of the working life, every person in the economy decides which sector to work for. Each individual is endowed with an innate ability $\theta_i \in [0, 1]$ that captures the attributes that are relevant for non-farm sector. So the higher is $\theta_i$ the better suited an individual is for non-farm employment. A fundamental source of intergenerational linkage arises from the fact that the genetic endowments of a child ($\theta_i$) are likely to be correlated with those of parents. The innate ability parameter $\theta_i$ is not known with certainty and every individual has to form an estimate utilizing all the available information contained in an appropriately defined information set.

In addition to ability, every individual is endowed with a vector of initial capital stock $k_i$ comprised of human, financial and physical, and social capital. The higher is the level of $k_i$ the higher is the probability of getting a better paid non-farm job. Parents can influence this initial capital stock $k_i$ through their investment in a child’s human capital (e.g. education) and their transfer of financial and physical capital. In addition, parental

15The model utilized here can be viewed as an extension of the celebrated contributions of Becker and Tome (1979 and 1986) and the recent extensions proposed in Sjogren (2000).
occupation can also influence their offsprings’ human capital as children can gain valuable skills and experience by observing their parents at work, and by informal apprenticeship in parents’ workplace, especially when the nature of occupation is such that the workplace is in close proximity to home. The parents, when successful in self-employment, often transfer significant reputation capital and a rich social network (social capital) to their children. An individual’s earning from an occupation, $Y_i$, is determined by innate ability ($\theta_i$) and endowment of capital stock ($k_i$).

At the beginning of the working life, individual $i$ takes the endowment of capital and the estimate of ability ($k_i$, $\hat{\theta}_i$) as given, and optimally chooses the occupation $d_i \in \{A, N\}$. Let the information set available to individual $i$ choosing occupation is denoted as $\Omega_i$ which include $k_i$ and $\hat{\theta}_i$. Let $F(Y_i \mid A; \Omega_i)$ denote the conditional distribution of income ($Y_i$) when individual chooses agriculture and the information set is $\Omega_i$. The associated probability density function is denoted as $P(Y_i \mid A; \Omega_i)$. The preference of an individual $i$ is represented by a concave utility function, $U_i(\cdot)$, that reflects, among other things, the risk preference.

We define the expected utility from choosing agriculture as:

$$V_i(A, \Omega_i) \equiv \int U_i(Y_i)P(Y_i \mid A; \Omega_i)dY_i$$

Analogously the expected utility from choosing non-farm sector is:

$$V_i(N, \Omega_i) \equiv \int U_i(Y_i)P(Y_i \mid N; \Omega_i)dY_i$$

---

16 As noted by Lentz and Laband (1983), this proximity of work place to home is an important factor behind the observed strong intergenerational following in occupations like agriculture. This proximity is also important in case of the household based activities common in the microcredit programs.

17 The preferences of a child are likely to be correlated with those of her parents. In addition, parents can also induce changes in children’s preferences by acting as their role models (Durlauf, 2000). The intergenerational correlation in preferences implies, for example, that, on an average, the children of the parents more inclined to taking risk will themselves be risk takers, and thus are more likely to become non-farm entrepreneurs.
The individual chooses non-farm employment iff the following holds:  

\[ V_i(N, \Omega_i) - V_i(A, \Omega_i) \geq 0 \]  

(1)

The probability that an arbitrary individual \( i \) drawn from the population will decide to work in the non-farm sector is \( \Pr(V_i(N, \Omega_i) - V_i(A, \Omega_i) \geq 0) \). At the heart of the occupation selection process is the formation of expectation about payoffs from different options using the information set \( \Omega_i \). A critical element of the information set is the occupational choices of the parents as they reveal two types of relevant information: (i) information about one’s own genetic endowment (or innate ability), (ii) information about the characteristics of a certain occupation. For example, if parents are successful (unsuccessful) non-farm entrepreneurs, the estimate of children’s ability to be successful in similar occupation will be revised upward (downward). Another important channel is that revelation of information might reduce the uncertainty about the parental occupation, and thus induce risk-averse children to prefer the parental occupation to other alternatives. Thus, the information revealed by parental choices (and their outcomes) can influence children’s occupation decision through their effects on the conditional distribution function of income \( Y_i \) giving rise to role model effects (Manski 1993; Streufert, 2000). For example, consider a child’s participation decision in non-farm sector \( (d_i = N) \). The parental role model effects imply that the conditional distribution of income when parents are in non-farm \( F(Y_i \mid N; N^p, \Omega_i) \) is stochastically dominant over the conditional distribution of income with neither of the parents is in non-farm \( F(Y_i \mid N; A^p, \Omega_i) \).  

The model presented above can also be used to explain intergenerational correlations running along gender line. First, the genetic transmissions might have a gender dimension. For example, the preference of a daughter (son) is likely to be more aligned with that of her (his) mother (father) compared to that of her (his) father (mother). Second, and probably

---

\(^{18}\)Assuming that the tie is broken in favor of non-farm sector.

\(^{19}\)Note that given a concave utility function both first and second order stochastic dominance are sufficient.
the most important factor behind gender effects in intergenerational linkages in occupational choices, is the gender dimension in role model effects. The information revealed by the choices (and consequent outcomes) of an older member of a society will be more informative for the choices of a given younger member the closer he/she is to the younger person in an appropriately defined socioeconomic space. The individuals can be grouped together by partitioning the socioeconomic space according to different exogenous (like ethnicity, gender) or endogenous (like schooling) characteristics. The finer the partitioning the more informative is the information revealed by the choices of a member of a given group for the other members of that same group. It follows that, given the membership in a family, gender creates a finer partitioning, and the mother becomes the natural role model for the daughter, and the father for the son. This has also implications for learning by doing and observing as the daughter (son) ‘sees’ and ‘hears’ primarily what her (his) mother (father) does and says. Another potential channel for gender effects is that the effects of parental social capital might run predominantly along gender lines; the mother’s social network might be more easily accessible to a daughter. Moreover, social norms regarding gender roles might also contribute to gender effects in occupation choice. The existence of gender effects for a daughter means that the conditional distribution of income from non-farm employment when mother is in non-farm $F(Y_i \mid N; N^m, \Omega_i)$ stochastically dominates the conditional distribution with father in non-farm $F(Y_i \mid N; N^f, \Omega_i)$.

