Child Schooling in India: An Empirical Analysis with Cluster Based Sample.

Itismita Mohanty

Department of Economics
University of Sydney
NSW 2006, Australia
Phone: +61 2 9036 9196
Fax: +61 2 9351 4341
e-mail:I.Mohanty@econ.usyd.edu.au

Abstract

Realising the significance of education in economic development and growth, the persisting problem of low child schooling in developing countries over decades, has been an area of policy concern. In this process, India is experiencing a combination of problems, with the access to primary schooling and also with the retention and achievement at the primary, secondary and higher secondary levels. In this paper I have used the same set of individual, household and community characteristics to explain these different stages of schooling succession independently with a Heckman Sample Selection Model. Further I have also analyzed the model separately for male, female and rural samples. I have used NFHS, 1998-99 database on India, and in an effort to account for the cluster design of this database, I have controlled for cluster fixed effects in a separate model. There is significantly strong evidence of gender difference favouring the male child at both the stages of schooling. But, in the cluster fixed effects model, though the gender effect is still very significant, yet it is favouring girls. Inside a cluster girls are estimated to acquire more years of schooling relative to boys provided that they are already enrolled. Parental schooling has a very positive and statistically significant impact on child schooling. There is also very significant wealth effect, community specific effect and regional disparities between states in India in case of child schooling

Introduction

Despite continuous efforts by developing countries over the past five decades to expedite the process of development through emphasis on education, most of them have not yet succeeded in their goal. The problem is much more complex in India, Where all levels of education from primary to higher are heavily subsidised by the government. In recent years, a number of initiatives have been undertaken to make higher education partially self-financing. The country is still dealing with challenges in its primary education, where after five decades of concerted efforts through the planning process, in 2002/2003, an estimated 82% of children in the age group of 6-14 were enrolled in school (Wikipedia, 2007). However, the drop-out rates both for males and females at the primary and at the secondary school level continue to remain high. This creates a kind of mismatch in the schooling progression, where the problem lies not only with the access to primary education but also with the retention and achievement at the primary, secondary and higher secondary levels.

To increase the educational level of the population we need to address the sequential nature of this problem like access to schooling, retention in schooling and achievement through schooling. Despite a plethora of research on child schooling in India, previous studies on this issue have not quite successful in dealing with this situation in one integrated model; they have either focused on current enrolment, grade attainment or failure. The main objective of this research is to study this sequential problem of schooling progression in one combined effort through primary focus on demand for schooling, though at the same time it would deal with the supply side considerations. This paper uses NFHS 1998-99 data from India to explore the

economic and socio-economic factors determining demand for schooling at the household level.

In this attempt we need to proceed in two different steps, first we need to understand the factors responsible for access to primary school, and second contingent on this the factors responsible for continuing with further schooling. This study has used the same set of individual, household and community characteristics to explain these different stages of schooling succession independently with a Sample Selection Model with/without cluster-fixed effects. Further it has also analyzed the model separately for male, female and rural samples to interpret any dissimilarity and to identify any peculiarity.

The rest part of the paper is organized in the following order. In the next section, I review the relevant literature. Section 3, covers the methodology and econometric estimation strategy used in this paper. The dataset is discussed in section 4. Following the discussion on dataset in section 4, I interpret the empirical results in section 5. Finally, the main conclusions are presented in Section 6.

2. Literature Review

There is extensive literature on factors responsible for low child schooling (Kingdon, 2005; Kremer *et al*, 2005; Kochar, 2004; Pal, 2004a; Lillard and Willis, 1994) in developing countries; some analyzing the factors responsible for lower enrolment rates at the primary, secondary and higher secondary schools (Borooah and Iyer 2005; Pal, 2004b; Glewwe and Jacoby, 2004; Kambhampati and Pal, 2001; Drez and Kingdon, 2001; Kingdon, 1998; Tansel, 1997; Handa, 1996; Deolalikar, 1993; Singh, 1992) and some analyzing the factors for highest grade attainment(Maitra, 2003; Glick and Sahn, 2000) and some for current enrolment rates and drop-outs (Maitra,

2003; Glick and Sahn, 2000; Patrinos and Psacharopoulos, 1997) . Different sets of individual, household and community characteristics are analyzed and interpreted as possible reasons behind these low schooling outcomes.

Cluster Fixed Effects

One of the important considerations in case of cross-sectional household surveys in developing countries is the cluster design of data. The households within a cluster might significantly vary on the basis of ethnicity, religion, wealth, occupation and educational status, yet there might be some neighbourhood effects and correlation between them. Like the local eccentricities are copied by those who live near one another and become more uniform within a cluster. There will usually be more homogeneity within clusters than between clusters (Dancer and Rammohan, 2007; Deaton, 2000). However, it is also not always possible to observe all the common characteristics shared by the households within a cluster, and ignoring some of those unobserved heterogeneity in the regression model would result in omitted variable bias in the estimated coefficients. For this purpose, the use of cluster fixed-effects estimation models would control for all of the common observed as well as unobserved characteristics shared by children from the same cluster and then run within cluster regressions to account for any child level heterogeneity within the cluster.

3. Methodology

The empirical model in this paper is based on the collective household framework (Alderman et al, 1995; Hoddinott and Haddad, 1995; Chiappori, 1988 and McElroy and Horney, 1981). Under this framework, investment in the human capital of a daughter or son in the household depends on the relative bargaining power of each parent. Again, the level of parental investment in the schooling of their children is

driven by their desire to equate the marginal benefits of the schooling investments to their costs.

Following, Glewwe and Jacoby (2004), I divide the households' investment into human capital (H) and physical capital (K). The human capital is accumulated through schooling investment in school age children (e), purchasing a schooling input such as books, uniforms and transportation costs etc. (x), at price p.

$$H = \psi (G^m(e, x) + G^f(e, x)) \quad (1)$$

G^m and G^f are neo-classical investment functions of the parents (mother and father) in households on their children and ψ is the learning productivity parameter that reflects school quality and child ability and motivation.

The households generate their current income Y by combining their physical capital with their labor in the production function,

$$Y = \theta F(K^m, K^f, L^m, L^f).$$

θ is a productivity parameter reflecting the state of technology, K^m and K^f are the physical capital already under the possession of mother and father in the household and, L^m and L^f are their respective labor. The household finances all their investment by forgoing their current consumption c.

Now the physical capital in the household accumulates according to

$$K = \theta F(K^m, K^f, L^m, L^f) - px - c \quad (2)$$

Further, we have set $K \geq 0$, suggesting that households cannot hold negative amount of physical capital. This is like a 'borrowing constraint'.

Suppose parents derive utility from both investment in their children's schooling and current consumption, the utility of the mother and father can be modeled as,

$$U^i = u\,(c, e); \qquad i=m,f \tag{3}$$

Further the reservation utility levels of both the parents are represented by $\tilde{U}^m$ and $\tilde{U}^f$, which represents their threshold utility levels or options outside the marriage if their marriage breaks down. An improvement in the reservation utility of any one of the parents would be expected to improve his/her bargaining power in the household. Hence, mothers with better earning abilities, education and improved social and economic status are considered to have better bargaining power in the household decision making process in the allocation of resources towards their children's schooling. The household's objective is then to maximize the collective utility of both the parents,

$$U^H = [U^m\,(c, e) - \tilde{U}^m\,(p, K^m; \gamma^m)] \times [U^f\,(c, e) - \tilde{U}^f\,(p, K^f; \gamma^f)] \tag{4}$$

subject to the constraints (1) and (2) along with a borrowing constraint.

The schooling investment of a child, h_i, is then given by a simple reduced form demand function:

$$h_{i=} f(P, I, z, \varepsilon_i) \tag{5}$$

P, is a vector of parental characteristics, which include the age of both parents, their educational status, their labour-market status, father's occupational status and finally mother's economic and social status. I, is a vector of household characteristics, like wealth, household size and proportion of daughters in the household. z, is a vector of individual child characteristics, such as age, age squared, gender, birth-order and, ε represents unobserved individual, household and community characteristics influencing the child's schooling. It is hard to observe the schooling preferences in a

household. Instead, we observe the actual number of years of schooling that the child has currently attained. Hence, considering that these choices are the representation of household's preference for children's schooling, I have used the 'number of years of schooling' as the dependent variable for this analysis.

There is however one important limitation for this measure to be used with the NFHS data on India, as the data has a large number of zeros. There is a significant proportion (13%) of children with no schooling at all. The variable years of schooling, is a continuous variable but, it is censored at zero. To deal with this issue of sample selection bias, I have adopted Heckman sample selection model, where, in stage 1, I estimate a probit equation examining the probability of a child going to school and in the stage 2, I estimate the determinants of years of schooling using Maximum Likelihood Estimation method (MLE), choosing only those children who attended school at stage 1. So, I am interested in estimating

$$h_i = \beta x_i + \varepsilon_i \qquad (6)$$

Here, h_i represents the child's years of schooling and it is a continuous variable, x_i is a vector of explanatory variables, β is the vector of coefficients that will be estimated and ε_i is a random error term. Now, the selection equation is

$$h_{1i}^* = \sigma x_{1i} + \upsilon_i \qquad [x \subset x_1] \qquad (7)$$

h_{1i}^* is a latent variable representing the household desire or preference to enroll a child in school, which can be expressed as a linear function of variables that affect the probability of a child attending school. However, we do not observe h_1^* and instead we observe the dummy variable h_{1i}, which takes the value 1 if the child is enrolled in a school and 0 otherwise. Hence,

$$h_{1i} = 1 \quad \text{iff} \quad h_{1i}^* > 0$$

$$h_{1i} = 0 \quad \text{iff} \quad h_{1i}^* \leq 0 \qquad (8)$$

and h_i is observed when h_{1i} is 1. The sample selection model allows for a correlation coefficient between the disturbances of the two equations. That is the distribution of υ_i and ε_i is given as,

$$\begin{bmatrix} v_i \\ e_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} s_e^2 & s_e r \\ s_e r & 1 \end{bmatrix} \right)$$

if these disturbances are uncorrelated i.e. if the estimated ρ is not significant, then the above h_i equation could be estimated by the OLS. However, in this model ρ is very significant.

