Does Access to Secondary Education affect Primary Schooling?: Evidence from India*

Abhiroop Mukhopadhyay† and Soham Sahoo‡

27th November, 2011

Abstract

This paper investigates if better access to Secondary Education increases enrollment in Primary Schools among children aged 6 to 10. Using the 64th round Education survey of the National Sample Survey (2007-2008), this paper finds support for this hypothesis. Controlling for distance to primary schools, we find that if secondary schools are more than 5 kilometer away, there is a 2.5 percentage point decrease in share of primary enrolled children. To counter problems that emanate from using cross sectional data, this paper gives further evidence using a longitudinal survey of households in 43 villages in Uttar Pradesh, first surveyed as a part of LSMS by the World Bank. We find that a kilometer decrease in distance to the nearest secondary school increases the share of primary enrolled children by 4.6 percentage points. We find that the results do not change even after we account for endogeneity of school placement and measurement error. This result gives support to the assertion that parents find it worthwhile to educate their children only if they can reach the highest level, where the returns may justify the costs of a prolonged period of schooling. If the costs of post primary schooling are too high, as they would be if secondary schools are far away, parents have lesser interest in their children's education even at the primary stage.

*The authors wish to thank seminar participants at the Institute of Economic Growth (Delhi) for their comments. This research was funded by a research grant from the Planning and Policy Research Unit, Indian Statistical Institute (Delhi). Abhiroop Mukhopadhyay conducted research on this paper when he was Sir Ratan Tata Senior Fellow (2010-2011) at the Institute of Economic Growth (Delhi).

†Assistant Professor, Indian Statistical Institute (Delhi) and Sir Ratan Tata Senior Fellow (2010-2011), Institute of Economic Growth (Delhi)

‡Ph.D Candidate, Indian Statistical Institute (Delhi).

1 Introduction

The second Millenium Development Goal of the United Nations aims at universalization of primary education by 2015. Despite best attempts and significant rise in enrollment in most of the developing countries, recent findings suggest that the target is unlikely to be met[1]. While access to primary schooling has improved substantially, there are still very high drop out rates which has posed to be a crucial problem[2]. In India, the major public policy intiatives like *Sarva Shiksha Abhiyan* (Education for All), provision of midday meal, free textbooks and uniforms etc have aimed to universalize elementary education and reduce disparity across regions, gender and social groups. Two aspects have been mainly pointed out and emphasized in debates: to reduce the access barrier through provision of community based primary schools and to improve quality of schools in terms of physical infrastructure, teacher quality etc. In this paper, we raise a third issue: does possibility of continuation into higher levels of schooling affect primary schooling outcomes? In particular, we seek to investigate the effect of access to secondary education on primary school participation.

The decision of investment in human capital of children is crucially linked to the perceived economic returns to education (Manski, 1996; Nguyen, 2008; Jensen, 2010). The received wisdom from studies conducted in the last two decades is that the returns to education are concave: i.e., the marginal effect of an increment in years of schooling at the primary level is larger than the effect at higher levels (Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004). However, several studies in recent times show that the private returns to each extra year of education, in fact, rises with the level of education. A very recent paper by Colclough et al (2010) refer to several studies on different countries that find this changing pattern of returns to education. Schultz (2004) finds that private returns in six African countries are highest at the secondary and post-secondary levels. Kingdon et al (2008) find a convex shape of the education-income relationship in eleven countries. Studies specific to India has also

[1] A Fact Sheet published by the United Nations in 2010 reveals that although enrollment in primary education in developing regions has increased from 83% in 2000 to 89% in 2008, yet this pace of progress is insufficient to meet the target by 2015. About 69 million school age children were out of school in 2008.

[2] Statistics published by the United Nations show that in India, proportion of pupils starting grade 1 who reach last grade of primary was only 68.5% in 2006 (http://unstats.un.org/unsd/mdg/SeriesDetail.aspx?srid=591). The report on elementary education in India published by the District Information System for Education (DISE) shows that the average dropout rate in primary level (grade 1 to 5) between 2006-07 and 2007-08 has been 9.4%.

shown similar results (Kingdon, 1998; Duraisamy, 2002; Kingdon and Unni, 2001). At the same time, some other studies show that while the actual rate of return to primary schooling is quite high, parents tend to believe that the first few years of schooling have very low returns as compared to the later years (Banerjee and Duflo, 2005, 2011). These observations suggest that households may perceive education investment as lumpy; that significant economic returns require continuation to at least high school, which, in turn, gives rise to the possibility that households may find it worthwhile to educate their children only if they can reach that level. In light of this, better access to a post-primary school represents reductions in the cost of post-primary schooling and increases the possibility of continuation into higher levels of schooling. In doing so, it can become an important determinant of primary school participation.