For the econometric estimation, we can now employ a standard probit model taking inequality (1) as the basis for our empirical specification. Specifically, we consider the binary response model (with slight abuse of notation):

$$N_i = 1 \{N^*_i \equiv V_i(N, \Omega_i) - V_i(A, \Omega_i) \geq 0\}, \quad (2)$$

For estimation we impose linearity and assume that the latent variable $N^*_i$ is generated from a model of the form
\[ N_i^* = \alpha_p N^p_i + X'_i \gamma + \varepsilon_i \] (3)

Where \( X_i \subseteq \Omega_i \) is a vector of explanatory variables and \( \varepsilon_i \) is the idiosyncratic random disturbance term. In the econometric analysis, the vector of explanatory variables \( X_i \) is required to include regressors that can control for heterogeneity across individuals in terms of preferences (\( U_i \)), and the productivity, and pay-off information contained in \( \Omega_i \). Equation (2) forms the basis of much of our empirical analysis. A complete list of explanatory variables \( X_i \) is provided in appendix table A.1.

(3) The Data

The data for our analysis come from the Nepal Living Standard Survey (NLSS) 1995/96. The NLSS consists of a nationally representative sample of 274 primary sampling unit (PSUs) selected with probability proportionate to population size, covering 73 of the 75 districts in Nepal. In each of the PSUs, 12 households were also selected randomly (16 households in the Mountain regions) providing a total sample size of 3373 households. With an average household size of about 5.6, the survey collected detail information for 18855 individuals. Of the total individual level sample of 18347 individuals for whom parents can be identified from the data, nearly 71 percent reported participation in the labor force, but about 20 percent did not report any occupation.\(^{20}\) For the rest of the sample (9417 observations), 10 percent are either child labor (less than 14 years of age, 9 percent) or too old (more than 70 years of age) and thus are dropped.\(^{21}\) But some of the parents did not either participate in the labor force or report their labor force participation, further reducing the size of the sample.\(^{22}\) Moreover, a number of the PSUs showed no employment

\(^{20}\) A large fraction of those not reporting occupation are in fact child labor with age 14 years or less.
\(^{21}\) Note that the empirical results reported in the following sections remain unchanged even if we use any other cut-off age (e.g. dropping those below 20 years of age and above 65 years of age and so on).
\(^{22}\) For 8394 individuals left in the sample, both parents reported participation in the labor force in the case of 6874 individuals, and for rest of the observations, employment status of either father (347 observations) or mother (1173) are missing. Note that if we use dummies to capture the missing parental information, the sample size can be increased but the qualitative results remain unaffected.
diversification which are dropped to avoid perfect fits in the regression analysis. Splitting the sample between males and females, we have a final sample of 2037 observations (in 152 PSUs) for daughters and 2919 observations (in 242 PSUs) for sons.

The NLSS contains detailed information on employment by sectors and by occupations at individual levels. The survey is unique in the sense that it contained an entire section of questionnaire on parental information, including level of education, sector of employment and place of birth. From the occupation information, we define our dependent variable as a binary variable taking the value of one if an individual is employed in nonfarm activities and zero otherwise. Similarly we define separate indicator variables for mother and father showing their employment in nonfarm sector.

Table 1 reports the basic statistics on employment status of daughters and sons. The (unconditional) probability of being employed in nonfarm sector is estimated to be around 44 percent for a man and 16 percent for a woman. In our data set, average participation rates of father and mother are around 20 percent and 8 percent respectively. A comparison of sons and daughters’ employment status conditional on father’s and mother’s employment status reveals that the probability of being employed in non-agriculture sector is markedly higher for both sons and daughters if father or mother were employed in non-agriculture as well. We also tested the significance of difference between probabilities of being employed in nonfarm sector by parent’s employment status (farm vs nonfarm). The test results reported in Table 1 indicate that in all cases, the null hypothesis of no difference can be rejected with a P-value equal to 0.00. According to Table 1, mother’s participation in non-farm sector appears to have a larger effect, compared with father’s non-farm participation, on both sons’ and daughters’ probability of participation in non-farm sector. The intergenerational occupational linkage appears to be much stronger for daughters relative to sons.

\[\text{Non-farm is defined as non-agricultural, i.e., excludes SIC one digit code ‘0’. Non-farm thus includes rural industries and services.}\]
Preliminary Evidence

With some indication of positive intergenerational correlations between parents’ and children’s occupational choices, we turn to more formal regression analysis. Starting from a simple Probit regression of son’s and daughter’s occupations on parental occupations, we take a sequential approach in presenting the results, introducing an array of control variables in subsequent steps. The upper panel in Table 2 reports the regression results for daughters and lower panel for sons.

Column (1) in Table 2 reports the coefficients of $N_f$ (father in non-farm) and $N_m$ (mother in non-farm) in the regression for son’s and daughter’s participation in nonfarm sector. The results from the probit regression without any controls show that mother’s non-farm participation has a significant positive influence on daughter’s probability of participation in the same sector. The marginal effect of a mother’s participation in nonfarm sector ($N^m$) is estimated to be 0.43 which is large compared to the daughter’s average probability of participation in non-agriculture of 0.16. In contrast, father’s participation in nonfarm ($N^f$) appears to have no statistically significant effect on daughter’s likelihood of being employed in the same sector. The results for sons reported in the lower panel of Table 2 indicate significant positive correlation between father and son’s employment in the nonfarm sector. The marginal effect of father’s employment in nonfarm sector ($N^f$) is around 0.15 which is statistically significant at 1 percent significance level. Compared with father, mother’s nonfarm participation ($N^m$) has a smaller marginal effect (0.10) which is significant at 5 percent level.

The next set of results reported in column (2) of Table 2 includes a large number of household and individual level characteristics as control variables. The access to non-farm jobs may depend on the personal networks that often run along ethnic group/caste (see, for example, Dreze, Lanjouw and Sharma, 1998). To capture the variations in access to non-farm jobs, we include a set of dummies depicting the ethnicity (caste and tribe) of the individual in the regression. We also include dummies showing if there is any short/long-
term migrant in the household, as migration frequently occurs on the basis of personal networks. A set of household variables including household size and composition are also added to the set of control variables. As discussed in the conceptual framework, human and financial capital variables are important links in the intergenerational transmissions of socioeconomic status. In addition to the level of education, we include the age of an individual as a human capital variable representing the work experience. The education levels of parents and spouse are also included as additional human capital controls. The inherited land (as the most important form of collateral), remittances received, and travel time to the nearest commercial bank are included as controls for access to capital. We include an individual’s marital status to account for taste and/or life-cycle related heterogeneity. The summary statistics for these explanatory variables are presented in appendix Table A.1. A comparison of daughters’ and sons’ samples in Table A.1 shows that the difference in the means of individual and family characteristics across parental occupation is much more pronounced in case of sons sample. This indicates that selection might be relatively more important for the sons’ sample.