Cluster Fixed Effects Model

The NFHS India 1998-99, is a cluster based survey, where it is based on two-stage stratified sample of households. The stratification process breaks down a single survey into multiple independent surveys, one for each stratum. In this case, a uniform sample design was adopted in all the states. In each state, the rural sample was selected in two stages: the selection of primary sampling units (PSUs), which are villages with probability proportional to population size (PPS) at the first stage, followed by random selection of households using within each PSU in the second stage. The urban sample was selected in a three-stage procedure. In the first stage wards were selected using PPS sampling. Then one census enumeration block (CEB) was randomly selected from each sample ward and finally households were randomly selected within each sample CEB. So each cluster in the data represents a PSU and there are a total of 2495 clusters in my full sample. In this manner the households in a single cluster live near to one another, they are not randomly distributed over space rather they are geographically grouped. Furthermore, my database has no information on community characteristics such as the proximity and availability of schools, quality of schools for the full sample. Ignoring these supply-side community or cluster specific characteristics is likely to result in biased estimates. Considering this, an

estimation of cluster fixed effects model helps to overcome this cluster specific correlation between the households which is likely to result in inconsistent estimates, ignoring the cluster level unobserved heterogeneity in the model which is also likely to result in omitted variable bias and simultaneously it would investigate for the presence of any possible intra-cluster variation in the characteristics.

However, recall the fact that a non-trivial percentage (13%) of children in the database has no schooling at all. Consequently, the dependant variable, years of schooling, is a continuous one which is censored at zero. Hence, the optimizing behavior in this model leads to a corner solution, which necessitates the need to care for sample selectivity bias in the model. I have estimated the model in two different stages to correct for selectivity bias. In the first stage, I estimate a standard probit regression to examine the probability that, a child in the household is enrolled in school. Including the *Inverse Mills Ratio* obtained from this first stage probit regression, I have then calculated a cluster fixed effects regression model with unbalanced clusters in the second stage. Here, I am interested in estimating

$$h_{ij} = \beta x_{ij} + a_i + \varepsilon_{ij.} \qquad (9)$$

However, I can't estimate this equation directly, instead in the first stage, I estimate the selection equation,

$$h_{1ij}^* = \sigma x_{1ij} + \upsilon_{ij} \qquad [x_{ij} \subset x_{1ij}] \qquad (10)$$

h_{1ij}^* is a latent variable representing the household desire or preference to enroll a child in school, which can be expressed as a linear function of variables that affect the probability of a child attending school. However, we do not observe h_{1ij}^* and instead we observe the dummy variable h_{1ij}, which takes the value 1 if the child is enrolled in a school and there are at least 2 children per cluster present in the sample, and 0 otherwise. Hence,

$$h_{1ij} = 1 \quad \text{iff} \quad h_{1ij}^* > 0$$

$$h_{1ij} = 0 \quad \text{iff} \quad h_{1ij}^* \leq 0 \qquad (11)$$

and h_{ij} is observed when h_{1ij} is 1.

4. Data

For this study, I am using the *National Family Health Survey (NFHS)*, 1998-99 dataset from India to empirically analyse the demand for schooling. I focus both on the access to schooling and contingent on this on the demand for continuing in school. The dataset is rich, unique, and is nationally representative. It has not been previously used to analyse this issue. The survey is administered on 90,303 ever-married women aged 15-49 years and it contains detailed information on household structure, labour market participation, asset ownership, health and educational characteristics for all the household members. For the purpose of this study I focus on the fifteen major states of India, which represents more than 80% of the population in the country.

The descriptions and definitions of the variables used in the regression model are presented in Tables 1 and 2. The paper focuses on children aged between 10-20 years. There are a total of 61,022 children in this age group, of which 13 % of these children in the sample have no education. From Table 1.1 below, we observe that among the main reasons cited for not been to school, 25% report that it costs too much, another 25.7 % of children do not attend school because they are required for either household/farm/family business work or for earning an income outside the family, 21% are not interested in studies, whereas only 3.6% cited the school is too far away.

The remaining 87 % have some level of education. Among those with some level of schooling, 36.5 % are with primary school education, 42.5 % are with secondary school education and only 7.9 % are with higher secondary school education. This

indicates very high drop-out rates at primary and secondary levels of schooling in India.

Another striking feature of the data is that, of the whole sample, 33,506(55 %) children are male. Among those with zero-years of schooling, we observe that a major proportion of them are female (63.5 %), whereas, among those with some level of schooling only 42.3 % are female.

There are 42,733 rural and 18,289 urban children in our sample. The supply-side community level variables representing school characteristics are only available for rural areas and considering the missing values in these variables in the village file, we are only left with 37, 758 children in the rural sample.

In the Indian educational system there are five grades within the primary school and five grades within the secondary school followed by two years of higher secondary school. Six is typically considered to be the normal primary school enrolment age although late enrolments are a common feature in many rural areas. Children in this age group, either an offspring of the household head or any other subsidiary member of the household are included in the sample, allowing for the presence of joint-family households a fairly common feature in India. In this way, only households with at least one child in this age cohort contribute for a total of 29,299 households in the sample.

Explanatory variables

The primary focus of our analysis is to examine the factors responsible for schooling of 10-20 year old children in India. We therefore include a range of individual,

parental and household level characteristics (including wealth) along with community level variables and fifteen state level dummies as control variables.

The individual characteristics considered here are the child's age, age squared (to capture a possible non-linear effect), gender and the birth order. The child's gender is a dichotomous variable that takes on a value of '1' for males and '0' for females. Similarly, the child's birth order is captured through a set of dummies representing absolute birth order variables, if he/she is the 1st, or 2nd, or 3rd, or 4th, or 5th and above child in the family.

The parental characteristics include the age of both parents, which are continuous variables; a set of dummies representing their educational status; dummies for their employment status like if they were employed in the last 12 months prior to the survey; a set of dummies representing the father's occupation, whether he is employed in professionals, clerical and sales jobs/agriculture-self employed/skilled manual jobs/unskilled manual jobs and finally the dummies representing the mother's economic and social status in the household.

The household characteristics play very important role in this type of analysis, where the schooling decision of the child is dependant on the economic status of the household. Hence, variables including the household size, sibling characteristics in the household like the number, age composition and gender composition of the siblings currently residing in the household are considered vital for analysis. The presence of large households sometimes considered to be imposing financial burden on the household resources. Hence, the household size is included in this study as a proxy for resource constraints in households in addition to the wealth index. It is expected to be

exerting an inverse pressure on the available resources for schooling. Though, it can also be interpreted as the presence of more earning members in the household there by exerting positive effect on the schooling outcome of the children. Similarly the proportion of female children in the family is also considered in this analysis to capture if the resource constrained households are really in disadvantage for availing resources for schooling of children, when the proportion of female children in the household increases.

The household financial resources are controlled through a set of dummy variables representing the household wealth. Though a major shortcoming of the NFHS dataset on India is that it contains no information on earnings and household expenditure patterns yet, it contains a household wealth index that divides households into five different wealth quintiles (WEALTH1-WEALTH5), with WEALTH1 representing the poorest quintile. The wealth index is calculated using the households' assets ownership, so it is not endogenous to the model, neither is it affected by the transitory nature of the labour income in the database, there by providing a reasonably reliable measure of the household economic status. Using this index, it is observed that among those children with zero-years of schooling in the sample, 47% are from households belonging to the lowest quintile and 28.3% from the second lowest quintile, and only 1.2% from the highest wealth quintile. But, among those children with some years of schooling only 13.1% are from the lowest quintile and 51% from households belonging to the highest two wealth quintiles.

The effect of caste and religion, which are observed exerting very strong influence in Indian society (Borooah and Iyer, 2005; Drez and Kingdon, 2001), is also included in this study through a set of caste dummies and a dummy variable representing if the

household is Hindu or non-Hindu. Using the caste structure, it can be interpreted that among those children with zero years of schooling 43% are from scheduled caste and scheduled tribe households, whereas among those with some years of schooling their proportion is only 24.7%.

Finally, a dummy variable representing the rural or urban residence of the household and a set of fifteen state level dummies are included to capture the effect of urbanisation and regional variations in schooling outcomes in Indian states.

5. Estimation results

Heckman Selection Model without Cluster Fixed Effects

The main results of this analysis are presented in Tables 3-5. The results for the Two-stage sample selection models with and without cluster fixed effects are presented in Table 3, where the first two columns represent the estimates of Heckman Sample Selection Model using the full information Maximum Likelihood Estimation (MLE) without controlling for cluster fixed effects. The decision to estimate this model using Heckman's Two Step full Maximum Likelihood Estimation procedure, is ofcourse supported by a very significant r, the correlation coefficient between the error terms of two equations, in the second-stage MLE of the model.

As already discussed, with the sample selection model we are able to analyse the factors that affect a child's years of education contingent upon whether or not the child is enrolled in a school. Considering first the probit and then the second stage maximum likelihood results, both the age and gender of the child do matter in the decision to enrol and once enrolled in the decision to continue in school. As the age of the child increases it is less likely that he/she would get enrolled in a school whereas, in the second stage a rise in the age of the child is an indicator of more years of

education being acquired. The child's gender is significant in both the stages. While in the first stage being a male child, profoundly raises the probability (7.3%) that he would be enrolled in a school; in the second stage, it also raises the estimated period of schooling by 0.178 years more than a female child. So, in both the stages the child's gender is coming out as a very significant determining factor in the decision to enrol as well as in the decision to continue in school. The variables representing the absolute birth orders of children in the family are also very significant at 1% level in both the stages of schooling. In case of a second born child in the family, it is 0.5% less likely that he/she would enrol in a school in comparison to a first born child. Similarly, in case of a third, fourth, and fifth-and-above born child, it is not only less likely that he/she would enrol in a school, but also his/her estimated years of schooling are lower in relation to the first born child.

In our sample there is strong evidence of significant resource constraints on child school enrolment and further on child's continuity in school. The variables, those proxies for household resources, such as the household size, proportion of daughters present in the family, parental educational status and labour market status are significant. Although we recognise the potential endogeneity of these variables, they are still included in the model due to their importance and also due to the lack of suitable instruments. Our results show that, household size is very significant even at 1% level with negative coefficients in both stages of the regression, where a single member increase in the household size makes 0.2% less probable that the child would enrol in a school, in the second stage it also decreases the number of years in school by 0.026 years. Similarly, proportion of daughters in the family, is also very significant at 1% level but with positive coefficients, where an increase in the

proportion of daughters in the family, 1.3% enhances the probability of a child to enrol in school, it also increases the estimated schooling by almost 0.197 years.

Furthermore, in case of the variables representing parental schooling status, relative to the base category of no schooling, both mother's and father's rising schooling levels like primary, secondary and higher secondary, do have significantly positive effects on the likelihood of children's attending school and hence, on the number of years continuing in school. There is 3% more probability for a child with primary-schooled father to enrol in a school and 0.139 more estimated years of schooling in relation to a child whose father has no-schooling at all. Whereas, for a child with secondary-schooled father there is 5% more probability to enrol in a school and 0.561 more estimated years of schooling and this figure even rises to 0.977 more estimated years of schooling in case of a child with higher-secondary-schooled father. Mother's schooling levels have almost the similar impact in case of primary schooling (3.2%) or less in case of secondary (3.6%) and higher secondary (3.3%) schooling than father's schooling levels on the likelihood of child's enrolment in school. In terms of child's estimated years of schooling, mother's primary schooling has almost double (0.274) the impact than father's, though the impact of mother's secondary schooling is almost closer (0.557) to that of father's and mother's higher secondary schooling has much lesser (0.599) impact while weighing against that of father's, in relation to a child whose mother has no-schooling at all.