The results reported in column (2) in Table (2) show that the added set of regressors are powerful determinants of nonfarm participation decision. With the inclusion of these controls, the Pseudo $R^2$ increases from 0.10 to 0.52 in daughter’s sample and from 0.03 to 0.23 in son’s sample. However, the addition of the powerful set of controls does not affect the estimated intergenerational partial correlations in any significant way. The marginal effect of $N^m$ (mother in non-farm) is estimated to be 0.41 which is virtually identical to the estimate from the regression with no controls (0.43). It is still highly statistically significant ($t = 8.04$). In the sons sample, the marginal effect of father’s employment in nonfarm sector ($N^f$) is estimated to be 0.16 which is also nearly identical to our earlier estimate from regression with no controls (0.15).

---

24 In addition, age and its squared term capture any cohort effect.

25 The ethnicity dummy may also capture access to credit.

26 This implies that in our case the girls’ sample is more like the preferred C8 (catholic school 8th graders) sample in AET (2005).
Although the regressions in column (2) include a large set of individual and household level controls, the intergenerational correlation in occupation may still result, spuriously, from the fact that parents and children may face similar labor market opportunities. For instance, if both parents and children live in an area with better non-farm opportunities (say location of a textile mill), then intergenerational correlation in non-farm participation may be an artefact of not adequately controlling for non-farm opportunities in the regression. To control for unobserved location specific heterogeneity in non-farm opportunities, we included village level fixed effects in the estimation (151 and 241 village dummies in daughter’s and son’s regressions respectively). The village fixed effects may also capture other village specific determinants of occupational choice like peer effects and agglomeration forces. In addition, we define the share of non-farm employment in total employment of an individual’s age cohort in her district of birth as an additional control for labor market opportunity. This may capture the time varying part of labor market opportunities in a village.

The results from regressions with village fixed effects are reported in column (3) of Table 2. The addition of village level fixed effects as well as a measure of intertemporal labor market opportunity leads to an increase in the explanatory power of the regressions further. The Pseudo $R^2$ of the regression is 0.62 in daughter’s sample and 0.53 in son’s sample. Despite the inclusion of such a large number of controls (village plus household plus individual level controls), the qualitative results regarding intergenerational occupational correlations remain largely unchanged. Although the marginal effect of $N^m$ (mother in non-farm) on a daughter’s nonfarm participation declines, it is still large (0.30) and is statistically highly significant with a t-value=6.33. The marginal effect of $N^f$ (father in non-farm) on a son’s nonfarm participation is now 0.10 and is statistically significant at 1 percent level. Consistent with the available evidence in the literature on income mobility (see, for example, Solon, 2002), the evidence also indicates that the cross gender effects are not important as they are not statistically significant in column (3) of Table 2.
(5) Genetic Transmissions and Intergenerational Correlations

The results discussed so far show that the intergenerational occupational correlations between parents and children run along gender lines (father-son and mother-daughter). The evidence indicates that the estimated partial correlations are not solely due to the ‘tangible’ determinants of occupational choice like education and assets, as they are already controlled for in the regression. The results can not be driven by village level factors like peer effects and geographic agglomeration as we include fixed effect. However, as discussed before, an important question from the policy perspective is how much of the partial correlations uncovered in column 3 of table 2 is causal due to environmental factors like role model effects and learning externalities as opposed to pure genetic correlations in occupational choice.

In the absence of credible identifying instruments, we utilize a number of ways to ascertain whether the observed intergenerational correlations can be explained solely in terms of unobserved ability correlations (and other unobserved common determinants). They include (i) sensitivity analysis a la Rosenbaum and Rubin (1983), Rosenbaum(1995) and AET (2005, 2000); (ii) estimation of lower bounds on intergenerational correlations using the technique developed by AET (2005, 2000).

(4.1) Sensitivity Analysis

The regression results presented in the previous section (Table 2) demonstrate that the inclusion of a large and powerful set of controls does not lead to a substantial weakening of intergenerational occupational correlations especially for the daughter. This suggests that the selection on observables is dominant and a relatively small amount of selection is due to unobservables. We now explore the question whether a small amount of selection on unobservables can explain away the estimated partial correlations in intergenerational occupational choices in table 2. Consider the following bivariate probit model for individual i.
\[ N_i = 1(\alpha_p N_i^p + X_i \gamma_1 + \delta_j \omega_j + \xi > 0), \]  
\[ N_i^p = 1(X_i \beta_1 + \delta_j \omega_j + u > 0) \]  
\[
\begin{bmatrix}
    u \\
    \xi
\end{bmatrix}
\sim
\begin{bmatrix}
    0 \\
    0
\end{bmatrix},
\begin{bmatrix}
    1 & \rho \\
    \rho & 1
\end{bmatrix}
\]  

where \( N_i \) (also \( N_i^p \)) is a binary occupation choice variable which takes the value 1 for non-farm and zero otherwise, \( \omega_j \) is the village dummy (fixed effect) included to control for unobserved and observed community level determinants including labor market opportunities and peer effects. We estimate the magnitudes of intergenerational correlations for different values of the correlation (\( \rho \)) between the unobserved determinants of nonfarm participation of parents \( (u) \) and children \( (\xi) \).\(^{27}\) The vector of explanatory variables \( (X) \) is the same as that in the regression results presented in column (3) of Table 2. However, the inclusion of village level fixed effect in the regression describing parental participation in nonfarm sector \( (N^p) \) causes problem in estimation as there are cases of perfect fit due to the absence of parental occupational diversification in a village (all 0s or 1s for the occupation dummy). When we exclude such cases, the sample size reduces to 1126 in daughter’s sample and 2547 in son’s sample. The results for these restricted samples are reported in columns 2 and 4 of Table 3 for daughters and sons respectively. An alternative approach that keeps the sample size same as in Table 2 relies on an index of village fixed effects estimated from the simple probit regressions reported in column 3 of Table 2. The results for these unrestricted samples are presented in columns 1 and 3 of Table 3 for daughter’s and sons respectively. Following AET (2005), the sensitivity analysis is performed for \( \rho = 0.1, 0.2, 0.3, 0.4, 0.5 \). Note that the correlation coefficient \( \rho \) represents

\(^{27}\)As discussed in AET (2005, 2000) the bivariate probit model above is identified because of nonlinearity. However, such identification based on functional form alone in the absence of valid instruments is treated with skepticism in applied literature (termed “weak identification”). In what follows, the bivariate probit model is treated as underidentified and thus the sensitivity analysis is performed across alternative values of \( \rho \).
only that part of genetic correlation across generations which influence the occupational choice.