Similarly, the variables representing father's occupational status are also quite essential for child schooling. Compared to the base category of children, whose fathers do not work, to those children in the sample with fathers employed in official jobs are estimated to have 0.196 more years of schooling and, even the fraction of

children whose fathers are employed in agriculture and self-employed, have estimated 0.117 years more of schooling. These variables have however, no significant effect on the likelihood of child's enrolment. The labour market status of the father is also very significant even at 1% level. The children whose fathers worked all through the year prior to the study, are estimated to have 0.193 years more in school than the base category of children whose fathers were not employed in the year prior to the study. However, mother's employment status that is mother being employed in paid or unpaid employment relative to the base category of mother didn't work all through the previous year, are significant for decisions on children continuing in school, with negative effects. Children with mothers in paid employment are 1.7% less likely to attend school and even 0.332 less estimated years in school compared to those children whose mothers didn't work all through the year prior to the study. Whereas for children whose mothers are in unpaid employment, though do not face any significant consequence on the likelihood of enrolment yet, they are estimated to have 0.222 less years in school. This is probably a peculiar feature for predominantly rural-Indian dataset, where a working mother in the household means the household is resource constrained and the mother is mostly working as a casual labourer. This argument is in parlance with Behrman *et al*. (1999). Further the variables representing mothers' economic status in the household, such as if the mother is allowed to have money set aside and if the mother needs permission to go to market, have very significant and positive influence on the decision to enrol a child in school.

Again, the significance of all of these factors does not however, weaken the stance of household wealth, a direct indicator of household resources, in the model. Both the probit and second-stage ML estimates represent that an increase in household wealth significantly increases the probability of a child being enrolled in school and also the

child's estimated years in school. Not quite unexpected though, the greatest increase in the probability of enrolment as well as in the estimated years of schooling comes from belonging to the wealthiest quintile compared to the lowest wealth quintile. In relation to a child belonging to the first (poorest) wealth quintile, a child belonging to the second wealth quintile is nearly 2%, a child from third wealth quintile is 3.6%, a child from fourth wealth quintile is 5.3% and in the end a child from the fifth wealth quintile is 7.3% more likely to enrol in a school. Likewise, in terms of estimated years of schooling, a child from second wealth quintile is estimated to spend 0.304 years, a child from third wealth quintile is estimated to spend 0.636 years, a child from fourth wealth quintile is estimated to spend 0.990 years, a child from fifth wealth quintile is estimated to spend 1.56 years, more in school.

It is already established in literature, the influence of community norms in Indian society such as religion -Hindu or non-Hindu and, caste -scheduled or non-scheduled on child schooling (Borooah and Iyer, 2005; Drez and Kingdon, 2001) are very strong. My results at both the stages of schooling further substantiate this phenomenon. For being a child from a Hindu household, it is 2.5% more likely to enrol in a school and, also estimated to have 0.427 more years of schooling, in comparison to a child from non-Hindu household. On the contrary, scheduled caste and scheduled tribe children are significantly in disadvantage not only in case of access to school but also in continuing at school relative to the base category of general caste children.

Finally, most of the state level dummies are also very significant compared to the base category of Uttar Pradesh (U.P.), the biggest state in India[1]. While, children from states like Bihar and Rajasthan are less likely to enrol in school, they are estimated to obtain more years of schooling than those from U.P. Children from Himachal Pradesh, Karnataka, Kerala, Maharastra, Orissa, and Tamil Nadu are not only more likely to enrol in school, but also estimated to spend more years in school than those from U.P. However, we need more insight to explain a negatively significant coefficient for the state of Punjab in case of access to school and even more surprising the negative coefficient for West Bengal in terms of estimated years in school in relation to U.P.

Cluster Fixed Effects Model

In the cluster fixed effects model, in the first stage I am estimating a standard probit regression, to account for the sample selection bias and, in the second stage a cluster fixed effects model with unbalanced clusters. The estimates of these two stage regression are presented in the last two columns (3rd and 4th) of Table 3. Here, the decision to estimate the model in two different stages is supported by a very significant *Inverse Mills Ratio* (*the* variable that is included from the first stage probit regression) in the second stage estimation of the model.

As it is evident, in the cluster fixed effects model we are able to analyse the factors that affect a child's years of schooling dependent upon the fact that the child has already enrolled in a school. It is not possible to successfully infer about what happens in the initial stage, in case of access to school with cluster fixed effects[2], when the households are faced with a decision to enrol a child in the school or not. So here the

[1] Now this state has been divided into two different separate states, yet in our database it is represented as a single state.

[2] At this stage it is not appropriate to estimate a fixed effect probit regression to deal with the sample selection bias, instead we estimated a standard probit regression (Wooldridge, 2002).

first stage standard probit estimates are expected to be same as the first-stage probit estimates of the Heckman selection model without the cluster fixed effects.

In case of the cluster fixed effects model, like the Heckman Selection Model, the age of the child is very significant in the decision to continue in school. A rise in the age of the child is an indicator of more estimated years of schooling is to be acquired. Similarly, the child's gender is also significant yet, unlike the Heckman Selection Model, it is with a negative coefficient. So in this case, a female child is estimated to acquire 0.124 more years of schooling than a male counterpart. The variables representing the absolute birth-order of the child are also very significant in this model. Unlike, the Heckman selection model, in the cluster fixed effects model a second born child in the family also faces very significant and negative consequences in terms of the years of acquiring schooling in a similar way as the third, fourth and fifth-and-above born child, in comparison to the base category of a first born child, though they vary in the magnitudes of coefficients.

In case of the variables representing father's schooling status, relative to the base category of father with no-schooling, children with higher-secondary-schooled fathers are estimated to acquire 0.295 more years of schooling. This effect was significantly more, 0.977 more years, in case of Heckman Selection Model and even in contrast, is the negative and significant coefficient for children with primary-schooled fathers. Similarly, in case of father's occupational status, compared to the base category of children, whose fathers do not work, to those children in the sample with fathers employed in professional, clerical and sales jobs have estimated to spend 0.208 years more in school. However, different from the Heckman Selection Model, the occupational category such as, if the father employed in agriculture-self employed and

also the labour market status of the father such as, if he was employed in last twelve months, have no significance on children's estimated years in school.

Now considering, the variables representing mother's schooling, relative to the base category of mother with no-schooling, children with secondary and higher-secondary-schooled mothers are estimated to acquire significantly more years of schooling. The impact of these mother's schooling variables are even more than the impact of fathers schooling variables. It is 0.322years more in school in case of children with secondary-schooled mothers and 0.395years more in school in case of children with higher-secondary-schooled mothers. Distinct from the Heckman selection model, children with primary-schooled-mothers show no significant improvement in terms of years spend in school. Same is the case with mother's employment status that is, the mother being in paid or unpaid employment relative to the base category of mother didn't work all through the previous year, has no significant influence on a child continuing in school.

There is also very strong evidence of significant resource constraints on child's continuity in school. Almost all of the variables, as in the case with Heckman selection model, which are considered as indirect measures of household resources like the household size, proportion of daughters in the family, are significant. The variable household size is very significant (at 1% level) with negative coefficient (0.013) whereas, the variable proportion of daughters in the family, is also significant (at 5% level) with positive coefficient (0.035). The variables representing household wealth, a direct measure of household resources are also very significant determining factors for the child's continuity in school. The greatest increase in the yeas of schooling is ofcourse comes from belonging to the wealthiest quintile compared to the

lowest wealth quintile. However, the coefficients of these variables or the extent of impacts on child's number of years in school is considerably less when considered within a cluster. In addition to that, unlike in the Heckman Selection Model the variable representing a child from the second wealth quintile compared to the base category of a child from the first (poorest) wealth quintile has no significant influence on child's years of schooling inside a cluster.

The influence of community characteristics even within a cluster (village) such as religion -Hindu or non-Hindu or, caste -scheduled, non-scheduled and other-backward-classes on child schooling are very strong. Children belonging to non-Hindu households, scheduled-caste, scheduled-tribe and even other-backward-classes households are less likely to continue in school than general caste children. However, the influence of religion in notably less (0.191 years) within a cluster than what we have observed (0.427 years) in Heckman Selection Model. Similarly, the caste category such as, the other-backward-classes was only a significant influence at 10% level on child schooling in the Heckman selection model, it is in fact very significant influence at 1% level with a larger impact when considered inside a cluster.

Heckman Selection Model Independently for Male and Female Samples

The child's gender is apparently a very significant determining factor in the decision concerning schooling in both the stages of regression model, and in both the cases with and without controlling for cluster fixed effects. Considering this, I have estimated the model independently for male and female samples to accommodate for the gender-specific impact of these explanatory variables. The results of the male and female models, estimated separately using the Heckman's two-step procedure are presented in Table 4. A likelihood ratio test justifies this decision to estimate these

two separate models, as there is a significant difference in the male and female coefficients.

The age of the child is a very significant deciding factor in the decision to continue in school in both the male and female models; whereas it is also a significant decision (at 5% level) with negative effect at the enrolment level for a female child, that is with the rising age it is less likely that a female child would enrol in a school. The absolute birth order of a child is also very significant influencing factor for his/her schooling though not always. In case of a female child, even if she is the 2nd born in the family still it has significant negative influence (8.6%) on the decision to enrol in a school relative to a 1st born. Similarly, a 3rd born female child is not only significantly (10% level) less likely (8.3%) to enrol in a school but also she is estimated to spend 0.204 less years in school. For the 4th and 5th born female children they are estimated to spend respectively 0.245 and 0.363 fewer years in school than a 1st born. Contrary to this, in case of a male child, while it does matter with negative consequences at both the stages of schooling for being a 3rd, 4th or 5th and above born in the family relative to the 1st born, it does not matter at all for being the 2nd born.