For daughters, the results from the unrestricted sample show that the marginal effect of the mother’s employment in nonfarm sector declines to 0.22 when $\rho = 0.10$, and to 0.15 when $\rho = 0.20$. The estimated marginal effect continues to decline with an increase in $\rho$ but is still positive though small in magnitude when $\rho$ is as high as 0.50. Interestingly, all the values of marginal effect are also statistically significant at 5 percent or less except for the case when $\rho = 0.50$. The conclusions derived from the restricted sample reported in column 2 are similar to that of the unrestricted sample; the marginal effects are, however, in general, larger in magnitude. These results suggest that barring sampling error, the unobserved genetic correlations pertinent to occupation choice would have to be greater than 0.50 to explain away the entire effect of $N^m$ (mother in nonfarm) on a daughter’s nonfarm participation.

In the case of sons, the marginal effect of father’s nonfarm participation becomes numerically small (0.05) and statistically insignificant in the unrestricted sample when $\rho = 0.10$. For values of $\rho$ equal to or greater than 0.2, the marginal effect becomes negative. The results are again very similar in the case of the restricted sample. These results suggest that the estimated effect of $N^f$ (father in nonfarm) on son’s nonfarm participation may be entirely driven by common unobserved factors like genetic transmissions.

(4.2) Lower Bounds on Intergenerational Occupational Linkages

The sensitivity analysis above indicates that the value of $\rho$ would have to be larger than 0.50 to completely explain the effect of mother’s nonfarm employment on that of daughters found in Table 2. But there is no estimate of $\rho$ in the literature which we can use as a benchmark. The available evidence from Behavioral Genetics shows that both the genetic transmissions and environmental factors are important in the correlation between the parents and children, especially for complex traits and behavior.\textsuperscript{28} The problems in

\textsuperscript{28}For example, the correlation between IQ scores of parents and children is around 0.5 (Plomin et. al.,
pinning down a plausible range for $\rho$ are more daunting in our case as other explanatory variables like education (both children’s and parents’), ethnicity (i.e., caste and tribe), and assets are likely to pick up a substantial part of this correlation.\footnote{This implies that the value of $\rho$ relevant for our analysis should be smaller than otherwise.} In the absence of any plausible way of judging the magnitude of the genetic correlations relevant for occupation choices in a rural economy as captured by $\rho$, we utilize an approach suggested by AET (2005). This allows us to estimate both the magnitude of $\rho$ and bounds for the intergenerational correlations.

To illustrate the basic insights behind AET (2005) approach, we consider equation (3) (with village fixed effects added). It defines the latent variable $N_i^*$ that determines children’s participation in nonfarm sector as:

$$N_i^* = \alpha_p N_i^p + X_i' \gamma_1 + \delta_j \omega_j + \epsilon$$  \hspace{1cm} (7)

where $N_i^p$ is the dummy variable for nonfarm participation by parents and $\omega_j$ is the village fixed effect for village $j$ where individual $i$ lives in. Let $N_i^{ps}$ is the latent variable such that $N_i^{ps} = 1$ if $N_i^{ps} > 0$ and zero otherwise. We can define the linear projection of $N_i^{ps}$ on $X_i' \gamma, \omega$ and $\epsilon$ as (for notational simplicity the subscript is dropped):

$$\text{Proj} \left( N_i^{ps} | X_i' \gamma_1, \omega, \epsilon \right) = \phi_0 + \phi_{X_i' \gamma_1} X_i' \gamma_1 + \omega \delta + \phi_{\epsilon} \epsilon$$  \hspace{1cm} (8)

Following AET (2005), we can interpret $\phi_{X_i' \gamma_1}$ as the “selection on observables” and $\phi_{\epsilon}$ the “selection on unobservable”. However, unlike AET (2005), we use a village level fixed effect to sweep off the observed and unobserved village level determinants. This implies that the selection on observables ($\phi_{X_i' \gamma_1}$) and unobservables ($\phi_{\epsilon}$) both represent only the individual characteristics. An advantage of this formulation is that it fits well with the notion that the ‘unobservables’ are like ‘observables’. An alternative approach is...
to include the village fixed effects as part of the observables. The argument is that the location of an individual is an observable characteristic.\footnote{We thank Chris Taber for pointing out the alternative interpretations of the fixed effects.} The linear projection of $N^p$ in this case becomes:

$$\text{Proj} (N^p \mid Z' \gamma_2, \varepsilon) = \phi_0 + \phi_{Z' \gamma_2} Z' \gamma_2 + \phi_\varepsilon \varepsilon; \quad Z = (X, \omega) \text{ and } \gamma_2 = (\gamma_1, \delta)$$

(9)

The advantage of this formulation is that it is more likely to satisfy the condition that selection on observables is dominant which helps in deriving the lower bound on intergenerational occupational linkage (see below).\footnote{Since location choice is endogeneous, the village fixed effects will capture some of the unobserved individual characteristics which are common to the villagers.} We perform the analysis under these alternative interpretations (equations (8) and (9)).\footnote{A third alternative is to exclude the fixed effects altogether and use village level observed controls (share of non-farm). The conclusions of this paper remain unchanged in this formulation, although the lower bound estimates are larger than reported here.} Note that in the case of univariate probit regressions, the maintained assumption is that there is no selection on unobservable, i.e., $\phi_\varepsilon = 0$.