In line with the literature, it is also quite evident in both the male and the female models that the parental characteristics have significant influence on children's schooling decision. The variables representing father's schooling status, like children with primary or secondary or higher-secondary schooled fathers are in significantly favourable circumstances in comparison to children with no-schooling fathers at both the stages of schooling, with only one exception of female children with primary-schooled fathers, they make no significant difference in terms of number of years in school. The influence of father's schooling status is also observed to be more on male

children than on female children, with the only deviation of children with higher-secondary-schooled fathers. Female children with primary schooled fathers are 42.4% more likely to enrol in school, where as in case of male children they are 56.5% more likely to enrol in school and also estimated to continue 0.228 more years in school than those with fathers without any schooling. Likewise, female children with secondary-schooled fathers are 75.2% more likely to enrol in school and estimated to continue 0.447 more years in school and these two figures for male children rises to 83.5% in terms of probability of enrolment and 0.678 years more in school. However, father's higher-secondary schooling though beneficial for both the males and females, it has stronger influence on female children's likelihood of enrolment (111.9%) than that of male children (89.8%), and male children's number of years in school, (1.113more years) than that of female children (0.858 more years).

Now, fathers occupation such as, if the father works in professionals, clerical and sales jobs or if he works in agriculture-self employed relative to the base category of not-working at all puts the male child in a significantly advantageous situation at both the stages of schooling though not the female child. Conversely, fathers employment status such as, if he was employed in last 12 months, is very significant (at 1% level) with positive influence only for the female child in terms of years in school, she is estimated to spend 0.252 years more in school. On top, if the father was employed for last 12 months in official job then it is favourable (significant at 5% level)for a female child to access school, it is 13.4% more likely that she would enrol, relative to the base category of father not employed in the last 12 months.

In an exactly similar manner to father's schooling status, the mother's schooling status such as if she is primary-schooled or secondary-schooled or higher-secondary-

schooled, relative to the base category of she is with no-schooling, are very significant with positive effects even at 1% level for both the male and female child in terms of both access to school as well as years in school. Female (Male) children with primary-schooled mothers are 59.7% (47.2%) more probable to enrol in school and approximated to spend 0.331(0.243) years more in school. Likewise, female (male) children with secondary-schooled (secondary school and above) mothers are 80.6% (44.2%) more likely to enrol in school and estimated to spend 0.68 (0.46) years more in school. More than this all these mothers schooling variables have greater impact on a female child than on a male child. Their impacts are even stronger than the impacts of fathers schooling status for a female child and weaker for a male child. Now, in case of the variables for mother's employment status like, if she works in paid employment or unpaid employment compared to the base category of she is not working, has very significant negative impact on female schooling at both stages. If the mother of a female child works in paid (unpaid) employment and/or self employed, then she is 30% (22.8%) less likely to enrol in school and spend 0.404 (0.305) fewer years in school

Now, the household characteristics basically both the indirect and direct measures of household resource potential, such as household size and proportion of daughters in the family and household wealth index are very significant with exactly the same characteristics as in the previous models for both male and female schooling. However the impact of household wealth index is stronger on female schooling than on male schooling, with certainly more rise observed in case of the wealthiest quintiles than the lowest wealth quintiles.

The community characteristics are also quite important for both male and female child schooling. Female children belonging to non-Hindu households and scheduled castes or scheduled tribes households are significantly less likely to enroll in school and also spend fewer estimated years in school than those from general caste. At the same time, female children from other-backward-classes households are only less likely to enroll in school (significant at 5% level) and they face no disadvantage in terms of estimated years in school. Similarly male children from non-Hindu households are also in disadvantage in both stages of schooling, with relatively additional disadvantageous to female children. But, male children from schedule caste households or from other backward classes are in disadvantage only in terms of years in school while those from schedule tribe households are in disadvantage at the enrolment level.

Most of the regional (state level) variables are significant in both the male and female models. Like a female child from the state of Andhra Pradesh, Haryana, Himachal Pradesh, Karnataka, Kerala, Maharastra, Orissa, Punjab and Tamilnadu is more likely to be enrolled in a school and also expected to spend more years in school than those from the state of Uttar Pradesh (base category). In case of Bihar, a female child is estimated to continue more years in school but it is less likely that she would enroll in a school; whereas in Rajasthan it is less likely that she would even get enrolled in a school than in Uttar Pradesh. However, male children only from states like Himachal Pradesh, Maharastra and Tamilnadu are in significantly favorable circumstances at both stages of schooling than those from Uttar Pradesh while those from Kerala and Orissa are only better of in terms of years in school. Male children from Andhra Pradesh and Karnataka are less likely to enroll in school but are estimated to continue more years in school in comparison to Uttar Pradesh. A rather surprising finding is

that, male children from the prosperous states like Gujarat and Punjab, also with those from Bihar (one of the poorest states) are less likely to enroll in a school. Nonetheless negative and significant coefficients for both female and male children in terms of years in school in the state of West Bengal in relation to those in Uttar Pradesh are yet to be clarified.

Heckman Selection Model for Rural Sample

In view of the predominance of rural sample in the NFHS-India data and also the fact that some of the locality-specific supply-side information on availability of schools and on school quality are presented only at the village-level, I have estimated a Heckman Sample Selection model exclusively for the rural sample. I have included the supply-side variables like the distance to the school (primary, middle, secondary, higher-secondary and college); distance to the nearest town, district-headquarter; distance to other infrastructural facilities like the local post office, telegraph office, telephone booth and bank; distance from the nearest transport facility and finally information on availability of rural development programmes like the Indira Awas Yojana (IAY), the Integrated Rural Development Programme (IRDP), the National Rural Employment Programme (NREP), Training Rural Youth for Self Employment (TRYSEM), Employment Guarantee Scheme (EGS), Development of Women and Children of Rural Areas (DWARCA), Sanjay Gandhi Niradhar Yojana (SGNY) in this model. These variables are considered equally important in the literature on schooling. Even so the supply side variables that influence schooling decisions are also likely to be endogenous in nature (Dancer and Rammohan, 2007; Rosenzweig and Wolpin, 1988). That is households might be selective in their place of residence depending on the availability of these infrastructure (Rosenzweig and Wolpin, 1988). But, given that I have only selected the rural sample for this analysis and selective migration between villages in India in response to local infrastructure is not a

common practice, inclusion of these variables in the model appear reasonable. The estimates of this rural model are presented in Table 5.

In this rural model, the age and the gender of the child also significantly influence his/her schooling and they perform in a similar manner as in case of the Heckman model without controlling for cluster fixed effects. Yet, the influence of gender is strikingly high on both the stages of child schooling. A male child is 91% more likely to enrol in school and estimated to spend 0.32 additional years in school than a female child. The variables representing the absolute birth orders of children are also very significant for schooling but they are not as obvious as in the earlier models. That is a 2^{nd} born child is 6.3% less likely to enrol in a school, while the third born is 8% less likely to enrol and also estimated to obtain 0.183 fewer years of schooling and the 4^{th} and 5^{th}-and-above born children are estimated to obtain respectively 0.257 and 0.320 less years of schooling than the 1^{st} born in the family.

Once again the parental characteristics like both father's and mother's schooling status are very significant for child schooling at both the stages. Yet, the magnitudes of impacts of these variables are noticeably higher than what they were at the enrolment stage of schooling in case of the Heckman Selection Model without Cluster Fixed Effects. Then again, in case of years spent in school they exhibit more or less closer impacts as in case of the Heckman Selection Model without Cluster Fixed Effects. In case of father's occupational status, only the children with fathers working in professionals, clerical and sales jobs are significantly advantageous to those with fathers not-working, in terms of number of years in school. Similarly, father's employment status, indicating if he is employed in the last 12 months than being not employed all through this period, is also significant in terms of obtaining more years

of schooling. In contrast, variables for mother's employment status, that is if she is either in paid or in unpaid employment in last 12 months compared to her not working, are significant with negative coefficients at both the stages of schooling. Mother's economic status, that is, if she needs permission to go to market has very significant negative influence on child's enrolment decision.

As it has been always the case, the variables representing the measures of household resources, such as the household wealth, household size and proportion of daughters in the family are also very significant for child schooling in this rural model. But, the extent of impact of the variables indicating households belonging to different wealth quintiles are markedly stronger than those in case of access to school in Heckman Selection model without Cluster Fixed Effects. A child from a household belonging to the 2nd wealth quintile is 28.1%, from a household belonging to the 3rd wealth quintile is 61.2%, from a household belonging to the 4th wealth quintile is 99.6% and from a household belonging to the 5th wealth quintile is 150.8% more likely to enrol in school than a child from the poorest wealth quintile.

Likewise, the variables representing community characteristics like religion and caste in a village are also very significant with more or less closer influence to those in the cluster fixed effects model.

Some of the supply-side variables for schooling representing the school proximity and quality, availability of other basic infrastructural facilities and also the availability of government implemented rural credit and employment guarantee programmes in villages are very significant for child schooling in this model. Nearly 90.5% villages in my sample have a primary school inside the village. So, while the distance to

primary school does not matter, distance to the middle schools is only significant (10% level) with a negligible negative coefficient in case of number of years in school. The distance to secondary school is very significant with negative effect at both the stages of schooling but, again with very small coefficients. In case of other facilities such as the distance to the district head quarter and also the distance to the post office are very significant with negative coefficients only at the enrolment level. The distance to the available transport service is significant (5% level) at both the stages of schooling with negative effects. Likewise, some of the rural credit and employment guarantee programmes like number of *IRDP* beneficiaries in the village is also significant with negative impact on the decision to enrol a child in school, while number of *egs* beneficiaries in the village is significant with negative impact at both the stages of schooling and the number of *sgny* beneficiaries is significant also with negative impact for the decision to continue in school.

In the end, the regional variables infer that rural children from states like Himachal Pradesh, Karnataka, Kerala, Maharastra and Orissa are more likely to enroll in school and also estimated to spend more years in school than those from Uttar Pradesh. Contrary to this, rural children from Bihar and Rajasthan are less likely to enrol in a school and those from Madhya Pradesh are estimated to spend fewer years in school. A rather surprising finding is that, those from Harayana are only more likely to enroll in a school and those from Gujurat and Punjab are only estimated to spend more years in school.

6. Conclusion

In this paper I have analysed the factors responsible for low child schooling in India. Considering the sequential nature of this problem I have estimated a Heckman Sample Selection model for the full sample and also separately for male, female and rural

samples. In order to account for the cluster design of data in NFHS-India, 1998-99, I have also estimated a cluster fixed effects model in two different steps, where in the first step a standard probit model is estimated concerning the selectivity bias in the model and in the second step a cluster fixed effects model is estimated including the *inverse mills ratio* from the probit. A detailed analysis and comparison of the estimates of these models reveal, while the individual characteristics of the child like age, gender and absolute birth order in the family are significant determining factors for child schooling outcomes, the parental characteristics like the schooling level of the parents, their occupational structure and employment status are also equally significant and so also the household resource potential, community characteristics and regional characteristics such as different states in India. These variables ofcourse vary in the extent and nature of their impact not only between models but also at different stages of schooling.