AET (2005, 2000) and Altonji, Conley, Elder, and Taber (2005) (henceforth ACET (2005)) show that selection on observables can be used as a guide to selection on unobservables. They point out that in many applied economic applications, it is a natural assumption that the selection on observables dominates the selection on unobservables which leads to the following conditions in our case (analogous to condition (3) in AET (2005)):

$$\phi_{X' \gamma_1} \geq \phi_\varepsilon \geq 0$$

(10)

$$\phi_{Z' \gamma_2} \geq \phi_\varepsilon \geq 0$$

(11)

Following AET (2005), we can implement the econometric estimation under the above restriction(s) and treat the estimate of $\alpha_p$ (equation (7)) corresponding to the case of
equality of selection on observables and unobservables (i.e., for example, $\phi_{X'\gamma_1} = \phi_{\varepsilon}$) as the lower bound on the part of intergenerational occupational linkage that is not driven by genetic transmissions. The inequality conditions (10) and (11) above are eminently plausible in our case due to the following considerations.\textsuperscript{33} First, as pointed out earlier, the addition of a set of rich and powerful determinants of occupation choice affects the strength of intergenerational linkages only marginally although the Pseudo $R^2$ goes up dramatically. For example, in daughter’s sample, the Pseudo $R^2$ increases from 0.10 to 0.52 when we include a rich set of determinants of occupational choice including education levels of children, parents and spouse, inherited land, and ethnicity. The estimated partial correlation in non-farm participation by mother and daughter is, however, barely affected (it declines from 0.43 to 0.41). This indicates that (i) the observables explain a large part of the variations in non-farm participation, and thus leave room for only a limited role for the unobserved individual characteristics; (ii) the estimated partial correlation is robust to possible inclusion of additional controls (if such data were available). Second, the data for our analysis come from a multipurpose household survey which was conducted primarily for poverty assessment. Since the role of non-farm occupations as an avenue to escape poverty traps in a low income agrarian economy is much discussed (Lanjouw and Feder, 2001), it is only natural that the survey includes rich information on the determinants of non-farm participation identified in the recent literature. This means that these observable characteristics are likely to pick up a substantial part of the unobserved genetic correlations relevant for occupation choice, a point mentioned earlier, but worth emphasizing again here. This also means that the selection on unobservable genetic endowment captured in $\phi_{\varepsilon}$ will be much smaller in our analysis. Third, we can decompose the error term in the occupation choice by children as in equation (4): $\xi = \xi_1 + \xi_2$ where $\xi_1$ is the part of selection on unobservables that is common to both generations but is determined at the time of parental occupation choice, and $\xi_2$ represents the unobserved shocks that occur during

\textsuperscript{33}We are grateful to Chris Taber and Todd Elder for clarifying the relevance of the conditions (10) and (11) in our analysis.
the children’s occupation choice. As shown by AET (2005), this implies that selection on observables is greater providing additional justifications for inequality conditions (10) and (11) above.\footnote{We note that three assumptions need to be satisfied for the point identification within the AET approach. Two of the assumptions ensure that the selection on observables is equal to selection on unobservables (AET (2000), P.4). They are: (i) the observables are a random subset of a large set of potential determinants, (ii) the number of explanatory variables is large, none of the elements dominating the distribution of \( N_p \) or \( N_i \). The detailed regression results presented in appendix Table A.2 show that the second assumption is approximately satisfied in our context. As discussed above in the text, in our case, the assumption of random selection of observables does not hold, as the survey includes rich information on the determinants of non-farm participation given its focus on poverty assessment. Thus the selection on observables is likely to be larger than the selection on unobservables. This implies that when estimation is done under the restriction of equality of selection on observables and unobservables, the estimated causal effect of parental occupation will underestimate the true magnitude of the intergenerational linkage and thus can be treated as the lower bound. The third assumption below when coupled with the equality of selection on observables and unobservables can provide point identification. The third assumption essentially says that the observables and unobservables are nearly uncorrelated. This assumption (a stricter version of it) is invoked in most of the empirical applications in economics with its focus on the endogeneity of a single variable (like the extensive literature on returns to education) and thus assumes that all other controls are not correlated with the error term. This, however, is not a natural assumption and may be difficult to justify in many applications. As emphasized by AET (2000), even if this assumption is not satisfied exactly, the Monte Carlo evidence indicates that the resulting bias is small. Also, AET (2005) and ACET (2005) note that for actual implementation of the approach what is needed is that the relevant assumptions are satisfied approximately.}

In the case of bivariate probit (equations 4-6), the lower bound estimate of \( \alpha_r \) can be estimated by imposing the following conditions depending on the treatment of fixed effect:

\[
\rho = \frac{\text{Cov}(Z'\beta_2, Z'\gamma_2)}{\text{Var}(Z'\gamma_2)} ; \beta_2 = (\beta_1, \delta) \tag{12}
\]

\[
\rho = \frac{\text{Cov}(X'\beta_1, X'\gamma_1)}{\text{Var}(X'\gamma_1)} \tag{13}
\]

Table 4 reports the estimates of the lower bounds on the intergenerational partial correlation (i.e., lower bound estimates of \( \alpha_m \) and \( \alpha_f \)). The inclusion of village dummies leads to computational difficulties and convergence problems in the estimation of the bivariate probit model. The results discussed earlier in Table 3 show that the index of village fixed effects estimated from univariate probit model performs equally well as village level dummies in controlling for spatial labor market opportunities. To avoid the non-convergence
problems, we use this index in the regressions reported in Table 4. The first panel reports the results from bivariate probit model under the constraint defined in equation (12) and the third panel shows the corresponding results under equation (13). The central conclusions of this paper are, however, not sensitive to the treatment of fixed effects and we focus our discussion on the case defined by equation (12) for the sake brevity. The estimated magnitudes of correlations between unobserved determinants of parent and children’s nonfarm participation are similar: 0.21 for daughters and 0.25 for sons. The estimates show that the intergenerational correlation between mother’s and daughter’s nonfarm participation is highly statistically significant (t-value=4.81). The estimated coefficient $\alpha_m$ is positive and large in magnitude (0.685) with a marginal effect of 0.146. In contrast, for sons, the estimated $\alpha_f = -0.143$ with a t-value of 1.74. The results in panel 1 of Table 4 thus strengthen our central conclusion from the sensitivity analysis in section (4.1) above that the estimated partial correlation in the nonfarm participation of mother and daughter is not likely to be driven entirely by the genetic correlations; at least part of the occupational linkage seems causal, probably reflecting cultural inheritance through role model effects, learning externalities and transfer of reputation and social capital from mother to the daughter as discussed in the conceptual framework above. In contrast, the lower bound estimate for the correlation in father’s and son’s nonfarm participation is negative implying that the observed (positive) intergenerational correlation may be an artefact of genetic transmissions across generations.\footnote{We caution here that the fact that the lower bound estimate is negative for the father and son should not be taken as conclusive evidence for an absence of intergenerational linkage. As mentioned before the lower bound estimates are likely to underestimate the strength of occupational linkage given that the selection on observables is likely to dominate. This also implies that the evidence from the bounds estimates in favor of a causal effect of mother’s non-farm participation is very strong.}