There is very strong evidence of significant gender difference favouring the male child, at both the stages of schooling in both the full-sample model and also in the rural model. Considering this, I have separately estimated the male and female samples, where I encounter the same set of parental and household characteristics have differential impact on male and female child. Keeping in line with the literature, parental schooling has a very positive and statistically significant impact on child schooling, while the impact of mother's schooling is stronger on female children the impact of father's schooling is stronger on males, except for fathers with higher secondary schooling. But when I consider the cluster design of data in the cluster fixed effects model looking for the intra-cluster variation in the same set of explanatory variables, though gender is still very significant for child schooling, but girls are estimated to spend more years in school relative to boys. There is very strong

evidence of significant resource constrain in households for availing schooling resources to children and this result is all most unanimous in all the models. The analysis also supports that an increase in the wealth in households has stronger impact on girls schooling relative to boys. It is always the case in India that community characteristics like caste and religion has a very strong influence in the society so my results coincide with this prevalent concept in case of child schooling.

Reference:

Alderman, H., Chiappori, P.A., Haddad, L., Hoddinott, J. and R. Kanbur, 1995. "Unitary versus Collective Models of the Household: Time to Shift the Burden of Proof?" *World Bank Research Observer*, 10(1): 1 – 20.

Behrman, Jere R., Foster, A., Rosenzweig, M. and P. Vahsishtha, 1999, "Women's Schooling, Home Teaching, and Economic Growth", Journal of Political Economy, 107(4): 682-715.

Borooah, V. K. and Sriya Iyer, 2005, "*Vidya*, *Veda* and *Varna*: The Influence of Religion and Caste on Education in Rural India" *The Journal of Development Studies*, Vol. 41, No. 8, pp. 1369-1404.

Chiappori, P., 1988, "Nash Bargained Household Decisions." *International Economic Review*, 29(4):791-96.

Dancer, D and Rammohan, A (2007). "Maternal Education and Child Schooling in Nepal", *Research in Labor Economics*, Vol 26, pp.365-397.

Deaton, Angus, 2000, *The Analysis of Household Surveys*, The Johns Hopkins University Press, Baltimore, Maryland.

Deolalikar, A.B., 1993, "Gender Differences in the Returns to Schooling and in School Enrollment Rates in Indonesia" *The Journal of Human Resources*, XXVIII, 4. pp. 899-932.

Dreze, Jean and Gita Gandhi Kingdon, 2001, "School Participation in Rural India" *Review of Development Economics*, 5(1), pp1-24.

Glewwe, Paul and Hanan G. Jacoby, 2004, Economic growth and the demand for education: is there a wealth effect?" *Journal of Development Economics*, 74, pp. 33-51.

Glick, Peter and David E. Sahn, 2000, "Schooling of girls and boys in a West African Country: the effects of parental education, income and household structure" *Economics of Education Review*, 19, pp.63-87.

Handa, Sudhanshu, 1996, "The Determinants of Teenage Schooling in Jamaica: Rich vs. Poor, Females vs. Males", *The Journal of Development Studies*, Vol.32, No.4, pp. 554-580.

Hoddinott, J. and L. Haddad, 1995. "Does Female Income Share Influence Household Expenditures? Evidence from Cote d'Iviore." *Oxford Bulletin of Economics and Statistics*, Vol.57, No.1, pp.77-96.

Kambhampati, U.S. and Sarmistha Pal, 2001, "Role of Parental Literacy in Explaining Gender Difference: Evidence from Child Schooling in India" *The European Journal of Development Research*, vol. 13, No. 2, pp.97-119.

Kingdon, G.G., 2005, "Where Has All the Bias Gone? Detecting Gender Bias in the Intrahousehold Allocation of Educational Expenditure." *Economic Development and Cultural Cha*nge, pp.409-451.

Kingdon, G.G., 1998, "Does the Labour Market Explain Lower Female Schooling in India?" *The Journal of Development Studies*, Vol.35, No.1, pp. 39-65.

Kochar, Anjini 2004, "Urban influences on rural schooling in India", *Journal of Development Economics*, vol.74, pp.113-136.

Kremer *et al*, 2005, "Teacher Absence in India: A Snapshot", *Journal of European Economic Association*, 3(2-3):658-667.

Lillard, L.A. and Robert J. Willis, 1994, "Intergenerational Educational Mobility: Effects of Family and State in Malaysia", *The Journal of Human Resources* 29(4):1126-1167.

Maitra, Pushkar, 2003, "Schooling and Educational Attainment: Evidence from Bangladesh" *Education Economics*, Vol. 11, No. 2, pp.129-153.

McElroy, M.B. and M.J. Horney, 1981. "Nash Bargained Household Decisions." *International Economic Review*, 22(2): 333-50.

Pal, Sarmistha, 2004a, "Child schooling in Peru: Evidence from a sequential analysis of schooling progression", *Journal of Population Economics*, 17:657-680

Pal, Sarmistha, 2004b, "How much of the Gender Difference in Child School Enrolment Can be Explained? Evidence from Rural India." *Bulletin of Economic Research*, 56-2, 0307-3378, pp.133-158.

Patrinos, Harry Anthony and George Psacharopoulos, 1997, "Family size, schooling and child labour in Peru – An empirical analysis", *Journal of Population Economics,* 10:387-405.

Rosenzweig, M.R. and K.R. Wolpin, 1988, "Migration Selectivity and Effects of Public Programs", *Journal of Public Economics*, 37(3): 265-289.

Singh, Ram D., 1992, "Underinvestment, Low Economic Returns to Education, and the Schooling of Rural Children: Some Evidence from Brazil", *Economic Development and Cultural Change*, Vol. 40, No.3, 645-664.

Tansel, Aysit, 1997, "Schooling Attainment, Parental Education and Gender in Cote d' Ivoire and Ghana" *Economic Development and Cultural Change*, Vol.45, No.4, pp.825-856.

Wooldridge J.M., 2002, *Introductory Econometrics, A Modern Approach.* Thomson, Australia.

Wooldridge J.M., 2002, *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts, London, England.

Wikipedia, latest update, "The Education in India*", The Free Encyclopedia,* viewed on 19th Jan. 2007, <http://en.wikipedia.org/wiki/Education_in_India>

Table 1: Descriptive Statistics

Variables	Full Sample Mean (Std. Dev.) (N=61022)	Rural Mean (Std. Dev.) (N=37758)
Child Characteristics		
Child's school attendance(enrolled in a school =1)	0.87	0.83
Child's years of schooling (corrected for enrolled in school=1)	6.43 (3.01)	6.04 (2.90)
Gender (1 if male, 0 otherwise)	0.55	0.55
Age (in years)	14.15 (3.05)	14.04 (3.03)
Age squared	209.68 (89.72)	206.27 (88.92)
Child's birth order (1 if the 1st born, 0 otherwise)	0.28	0.27
Child's birth order (1 if the 2nd born, 0 otherwise)	0.25	0.24
Child's birth order (1 if the 3rd born, 0 otherwise)	0.19	0.19
Child's birth order (1 if the 4th born, 0 otherwise)	0.12	0.13
Child's birth order (1 if the 5th born, 0 otherwise)	0.15	0.16
Parental Characteristics		
Father's age	43.66 (7.67)	43.56 (8.15)
Father's education (1 if no-schooling, 0 otherwise)	0.29	0.37
Father's education (1 if primary-school, 0 otherwise)	0.21	0.22
Father's education (1 if secondary-school, 0 otherwise)	0.34	0.32
Father's education (1 if higher secondary, 0 otherwise)	0.15	0.09
Father's occupation (1 if not-working)	0.02	0.02
Father's occupation (1 if works in professionals, clerical & sales jobs)	0.22	0.13
Father's occupation (1 if works in agriculture-self employed)	0.40	0.55
Father's occupation (1 if works in household-domestic, services & other skilled manual jobs)	0.25	0.19
Father's occupation (1 if works in unskilled manual jobs)	0.11	0.11
Father's employment status (1 if he was employed in last 12 months)	0.39	0.45
Father's employment status (1 if in last 12 months he worked in office job)	0.05	0.04
Mother's age	37.16 (5.62)	36.95 (5.75)
Mother's education (1 if no-schooling, 0 otherwise)	0.59	0.71
Mother's education (1 if primary-school, 0 otherwise)	0.18	0.17
Mother's education (1 if secondary-school, 0 otherwise)	0.18	0.11
Mother's education (1 if higher secondary, 0 otherwise)	0.05	0.01
Mother's employment status (1 if the she didn't work in last 12 months)	0.59	0.53
Mother's employment status (1 if she works in paid employment and/or self emplo yed)	0.26	0.27
Mother's employment status (1 if she works in unpaid employment)	0.15	0.20
Mother's economic status (1 if she is allowed to have money set aside)	0.62	0.57

Table 1 continued… **variables**	**Mean** **(Std. Dev.)**	**Mean** **(Std. Dev.)**
Mother's economic status (1 if she needs permission to go to market)	0.64	0.72
Mother's social status (1 if she decides on her own or is involved in the decision about her staying with the) family	0.50	0.47
Mother's social status (1 if she needs permission to visit relatives or friends)	0.73	0.78
Household Characteristics		
Household size (number of household members)	6.88 (3.23)	7.08 (3.34)
Proportion of daughters in the family	0.42 (0.28)	0.42 (0.27)
Household wealth (1 if it is in 1st wealth quintile)	0.18	0.25
Household wealth (1 if it is in 2nd wealth quintile)	0.17	0.23
Household wealth (1 if it is in 3rd wealth quintile)	0.20	0.24
Household wealth (1 if it is in 4th wealth quintile)	0.22	0.20
Household wealth (1 if it is in 5th wealth quintile)	0.23	0.09
Community Characteristics		
Religion (1 if the household head is Hindu)	0.84	0.88
Caste (1 if the household belongs to Schedule Caste)	0.19	0.21
Caste (1 if the household belongs to Schedule Tribe)	0.08	0.10
Caste (1 if the household belongs to Other Backward Classes)	0.34	0.36
Caste (1 if the household belongs to other castes)	0.39	0.33
Village Characteristics		
Distance to educational facility (primary school)		0.74 (6.26)
Distance to educational facility (middle school)		1.97 (4.81)
Distance to educational facility (secondary school)		4.37 (6.20)
Distance to educational facility (higher secondary		8.51 (10.30)
Distance to educational facility (college)		19.76 (17.64)
Distance to nearest town (km)		14.27 (12.55)
Distance to district headquarters (km)		43.62 (26.81)
Distance to nearest railway station (km)		26.30 (25.73)
Distance to available transport service (km)		12.22 (26.34)
Distance from all-weather road (km)		12.96 (27.40)
Distance to other facility (post office)		2.16 (5.10)
Distance to other facility (telegraph office)		9.87 (10.68)
Distance to other facility (std booth)		9.73 (10.93)
Distance to other facility (bank)		5.58 (7.17)