To ensure robustness of our findings, we also check whether these results are driven by the joint normality assumption underlying the bivariate probit model. Following AET
(2005), we utilize the following semi-parametric specification for the error terms:

\[ u = \theta + u^* \]
\[ \xi = \theta + \xi^* \]

Where \( u^* \) and \( \xi^* \) are independent standard normals and \( \theta \) is unrestricted. Bivariate probit model estimated earlier is thus a special case where \( \theta \) is assumed to be distributed as normal. We estimated the model using nonparametric maximum likelihood method suggested by Heckman and Singer (1984) and AET (2005). The estimation method treats the distribution of \( \theta \) as discrete; in practice, we obtain two points of support for \( \theta \). The estimated \( \rho \) and \( \alpha_p \) are reported in the second and fourth panels of Table 4. Again, for the sake of brevity, we focus on the case when the village fixed effects are treated as part of the observables index (second panel in Table 4). The estimated magnitudes of \( \rho \) are smaller compared with those from the bivariate probit model, but as before they are similar for sons (0.178) and daughters (0.163). However, the overall results regarding the intergenerational effects remain unchanged. The effect of \( N^m \) (mother in nonfarm) on daughter’s nonfarm participation is statistically highly significant, and positive \( (\alpha_m = 0.665) \). The implied marginal effect is 0.135 which is virtually identical to that found in the bivariate probit model (0.146). For sons, the estimated lower bound on intergenerational correlation is positive but much smaller in magnitude \( (\text{marginal effect}=0.086) \) and statistically insignificant.

Sources of Gender Differences

The main finding from our empirical analysis is that at least part of the partial correlation in non-farm participation by mothers and daughters seems to survive when we take into account the unobserved common factors across generations, but may not be so in case of fathers and sons. Understanding the sources of this gender differences in the intergenerational occupation linkages is clearly important. But a proper analysis of the
issue is beyond the scope of this paper. We, however, note a few plausible hypotheses to be explored rigorously in future work. First, there is significant gender inequality in educational attainment in Nepal in favor of sons (more than double). Education is an important key to occupational mobility out of agriculture. The dramatically higher educational attainment of sons thus weakens the intergenerational link by enhancing their human capital suitable for non-agricultural occupation. Education also broadens the set of role models that include the teachers among others. Second, women in a traditional society such as Nepal are brought up within the confines of a household with limited exposure to the outside world. Naturally, they have their mother as the primary role model with a disproportionately large influence on a daughter’s occupation choice. For sons, the set of role models may extend well beyond the household; while the father may still exert some influence, his impact on a son’s occupation choice becomes diluted. Third, as emphasized in recent literature, occupational mobility may be linked to geographic mobility (see, for example, Long and Ferrie (2007). Although geographic mobility is, in general, very restricted in Nepal, there is clear gender effect in migration; the sons are, in general, more likely to choose their geographic location as a strategy to move out of agriculture. The geographic location of daughters is, in contrast, determined by the parental location (for unmarried) or by the location of the husband (for married).  

(5) Conclusions:

The economic literature on intergenerational mobility has witnessed a renewed interest in recent years. However, most of the existing economic research focuses on the income correlations between father and son(s) in the context of developed countries. Using data from a developing country, Nepal, we present evidence on the intergenerational occupational mobility from agriculture to nonfarm sector with an emphasis on the gender differences. It is extremely difficult, if not impossible, to find credible instrument(s) to address the

\footnote{We thank Bhashkar Mazumder for pointing out the possible role of geographic mobility. We, however, note that the role geographic location can play in our results is limited by the fact that we use village level fixed effects. Since the village fixed effects are employed in separate daughters’ and sons’ regressions they may not account for gender differences in locations completely.}
genetic correlations (ability and preference) or other common unobserved determinants of occupation choice. We employ the recent econometric approach developed by Altonji, Elder and Taber (2005, 2000) to ascertain if the observed partial correlation in non-farm participation can be attributed solely to genetic transmissions or at least part of the effect is likely to be causal. The approach uses the degree of selection on observables as a guide to the degree of selection on unobservables. It allows us to estimate lower bounds on the part of the intergenerational occupational correlations that can be attributed to intergenerational ‘cultural inheritance’ possibly due to factors like role model effects, learning externalities, and transfer of reputation and social capital, among other things. The results show that the observed partial correlation between the father and a son can be easily explained away by a moderate correlation in genetic endowments across generations. In contrast, for the mother and daughter(s), the intergenerational occupational linkage is very strong, and it is unlikely that the estimated partial correlation is driven solely by the unobserved genetic correlations. The evidence points to a causal effect of mother’s occupation choice on that of the daughter. The estimated lower bound for mother-daughter intergenerational occupational correlation shows a marginal effect of 0.14. The evidence that, beyond the genetic transmissions, there is persistence in nonfarm participation between the mother and a daughter has important implications. It provides a link in the analysis of poverty trap in rural households and brings into focus the intergenerational occupational linkage as an important factor in understanding the restricted economic mobility of women in developing countries.

References


March, 2005.


(38) Rosenbaum, P.R., 1995, Observational Studies, New York: Springer-Verlag.

(39) Rosenbaum, P.R., and D. B. Rubin, 1983, “Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome,” Journal of Royal Sta-


Table 1: Nonfarm participation of children conditional on Parent’s Employment Status (weighted mean)

<table>
<thead>
<tr>
<th>Probability of being Employed in Nonfarm Activities</th>
<th>Daughters</th>
<th>Sons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother's employment in Farm</td>
<td>0.13</td>
<td>0.43</td>
</tr>
<tr>
<td>Nonfarm</td>
<td>0.54</td>
<td>0.63</td>
</tr>
<tr>
<td>Difference</td>
<td>0.414***</td>
<td>0.20***</td>
</tr>
<tr>
<td>Father's employment in Farm</td>
<td>0.14</td>
<td>0.41</td>
</tr>
<tr>
<td>Nonfarm</td>
<td>0.24</td>
<td>0.57</td>
</tr>
<tr>
<td>Difference</td>
<td>0.10***</td>
<td>0.16***</td>
</tr>
<tr>
<td>Unconditional</td>
<td>0.16</td>
<td>0.44</td>
</tr>
</tbody>
</table>