Table 1 continued… **Variables**	**Mean** **(Std. Dev.)**	**Mean** **(Std. Dev.)**
irdp beneficiaries		20.60 (36.11)
nrep beneficiaries		3.02 (13.47)
trysem beneficiaries		2.70 (11.23)
egs beneficiaries		2.72 (13.48)
dwacra beneficiaries		4.84 (15.29)
iay beneficiaries		11.49 (21.34)
sgny beneficiaries		2.10 (11.21)
Regional Characteristics		
Place of residence (1 if the household resides in an urban area)	0.30	
State (1 if the state is Andhra Pradesh)	0.05	0.05
State (1 if the state is Bihar)	0.10	0.14
State (1 if the state is Gujarat)	0.06	0.05
State (1 if the state is Haryana)	0.05	0.05
State (1 if the state is Himachal Pradesh)	0.05	0.05
State (1 if the state is Karnataka)	0.05	0.06
State (1 if the state is Kerala)	0.03	0.02
State (1 if the state is Madhya Pradesh)	0.10	0.11
State (1 if the state is Maharastra)	0.07	0.05
State (1 if the state is Orissa)	0.06	0.08
State (1 if the state is Punjab)	0.05	0.04
State (1 if the state is Rajasthan)	0.10	0.11
State (1 if the state is Tamil Nadu)	0.05	0.04
State (1 if the state is West Bengal)	0.05	0.01
State (1 if the state is Uttar Pradesh)	0.13	0.15

Table 2: Descriptive Statistics

variables	Males Mean (Std. Dev.) (N=33506)	Females Mean (Std. Dev.) (N=27516)
Child Characteristics		
Child's school attendance(enrolled in a school =1)	0.91	0.82
Child's years of schooling (corrected for enrolled in school=1)	6.50 (2.97)	6.34 (3.06)
Age (in years)	14.35 (3.13)	13.92 (2.93)
Age squared	215.65 (92.64)	202.40 (85.46)
Child's birth order (1 if the 1st born, 0 otherwise)	0.28	0.29
Child's birth order (1 if the 2nd born, 0 otherwise)	0.25	0.25
Child's birth order (1 if the 3rd born, 0 otherwise)	0.19	0.19
Child's birth order (1 if the 4th born, 0 otherwise)	0.12	0.13
Child's birth order (1 if the 5th born, 0 otherwise)	0.15	0.15
Parental Characteristics		
Father's age	43.75 (7.59)	43.56 (7.76)
Father's education (1 if no-schooling, 0 otherwise)	0.30	0.28
Father's education (1 if primary-school, 0 otherwise)	0.21	0.21
Father's education (1 if secondary-school, 0 otherwise)	0.34	0.35
Father's education (1 if higher secondary, 0 otherwise)	0.15	0.16
Father's occupation (1 if not-working)	0.03	0.02
Father's occupation (1 if works in professionals, clerical & sales jobs)	0.21	0.23
Father's occupation (1 if works in agriculture-self employed)	0.41	0.39
Father's occupation (1 if works in household-domestic, services & other skilled manual jobs)	0.25	0.26
Father's occupation (1 if works in unskilled manual jobs)	0.11	0.10
Father's employment status (1 if he was employed in last 12 months)	0.39	0.39
Father's employment status (1 if in last 12 months he worked in office job)	0.05	0.05
Mother's age	37.26 (5.60)	37.03 (5.64)
Mother's education (1 if no-schooling, 0 otherwise)	0.61	0.57
Mother's education (1 if primary-school, 0 otherwise)	0.17	0.18
Mother's education (1 if secondary-school, 0 otherwise)	0.17	0.19
Mother's education (1 if higher secondary, 0 otherwise)	0.05	0.05
Mother's employment status (1 if the she didn't work in last 12 months)	0.60	0.59
Mother's employment status (1 if she works in paid employment and/or self employed)	0.26	0.26
Mother's employment status (1 if she works in unpaid employment)	0.15	0.15
Mother's economic status (1 if she is allowed to have money set aside)	0.62	0.62
Mother's economic status (1 if she needs permission to go to market)	0.65	0.64

Table 2 continued… **variables**	**Mean** **(Std. Dev.)**	**Mean** **(Std. Dev.)**
Mother's social status (1 if she decides on her own or is involved in the decision about her staying with the) family	0.49	0.50
Mother's social status (1 if she needs permission to visit relatives or friends)	0.74	0.72
Household Characteristics		
Household size (number of household members)	6.78 (3.26)	7.00 (3.19)
Proportion of daughters in the family	0.29 (0.24)	0.59 (0.23)
Household wealth (1 if it is in 1st wealth quintile)	0.18	0.17
Household wealth (1 if it is in 2nd wealth quintile)	0.18	0.17
Household wealth (1 if it is in 3rd wealth quintile)	0.20	0.19
Household wealth (1 if it is in 4th wealth quintile)	0.22	0.22
Household wealth (1 if it is in 5th wealth quintile)	0.22	0.24
Community Characteristics		
Religion (1 if the household head is Hindu)	0.84	0.84
Caste (1 if the household belongs to Schedule Caste)	0.20	0.19
Caste (1 if the household belongs to Schedule Tribe)	0.08	0.08
Caste (1 if the household belongs to Other Backward Classes)	0.34	0.34
Caste (1 if the household belongs to other castes)	0.38	0.39
Regional Characteristics		
Place of residence (1 if the household resides in an urban area)	0.30	0.30
State (1 if the state is Andhra Pradesh)	0.05	0.05
State (1 if the state is Bihar)	0.10	0.10
State (1 if the state is Gujarat)	0.05	0.06
State (1 if the state is Haryana)	0.05	0.05
State (1 if the state is Himachal Pradesh)	0.05	0.05
State (1 if the state is Karnataka)	0.05	0.05
State (1 if the state is Kerala)	0.03	0.04
State (1 if the state is Madhya Pradesh)	0.10	0.10
State (1 if the state is Maharastra)	0.08	0.07
State (1 if the state is Orissa)	0.06	0.06
State (1 if the state is Punjab)	0.05	0.04
State (1 if the state is Rajasthan)	0.11	0.10
State (1 if the state is Tamil Nadu)	0.05	0.05
State (1 if the state is West Bengal)	0.05	0.05
State (1 if the state is Uttar Pradesh)	0.13	0.13

Table 3: Results for the Sample Selection Model with/without Cluster Fixed Effects

Variables	Heckman Selection Model (Full Maximum Likelihood Estimation)		Sample Selection Model with Cluster Fixed Effects	
Child Characteristics	**Probit Marginal Effects (SE) (N=61022)**	**Second-stage MLE Estimates Coef. (SE) (N=53060)**	**Probit Marginal Effects (SE) (N=61022)**	**Fixed effects Estimates Coef. (SE) (N=53057)**
Gender (1 if male, 0 otherwise)	0.073***	0.178***	0.073***	-0.124***
	(0.003)	(0.036)	(0.003)	(0.021)
Age (in years)	-0.005***	1.627***	-0.005**	1.642***
	(0.001)	(0.046)	(0.002)	(0.029)
Age squared	-0.000	-0.035***	0.000	-0.035***
	(0.000)	(0.002)	(0.000)	(0.001)
Child's birth order (1 if the 2nd born, 0 otherwise)	-0.005***	-0.098	-0.005**	-0.090***
	(0.001)	(0.078)	(0.002)	(0.021)
Child's birth order (1 if the 3rd born, 0 otherwise)	-0.007***	-0.161**	-0.007***	-0.145***
	(0.001)	(0.072)	(0.002)	(0.025)
Child's birth order (1 if the 4th born, 0 otherwise)	-0.007**	-0.242***	-0.007***	-0.220***
	(0.004)	(0.069)	(0.003)	(0.032)
Child's birth order (1 if the 5th born, 0 otherwise)	-0.006***	-0.320***	-0.006**	-0.276***
	(0.002)	(0.060)	(0.003)	(0.035)
Parental Characteristics				
Father's age	0.000	-0.004**	0.000	-0.003
	(0.000)	(0.002)	(0.000)	(0.002)
Father's education (1 if primary-school, 0 otherwise)	0.030***	0.139***	0.031***	-0.256***
	(0.001)	(0.028)	(0.002)	(0.032)
Father's education (1 if secondary-school, 0 otherwise)	0.052***	0.561***	0.054***	-0.013
	(0.002)	(0.034)	(0.002)	(0.033)
Father's education (1 if higher secondary, 0 otherwise)	0.045***	0.977***	0.047***	0.295***
	(0.002)	(0.038)	(0.002)	(0.040)
Father's occupation (1 if works in professionals, clerical & sales jobs)	0.003	0.196***	0.002	0.208***
	(0.004)	(0.061)	(0.005)	(0.062)
Father's occupation (1 if works in agriculture-self employed)	0.006	0.117**	0.005	0.098
	(0.004)	(0.052)	(0.004)	(0.061)
Father's occupation (1 if works in household-domestic, services & other skilled manual jobs)	0.004	0.040	0.004	0.034
	(0.003)	(0.046)	(0.004)	(0.061)
Father's occupation (1 if works in unskilled manual jobs)	-0.008	-0.007	-0.009*	0.064
	(0.006)	(0.035)	(0.006)	(0.065)
Father's employment status (1 if he was employed in last 12 months)	0.002	0.193***	0.001	0.095
	(0.005)	(0.061)	(0.005)	(0.078)
Father's employment status (1 if in last 12 months he worked in office job)	0.005	-0.005	0.005	-0.046
	(0.003)	(0.040)	(0.004)	(0.042)
Mother's age	-0.000***	0.022***	0.000**	0.023***
	(0.000)	(0.002)	(0.000)	(0.003)
Mother's education (1 if primary-school, 0 otherwise)	0.032***	0.274***	0.034***	0.033
	(0.002)	(0.022)	(0.002)	(0.025)
Mother's education (1 if secondary-school, 0 otherwise)	0.036***	0.557***	0.037***	0.322***
	(0.002)	(0.013)	(0.002)	(0.027)
Mother's education (1 if higher secondary, 0 otherwise)	0.033***	0.599***	0.031***	0.395***
	(0.005)	(0.075)	(0.005)	(0.041)
Mother's employment status (1 if she works in paid employment and/or self employed)	-0.017***	-0.332***	-0.016***	-0.122
	(0.005)	(0.097)	(0.006)	(0.078)
Mother's employment status (1 if she works in unpaid employment)	-0.009	-0.222***	-0.009	-0.074
	(0.007)	(0.047)	(0.006)	(0.079)
Mother's employment and child's gender (1 if the mother is in paid employment and the child is male)	-0.006***		-0.006**	
	(0.001)		(0.003)	
Mother's economic status (1 if she is allowed to have money set aside)	0.004**		0.004***	
	(0.002)		(0.001)	
Mother's economic status (1 if she needs permission to go to market)	-0.006***		-0.006***	
	(0.001)		(0.002)	
Mother's social status (1 if she decides on her own or is involved in the decision about her staying with the) family	0.003		0.002	
	(0.002)		(0.001)	