* significant at 1%; ** significant at 5%; *** significant at 1%
### Table 2: Probit Estimates of Intergenerational Correlations

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Father in Non-Farm (n_f)</strong></td>
<td>-0.072</td>
<td>-0.062</td>
<td>-0.068</td>
</tr>
<tr>
<td></td>
<td>[0.016]</td>
<td>[0.012]</td>
<td>[0.009]</td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td>(0.49)</td>
<td>(0.46)</td>
</tr>
<tr>
<td><strong>Mother in Non-farm (n_m)</strong></td>
<td>1.286</td>
<td>1.283</td>
<td>1.157</td>
</tr>
<tr>
<td></td>
<td>[0.433]</td>
<td>[0.406]</td>
<td>[0.299]</td>
</tr>
<tr>
<td></td>
<td>(8.21)**</td>
<td>(8.04)**</td>
<td>(6.33)**</td>
</tr>
<tr>
<td><strong>Pseudo-R²</strong></td>
<td>0.1</td>
<td>0.522</td>
<td>0.622</td>
</tr>
</tbody>
</table>

### Daughter's sample

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Father in Non-Farm (n_f)</strong></td>
<td>0.372</td>
<td>0.402</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>[0.146]</td>
<td>[0.157]</td>
<td>[0.099]</td>
</tr>
<tr>
<td></td>
<td>(4.92)**</td>
<td>(4.85)**</td>
<td>(2.79)**</td>
</tr>
<tr>
<td><strong>Mother in Non-farm (n_m)</strong></td>
<td>0.258</td>
<td>0.252</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>[0.102]</td>
<td>[0.099]</td>
<td>[0.078]</td>
</tr>
<tr>
<td></td>
<td>(2.16)*</td>
<td>(2.00)*</td>
<td>(1.38)</td>
</tr>
<tr>
<td><strong>Pseudo-R²a</strong></td>
<td>0.035</td>
<td>0.227</td>
<td>0.527</td>
</tr>
</tbody>
</table>

### Son's Sample

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual and household characteristics b</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Village fixed effect</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Note.** - Entries are probit coefficients. Standard errors are corrected for intra-cluster correlations due to clustered sampling. t-values are in parentheses and marginal effect of each variable (evaluated at sample means) is shown in bracket.

a. Pseudo $R^2$ is defined as $\frac{\text{Var}(X'\gamma)}{1+\text{Var}(X'\gamma)}$

b. Regressors in column 2 include level of education, age, age squared, dummy for married, household size & composition, inherited land, distance to bank, un-earned income, dummy for migrant member in the household, 3 ethnicity dummies, father, mother and spouse's education level. Regressors in column (3), in addition to above regressors, include share of nonfarm employment by age cohort, and an index of village fixed effect.

* significant at 5%; ** significant at 1%
<table>
<thead>
<tr>
<th>Correlation of Disturbances</th>
<th>Daughter's Sample</th>
<th>Son's Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unrestricted Sample</td>
<td>Restricted Sample</td>
</tr>
<tr>
<td>ρ = 0</td>
<td>1.116 [0.284] (7.68)**</td>
<td>1.260 [0.353] (8.02)**</td>
</tr>
<tr>
<td>ρ = 0.1</td>
<td>0.917 [0.216] (6.35)**</td>
<td>1.075 [0.288] (6.88)**</td>
</tr>
<tr>
<td>ρ = 0.2</td>
<td>0.715 [0.154] (5.00)**</td>
<td>0.886 [0.225] (5.73)**</td>
</tr>
<tr>
<td>ρ = 0.3</td>
<td>0.509 [0.100] (3.63)**</td>
<td>0.692 [0.165] (4.56)**</td>
</tr>
<tr>
<td>ρ = 0.4</td>
<td>0.299 [0.053] (2.19)*</td>
<td>0.492 [0.11] (3.33)**</td>
</tr>
<tr>
<td>ρ = 0.5</td>
<td>0.085 [0.013]</td>
<td>0.286 [0.06]</td>
</tr>
</tbody>
</table>

N | 2037 | 1126 | 2919 | 2547 |

Note.- Entries are probit coefficients. Standard errors are corrected for intra-cluster correlations due to clustered sampling. t-values are in parentheses and marginal effect of each variable (evaluated at sample means) is shown in bracket.

a. Regressors include level of education, age, age squared, dummy for married, household size & composition, inherited land, distance to bank, un-earned income, dummy for migrant member in the household, 3 ethnicity dummies, father, mother and spouse's education level, share of nonfarm employment by age cohort, and an index of village fixed effect.

b. Regressors include level of education, age, age squared, dummy for married, household size & composition, inherited land, distance to bank, un-earned income, dummy for migrant member in the household, 3 ethnicity dummies, father, mother and spouse's education level, share of nonfarm employment by age cohort, and an index of village level dummies.

* significant at 5%; ** significant at 1%
**Table 4: Estimates of Lower Bounds for the Intergenerational Correlations**

<table>
<thead>
<tr>
<th></th>
<th>Daughter's Sample</th>
<th>Son's Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
<td>$\alpha_a$</td>
</tr>
<tr>
<td><strong>Bivariate Probit Estimation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = \text{Cov}(Z_{Y_2}, Z_{\beta Y_2})/\text{Var}(Z_{Y_2})$</td>
<td>0.214 (1.70)</td>
<td>0.685 [0.146]</td>
</tr>
<tr>
<td></td>
<td>(4.81)**</td>
<td></td>
</tr>
<tr>
<td><strong>Nonparametric Maximum Likelihood Estimation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = \text{Cov}(Z_{Y_2}, Z_{\beta Y_2})/\text{Var}(Z_{Y_2})$</td>
<td>0.163 (0.24)</td>
<td>0.665 [0.135]</td>
</tr>
<tr>
<td></td>
<td>(3.50)**</td>
<td></td>
</tr>
<tr>
<td><strong>Bivariate Probit Estimation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = \text{Cov}(X_{Y_1}, X_{\beta Y_1})/\text{Var}(X_{Y_1})$</td>
<td>0.219 (1.43)</td>
<td>0.677 [0.144]</td>
</tr>
<tr>
<td></td>
<td>(4.75)**</td>
<td></td>
</tr>
<tr>
<td><strong>Nonparametric Maximum Likelihood Estimation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho = \text{Cov}(X_{Y_1}, X_{\beta Y_1})/\text{Var}(X_{Y_1})$</td>
<td>0.177 (0.20)</td>
<td>0.665 [0.135]</td>
</tr>
<tr>
<td></td>
<td>(3.50)**</td>
<td></td>
</tr>
</tbody>
</table>

Note: Entries are probit coefficients. Standard errors are corrected for intra-cluster correlations due to clustered sampling. t-values are in parentheses and marginal effect of each variable (evaluated at sample means) is shown in bracket.
Regressors include level of education, age, age squared, dummy for married, household size & composition, inherited land, distance to bank, un-earned income, dummy for migrant member in the household, 3 ethnicity dummies, father, mother and spouse's education level, share of nonfarm employment by age cohort, and an index of village fixed effect.
* significant at 5%; ** significant at 1%
### Table A.1: Summary Statistics by parental employment status.