Table 3 continued…	**M.E (SE)**	**Coef. (SE)**	**M.E (SE)**	**Coef. (SE)**
Mother's social status (1 if she needs permission to visit relatives or friends)	0.005** (0.003)		0.006** (0.003)	
Household Characteristics				
Household size (number of household members)	-0.002*** (0.000)	-0.026*** (0.005)	-0.002*** (0.000)	-0.013*** (0.003)
Proportion of daughters in the family	0.013*** (0.003)	0.197*** (0.053)	0.013*** (0.003)	0.086** (0.035)
Household wealth (1 if it is in 2nd wealth quintile)	0.019*** (0.003)	0.304*** (0.063)	0.020*** (0.001)	0.019 (0.037)
Household wealth (1 if it is in 3rd wealth quintile)	0.036*** (0.003)	0.636*** (0.073)	0.038*** (0.002)	0.126*** (0.042)
Household wealth (1 if it is in 4th wealth quintile)	0.053*** (0.003)	0.990*** (0.089)	0.055*** (0.002)	0.406*** (0.047)
Household wealth (1 if it is in 5th wealth quintile)	0.073*** (0.004)	1.563*** (0.112)	0.074*** (0.003)	0.913*** (0.053)
Community Characteristics				
Religion (1 if the household head is Hindu)	0.025*** (0.004)	0.427*** (0.053)	0.025*** (0.003)	0.191*** (0.031)
Caste (1 if the household belongs to Schedule Caste)	-0.004*** (0.001)	-0.102*** (0.014)	-0.005** (0.002)	-0.124*** (0.028)
Caste (1 if the household belongs to Schedule Tribe)	-0.022*** (0.003)	-0.176*** (0.019)	-0.023*** (0.003)	-0.110** (0.049)
Caste (1 if the household belongs to Other Backward Classes)	-0.003 (0.004)	-0.043* (0.024)	-0.004** (0.002)	-0.063*** (0.023)
Regional Characteristics				
Place of residence (1 if the household resides in an urban area)	-0.012*** (0.004)	-0.150*** (0.016)	-0.012*** (0.003)	
State (1 if the state is Andhra Pradesh)	-0.004 (0.004)	0.779*** (0.032)	-0.004 (0.004)	
State (1 if the state is Bihar)	-0.039*** (0.008)	0.113** (0.050)	-0.040*** (0.004)	
State (1 if the state is Gujarat)	-0.006 (0.006)	0.102* (0.061)	-0.005 (0.004)	
State (1 if the state is Haryana)	0.007** (0.002)	0.068 (0.043)	0.007* (0.004)	
State (1 if the state is Himachal Pradesh)	0.031*** (0.003)	0.416*** (0.030)	0.032*** (0.002)	
State (1 if the state is Karnataka)	0.009*** (0.002)	0.925*** (0.037)	0.009*** (0.003)	
State (1 if the state is Kerala)	0.033*** (0.003)	1.342*** (0.061)	0.034*** (0.002)	
State (1 if the state is Madhya Pradesh)	0.002 (0.002)	-0.102** (0.041)	0.002 (0.003)	
State (1 if the state is Maharastra)	0.026*** (0.002)	0.610*** (0.023)	0.027*** (0.002)	
State (1 if the state is Orissa)	0.015*** (0.001)	0.593*** (0.047)	0.016*** (0.002)	
State (1 if the state is Punjab)	-0.012** (0.006)	0.126** (0.055)	-0.013** (0.006)	
State (1 if the state is Rajasthan)	-0.020*** (0.002)	0.056 (0.056)	-0.021*** (0.003)	
State (1 if the state is Tamil Nadu)	0.029*** (0.001)	0.858*** (0.040)	0.030*** (0.002)	
State (1 if the state is West Bengal)	0.012** (0.005)	-0.672*** (0.125)	0.013*** (0.003)	
_cons		-11.834*** (0.384)		-10.235*** (0.222)
rho		-0.196*** (0.006)		
mills				-2.675*** (0.094)

*** indicates significant at 1% level, ** indicates significant at 5% level, * indicates significant at 10% level

Table 4: Results for the Sample Selection Model: Independently for Males and Females

Variables	Heckman Selection Model Females		Heckman Selection Model with Males	
Child Characteristics	**Probit Marginal Effects(SE) (N=27516)**	**Second-stage MLE Estimates Coef. (SE) (N=22456)**	**Probit Marginal Effects(SE) (N=33506)**	**Second-stage MLE Estimates Coef. (SE) (N=30604)**
Age (in years)	-0.011**	1.492***	0.000	1.717***
	(0.005)	(0.054)	(0.002)	(0.058)
Age squared	0.000	-0.030***	0.000**	-0.038***
	(0.000)	(0.002)	(0.000)	(0.002)
Child's birth order (1 if the 2nd born, 0 otherwise)	-0.010**	-0.092	-0.003	-0.107
	(0.005)	(0.091)	(0.002)	(0.074)
Child's birth order (1 if the 3rd born, 0 otherwise)	-0.010	-0.204**	-0.006**	-0.136*
	(0.006)	(0.091)	(0.003)	(0.078)
Child's birth order (1 if the 4th born, 0 otherwise)	-0.005	-0.245**	-0.009**	-0.250***
	(0.005)	(0.104)	(0.004)	(0.052)
Child's birth order (1 if the 5th born, 0 otherwise)	-0.004	-0.363***	-0.009**	-0.300***
	(0.005)	(0.069)	(0.004)	(0.089)
Parental Characteristics				
Father's age	0.000	-0.003**	0.000*	-0.004
	(0.000)	(0.002)	(0.000)	(0.004)
Father's education (1 if primary-school, 0 otherwise)	0.040***	0.051	0.027***	0.228***
	(0.003)	(0.060)	(0.001)	(0.014)
Father's education (1 if secondary-school, 0 otherwise)	0.075***	0.447***	0.044***	0.678***
	(0.003)	(0.079)	(0.003)	(0.019)
Father's education (1 if higher secondary, 0 otherwise)	0.075***	0.858***	0.033***	1.113***
	(0.003)	(0.074)	(0.002)	(0.046)
Father's occupation (1 if works in professionals, clerical & sales jobs)	-0.005	0.067	0.007**	0.276**
	(0.012)	(0.048)	(0.003)	(0.120)
Father's occupation (1 if works in agriculture-self employed)	0.000	-0.070	0.010***	0.239**
	(0.011)	(0.075)	(0.003)	(0.109)
Father's occupation (1 if works in household-domestic, services & other skilled manual jobs)	0.000	-0.056	0.007	0.097
	(0.007)	(0.082)	(0.005)	(0.090)
Father's occupation (1 if works in unskilled manual jobs)	-0.019	-0.123*	-0.002	0.064
	(0.016)	(0.070)	(0.005)	(0.075)
Father's employment status (1 if he was employed in last 12 months)	0.008	0.252***	-0.004	0.147
	(0.006)	(0.036)	(0.006)	(0.121)
Father's employment status (1 if in last 12 months he worked in office job)	0.014**	0.012	0.001	-0.002
	(0.006)	(0.063)	(0.004)	(0.024)
Mother's age	-0.001***	0.024***	0.000**	0.021***
	(0.000)	(0.003)	(0.000)	(0.004)
Mother's education (1 if primary-school, 0 otherwise)	0.052***	0.331***	0.023***	0.243***
	(0.004)	(0.029)	(0.002)	(0.036)
Mother's education (1 if secondary-school, 0 otherwise)	0.065***	0.680***	0.022***[3]	0.460***
	(0.004)	(0.015)	(0.004)	(0.024)
Mother's education (1 if higher secondary, 0 otherwise)	0.040***	0.676***		
	(0.014)	(0.096)		
Mother's employment status (1 if she works in paid employment and/or self employed)	-0.039***	-0.404***	-0.008	-0.283*
	(0.008)	(0.075)	(0.006)	(0.162)
Mother's employment status (1 if she works in unpaid employment)	-0.030***	-0.305***	0.004	-0.168
	(0.011)	(0.059)	(0.006)	(0.113)
Mother's economic status (1 if she is allowed to have money set aside)	0.007**		0.003*	
	(0.003)		(0.002)	

[3]For the male sample this Mother's education variable is 1if mother's educational attainment is secondary school and above and 0 otherwise.

Table 4 continued…	**M.E (SE)**	**Coef. (SE)**	**M.E (SE)**	**Coef. (SE)**
Mother's economic status (1 if she needs permission to go to market)	-0.006 (0.005)		-0.006*** (0.001)	
Mother's social status (1 if she decides on her own or is involved in the decision about her staying with the) family	0.005 (0.004)		0.001 (0.001)	
Mother's social status (1 if she needs permission to visit relatives or friends)	0.006 (0.006)		0.004*** (0.002)	
Household Characteristics				
Household size (number of household members)	-0.004*** (0.001)	-0.023*** (0.004)	-0.002*** (0.000)	-0.029*** (0.007)
Proportion of daughters in the family	0.018** (0.009)	0.156** (0.074)	0.008* (0.004)	0.182** (0.091)
Household wealth (1 if it is in 2nd wealth quintile)	0.033*** (0.004)	0.267*** (0.069)	0.014*** (0.003)	0.352*** (0.065)
Household wealth (1 if it is in 3rd wealth quintile)	0.059*** (0.005)	0.619*** (0.097)	0.027*** (0.002)	0.696*** (0.084)
Household wealth (1 if it is in 4th wealth quintile)	0.087*** (0.006)	1.009*** (0.100)	0.038*** (0.002)	1.030*** (0.106)
Household wealth (1 if it is in 5th wealth quintile)	0.118*** (0.006)	1.677*** (0.149)	0.052*** (0.003)	1.529*** (0.103)
Community Characteristics				
Religion (1 if the household head is Hindu)	0.027*** (0.008)	0.330*** (0.042)	0.024*** (0.002)	0.504*** (0.088)
Caste (1 if the household belongs to Schedule Caste)	-0.017*** (0.003)	-0.116*** (0.030)	0.002 (0.002)	-0.103*** (0.024)
Caste (1 if the household belongs to Schedule Tribe)	-0.034*** (0.009)	-0.092** (0.042)	-0.016*** (0.004)	-0.242 (0.052)
Caste (1 if the household belongs to Other Backward Classes)	-0.015** (0.006)	-0.059 (0.045)	0.002 (0.003)	-0.040** (0.021)
Regional Characteristics				
Place of residence (1 if the household resides in an urban area)	-0.003 (0.007)	-0.087*** (0.027)	-0.019*** (0.003)	-0.204*** (0.042)
State (1 if the state is Andhra Pradesh)	0.017*** (0.003)	0.904*** (0.085)	-0.022*** (0.008)	0.692*** (0.045)
State (1 if the state is Bihar)	-0.042*** (0.014)	0.219** (0.091)	-0.044*** (0.010)	0.026 (0.055)
State (1 if the state is Gujarat)	0.005 (0.012)	0.151** (0.075)	-0.016*** (0.004)	0.080 (0.057)
State (1 if the state is Haryana)	0.018*** (0.006)	0.172** (0.073)	-0.001 (0.004)	-0.002 (0.030)
State (1 if the state is Himachal Pradesh)	0.057*** (0.004)	0.598*** (0.062)	0.015*** (0.003)	0.289*** (0.054)
State (1 if the state is Karnataka)	0.037*** (0.002)	1.111*** (0.065)	-0.014*** (0.003)	0.798*** (0.056)
State (1 if the state is Kerala)	0.060*** (0.004)	1.623*** (0.074)	0.015* (0.009)	1.106*** (0.072)
State (1 if the state is Madhya Pradesh)	0.015*** (0.006)	0.040 (0.046)	-0.008* (0.005)	-0.195*** (0.048)
State (1 if the state is Maharastra)	0.047*** (0.003)	0.790*** (0.055)	0.014*** (0.003)	0.489*** (0.034)
State (1 if the state is Orissa)	0.033*** (0.002)	0.724*** (0.082)	0.005 (0.003)	0.521*** (0.041)
State (1 if the state is Punjab)	0.016*** (0.005)	0.292*** (0.083)	-0.036*** (0.010)	0.001 (0.055)
State (1 if the state is Rajasthan)	-0.064*** (0.006)	-0.056 (0.042)	0.000 (0.002)	0.088 (0.063)

Table 4 continued…	**M.E (SE)**	**Coef. (SE)**	**M.E (SE)**	**Coef. (SE)**
State (1 if the state is Tamil Nadu)	0.053***	1.073***	0.017***	0.721***
	(0.002)	(0.035)	(0.003)	(0.051)
State (1 if the state is West Bengal)	0.028***	-0.605***	0.001	-0.709***
	(0.008)	(0.143)	(0.004)	(0.113)
_cons		-10.834***		-12.351***
		(0.344)_		(0.487)
rho		-0.210***		
		(0.024)		

*** indicates significant at 1% level, ** indicates significant at 5% level, * indicates significant at 10% level

Table 5: Results for the Sample Selection Model: Rural Sample

Variables	Probit Marginal Effects (SE) (N=37758)	Second-stage MLE Estimates Coef. (SE) (N=31316)
Child Characteristics		
Gender (1 if male, 0 otherwise)	0.111***	0.320***
	(0.002)	(0.047)
Age (in years)	-0.008***	1.681***
	(0.002)	(0.038)
Age squared	0.000	-0.038***
	(0.000)	(0.001)
Child's birth order (1 if the 2nd born, 0 otherwise)	-0.007***	-0.122
	(0.002)	(0.075)
Child's birth order (1 if the 3rd born, 0 otherwise)	-0.009***	-0.183**
	(0.003)	(0.076)
Child's birth order (1 if the 4th born, 0 otherwise)	-0.009	-0.257***
	(0.006)	(0.085)
Child's birth order (1 if the 5th born, 0 otherwise)	-0.006	-0.320***
	(0.005)	(0.076)
Parental Characteristics		
Father's age	0.000**	-0.005***
	(0.000)	(0.002)
Father's education (1 if primary-school, 0 otherwise)	0.043***	0.135***
	(0.002)	(0.042)
Father's education (1 if secondary-school, 0 otherwise)	0.070***	0.534***
	(0.002)	(0.033)
Father's education (1 if higher secondary, 0 otherwise)	0.062***	0.934***
	(0.002)	(0.028)
Father's occupation (1 if works in professionals, clerical & sales jobs)	0.005	0.190**
	(0.009)	(0.095)
Father's occupation (1 if works in agriculture-self employed)	0.005	0.119
	(0.009)	(0.114)
Father's occupation (1 if works in household-domestic, services & other skilled manual jobs)	0.002	0.034
	(0.007)	(0.129)
Father's occupation (1 if works in unskilled manual jobs)	-0.017	-0.051
	(0.014)	(0.110)
Father's employment status (1 if he was employed in last 12 months)	0.003	0.220**
	(0.004)	(0.106)
Father's employment status (1 if in last 12 months he worked in office job)	0.002	0.103
	(0.007)	(0.106)
Mother's age	-0.001***	0.019***
	(0.000)	(0.002)
Mother's education (1 if primary-school, 0 otherwise)	0.045***	0.263***
	(0.002)	(0.027)
Mother's education (1 if secondary-school, 0 otherwise)	0.055***	0.459***
	(0.003)	(0.033)
Mother's education (1 if higher secondary, 0 otherwise)	0.066***	0.446***
	(0.002)	(0.148)
Mother's employment status (1 if she works in paid employment and/or self employed)	-0.029***	-0.348***
	(0.004)	(0.129)
Mother's employment status (1 if she works in unpaid employment)	-0.014***	-0.265**
	(0.005)	(0.110)
Mother's employment and child's gender (1 if the mother is in paid employment and the child is male)	-0.008***	
	(0.002)	
Mother's economic status (1 if she is allowed to have money set aside)	0.002	
	(0.003)	
Mother's economic status (1 if she needs permission to go to market)	-0.011**	
	(0.005)	

Table 5 continued…	**M.E (SE)**	**Coef. (SE)**
Mother's social status (1 if she decides on her own or is involved in the decision about her staying with the) family	0.002 (0.002)	
Mother's social status (1 if she needs permission to visit relatives or friends)	0.007 (0.005)	
Household Characteristics		
Household size (number of household members)	-0.003*** (0.001)	-0.018*** (0.006)
Proportion of daughters in the family	0.019*** (0.006)	0.150*** (0.057)
Household wealth (1 if it is in 2nd wealth quintile)	0.027*** (0.006)	0.303*** (0.060)
Household wealth (1 if it is in 3rd wealth quintile)	0.052*** (0.006)	0.659*** (0.083)
Household wealth (1 if it is in 4th wealth quintile)	0.070*** (0.003)	1.053*** (0.087)
Household wealth (1 if it is in 5th wealth quintile)	0.067*** (0.002)	1.553*** (0.050)
Community Characteristics		
Religion (1 if the household head is Hindu)	0.034*** (0.007)	0.424*** (0.059)
Caste (1 if the household belongs to Schedule Caste)	-0.011*** (0.002)	-0.111*** (0.022)
Caste (1 if the household belongs to Schedule Tribe)	-0.036*** (0.002)	-0.103*** (0.016)
Caste (1 if the household belongs to Other Backward Classes)	-0.013*** (0.003)	-0.063** (0.028)
Village Characteristics		
Distance to educational facility (primary school)	0.000 (0.000)	0.005 (0.004)
Distance to educational facility (middle school)	0.001 (0.000)	-0.005* (0.003)
Distance to educational facility (secondary school)	-0.001*** (0.000)	-0.006*** (0.002)
Distance to educational facility (higher secondary	0.000 (0.000)	-0.002 (0.002)
Distance to educational facility (college)	0.000 (0.000)	-0.002 (0.001)
Distance to nearest town (km)	0.000 (0.000)	0.000 (0.000)
Distance to district headquarters (km)	0.000*** (0.000)	0.000 (0.000)
Distance to nearest railway station (km)	0.000 (0.000)	0.002** (0.001)
Distance to available transport service (km)	0.000** (0.000)	-0.001** (0.000)
Distance from all-weather road (km)	0.000 (0.000)	0.001 (0.001)
Distance to other facility (post office)	-0.001*** (0.000)	-0.002 (0.001)
Distance to other facility (telegraph office)	0.000*** (0.000)	0.005*** (0.001)
Distance to other facility (std booth)	0.000 (0.000)	-0.003** (0.001)
Distance to other facility (bank)	0.000 (0.000)	-0.002 (0.004)

Table 5 continued…	**M.E (SE)**	**Coef. (SE)**
irdp beneficiaries	0.000***	0.000
	(0.000)	(0.000)
nrep beneficiaries	0.000	0.001
	(0.000)	(0.002)
trysem beneficiaries	0.000	0.002
	(0.000)	(0.002)
egs beneficiaries	0.000**	-0.002**
	(0.000)	(0.001)
dwacra beneficiaries	0.000	0.001
	(0.000)	(0.000)
iay beneficiaries	0.000	0.001*
	(0.000)	(0.000)
sgny beneficiaries	0.000	-0.004***
	(0.000)	(0.001)
Regional Characteristics		
State (1 if the state is Andhra Pradesh)	0.003	0.723***
	(0.005)	(0.040)
State (1 if the state is Bihar)	-0.053***	0.034
	(0.010)	(0.045)
State (1 if the state is Gujarat)	0.004	0.083**
	(0.007)	(0.043)
State (1 if the state is Haryana)	0.016***	0.006
	(0.004)	(0.092)
State (1 if the state is Himachal Pradesh)	0.050***	0.432***
	(0.004)	(0.042)
State (1 if the state is Karnataka)	0.016***	0.978***
	(0.005)	(0.044)
State (1 if the state is Kerala)	0.051***	1.421***
	(0.004)	(0.076)
State (1 if the state is Madhya Pradesh)	0.008	-0.226***
	(0.005)	(0.049)
State (1 if the state is Maharastra)	0.039***	0.770***
	(0.001)	(0.046)
State (1 if the state is Orissa)	0.023***	0.492***
	(0.004)	(0.050)
State (1 if the state is Punjab)	-0.011	0.133**
	(0.009)	(0.068)
State (1 if the state is Rajasthan)	-0.021***	0.030
	(0.005)	(0.066)
State (1 if the state is Tamil Nadu)	0.043***	0.913***
	(0.003)	(0.069)
State (1 if the state is West Bengal)	0.018**	-0.739***
	(0.009)	(0.124)
_cons		-11.809***
		(0.417)
rho		-0.181***
		(0.012)

*** indicates significant at 1% level, ** indicates significant at 5% level, * indicates significant at 10% level