<table>
<thead>
<tr>
<th></th>
<th>Daughter's Sample</th>
<th></th>
<th>Son's Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Farm (N=1880)</td>
<td>Non-farm (N=157)</td>
<td>Difference</td>
<td>Farm (N=2309)</td>
</tr>
<tr>
<td>Participation in Nonfarm employment (proportion)</td>
<td>0.13</td>
<td>0.54</td>
<td>0.414***</td>
<td>0.39</td>
</tr>
<tr>
<td>Level of Education (Years)</td>
<td>1.63</td>
<td>2.66</td>
<td>1.03**</td>
<td>3.76</td>
</tr>
<tr>
<td>Father's level of education (years)</td>
<td>1.15</td>
<td>2.01</td>
<td>0.96*</td>
<td>0.81</td>
</tr>
<tr>
<td>Mother's level of education (years)</td>
<td>0.1</td>
<td>0.3</td>
<td>0.2</td>
<td>0.05</td>
</tr>
<tr>
<td>Spouse' level of education (years)</td>
<td>2.07</td>
<td>2.05</td>
<td>-0.03</td>
<td>0.334</td>
</tr>
<tr>
<td>Age</td>
<td>32.72</td>
<td>30.67</td>
<td>-2.05</td>
<td>37.13</td>
</tr>
<tr>
<td>Age squared</td>
<td>1230</td>
<td>1153</td>
<td>-77</td>
<td>1591</td>
</tr>
<tr>
<td>Married</td>
<td>0.83</td>
<td>0.63</td>
<td>-0.19***</td>
<td>0.8</td>
</tr>
<tr>
<td>Household size(log)</td>
<td>2.01</td>
<td>2</td>
<td>-0.001</td>
<td>1.77</td>
</tr>
<tr>
<td>Share of adult female</td>
<td>0.25</td>
<td>0.26</td>
<td>0.012</td>
<td>0.23</td>
</tr>
<tr>
<td>Share of children</td>
<td>0.17</td>
<td>0.17</td>
<td>0.0001</td>
<td>0.15</td>
</tr>
<tr>
<td>Share of Young</td>
<td>0.35</td>
<td>0.35</td>
<td>0.006</td>
<td>0.34</td>
</tr>
<tr>
<td>Share of Old</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Inherited land (value in million Rs.) (log)</td>
<td>9.26</td>
<td>6.65</td>
<td>-2.61***</td>
<td>8.39</td>
</tr>
<tr>
<td>Travel time to nearest bank</td>
<td>2.65</td>
<td>1.78</td>
<td>-0.87***</td>
<td>2.91</td>
</tr>
<tr>
<td>Un-earned income (million Rs)</td>
<td>0.008</td>
<td>0.007</td>
<td>-0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>Migrant in the household</td>
<td>0.39</td>
<td>0.48</td>
<td>0.09</td>
<td>0.32</td>
</tr>
<tr>
<td>Upper caste Hindu (Proportion)</td>
<td>0.36</td>
<td>0.21</td>
<td>-0.15**</td>
<td>0.36</td>
</tr>
<tr>
<td>Lowr caste Hindu (Proportion)</td>
<td>0.07</td>
<td>0.19</td>
<td>0.12*</td>
<td>0.07</td>
</tr>
<tr>
<td>Tribal (Proportion)</td>
<td>0.29</td>
<td>0.23</td>
<td>-0.06</td>
<td>0.26</td>
</tr>
<tr>
<td>Share of nonfarm in district by age cohort</td>
<td>0.14</td>
<td>0.2</td>
<td>0.06***</td>
<td>0.28</td>
</tr>
</tbody>
</table>

* significant at 1%; ** significant at 5%; *** significant at 1%
## Appendix Table A.2: Probit Estimation of Intergenerational correlations

| Employment Status               | Mother in Non-farm Employment | Father in Non-farm Employment | Level of education (year) | Age | Age Squared | Married (Yes=1) | Household Size (log) | Share of Adult Female | Share of Children | Share of Youth | Share of Old | Inherited Land (log) | Father's education (year) | Mother's Education (year) | Spouse's education(year) | Travel time to Bank | Unearned income | Migrant in the household (yes=1) | Upper Caste Hindu (yes=1) | Lower Caste Hindu (yes=1) | Belongs to tribe (yes=1) | Constant | Observations |
|--------------------------------|-------------------------------|-------------------------------|----------------------------|-----|-------------|----------------|-------------------|----------------------|-------------------|----------------|-------------|----------------|------------------------|-------------------------|---------------------------|------------------------|---------------------|----------------|-----------------------|------------------------|-----------------------|-----------------------|----------|--------------|
| Daughter in Non-Farm Sector    | 1.283                         | -0.062                       | 0.070                      | 0.055 | -0.001      | -0.075         | -0.293            | -0.150               | 0.337             | 0.452         | 0.118       | -0.048                  | -0.026                  | 0.034                     | 0.031                  | 0.007               | -12.758            | -0.005                 | 0.089                   | 0.100                  | 0.290                 | 0.290       | 2037         |
| Son in Non-Farm Sector         | 0.252                         | 0.402                        | 0.028                      | 0.107 | -0.001      | 0.375          | -0.291            | -0.385               | 0.385             | -0.193        | -0.423      | -0.008                  | 0.003                   | 0.098                     | 0.017                  | 0.008               | 0.038              | 0.504                  | -0.322                | 0.227                  | -0.170               | -1.887       | 2919         |

Note.- Entries are probit coefficients. Standard errors are corrected for intra-cluster correlations due to clustered sampling. t-values are in parentheses.

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