The objective of this paper is to empirically test the hypothesis that access to secondary schools affect primary schooling outcomes. Although the importance of access to schooling in determining the educational outcomes has been well recognized in the literature (Duflo, 2001, 2004; Glick and Sahn, 2006), most of them are on access to primary school and its effect on the children in the primary school going age. Therefore a major supply side intervention for the policy makers has been to increase availability of primary schools to the community in order to encourage more school participation by the children. One commonly used measure of access is distance to school. Different studies have found that a reduction in the distance to school improves enrollment, reduces dropout and improves test scores (Lavy, 1996; Bommier and Lambart, 2000; Brown and Park, 2002; Handa, 2002; Burde and Linden, 2010).

Most of the literature addressing access examine the linkage between schooling outcomes at a particular level with the access to that level of schooling. Therefore, they have a static view of concentrating solely on the current cost of schooling. However, difficult access to post-primary schooling which reflects the future cost of schooling, hinders the possibility of continuation to higher level of education and hence it can adversely affect the schooling decisions even at the primary level. There are only a few studies which acknowledge the importance of access to post-primary schooling in determining the outcomes at primary level. Lavy (1996) uses a cross-section data on rural Ghana and shows that distance to post-primary schools has a negative effect on primary school enrollment. He suggests that the effective fees for post-primary schooling should be reduced in order to induce more participation at the primary level. Results similar to Lavy (1996) have been found in studies by Burke and Beegle (2004)

on Tanzania, by Vuri (2008) on Ghana and Guatemala and by Lincove (2009) on Nigeria. Hazarika (2001) uses a cross-section data on rural Pakistan and finds no impact of access to post-primary school on primary school enrollment of girls. His study indicates that if gains from post-primary schooling is low, which was indeed the case for girls in Pakistan, then access to it has no effect on primary school outcomes. Almost all of these studies base their analysis on cross-sectional data, hence they are unable to control for time-invariant unobserved heterogeneity at the household level. Besides, they have large number of villages but small number of households in each village, so they can not control for village level fixed effects. Moreover, apart from Lavy (1996), there is no discussion on endogeneous placement of schools which can be a potential source of bias in the estimates.

In this paper we use a nationally representative data on rural India to find support for the hypothesis that better access to secondary education increases enrollment and attendance among the children in primary school going age group. Controlling for distance to primary schools, we find that if secondary schools are more than 5 kilometers away, there is a 2.5 percentage point decrease in share of primary enrolled children[3]. To counter problems that may arise from using cross-sectional data, this paper gives further evidence from a household level longitudinal survey on 43 villages in Uttar Pradesh. We find that one kilometer decrease in the distance to nearest secondary school increases the share of primary enrolled children by 4.6 percentage points on an average. The results do not change even after we account for endogeneous placement of secondary schools and measurement error issues. The effect is more prominent for girls and for poorer households.

This study contributes to the existing literature by extensively examining how developments at higher levels of education influence decisions at much junior levels. This is especially relevant in a developing country where access to higher education is not universal. This paper indicates a robust causal relationship between access to secondary level and participation at the primary level [4]. The analysis presented in this paper suggests that goals of universal primary education cannot be achieved unless access to higher levels of schooling are made more accesible.

[3] The effect is similar (at 2.6 percentage points) on attendance.

[4] Our study relates very well to the recent policy environment in India, where the Ministry of Human Resource Development has developed a framework for universalization of access to and improvement of quality at the secondary stage. Following *Sarva Shiksha Abhiyan*, setting up of a national mission on *Rashtriya Madhyamik Shiksha Abhiyan* (RMSA), or universalization of secondary education, is in process.

The paper is organized as follows: Section 2 describes the two data sets we have used for analysis. In section 3, we explain the empirical methodology. Section 4 interprets the results obtained from regressions and Section 5 shows that the results are robust. In Section 6 we compare the results across different groups based on age, caste, gender and landholding. In the final section we discuss the conclusions of the paper.

2 Data

To test our hypothesis, we provide evidence using data from two surveys: the first one is a nation-wide survey on education conducted by the National Sample Survey Organization (NSSO), India; the second survey is a longitudinal follow up of households first surveyed as a part of LSMS in Uttar Pradesh (also called Survey of Living Conditions: SLC). We discuss the two data sets briefly and provide some description on key variables of interest in this section.

The National Sample Survey Organization (NSSO) conducts nationally representative household surveys on "participation and expenditure in education" once in ten years. We use the most recent education survey (64th round) by NSSO conducted in 2007-08. Detailed information about the schooling status of each child in a household are reported in the dataset. The dataset also contains standard socio-economic characteristics of the household. Crucial to the paper, it contains information on another key variable of interest: distance to the nearest primary, upper-primary and secondary school for each household. This information allows us to examine the relationship between distance to the nearest secondary school and primary school participation. We focus on rural India for our analysis.

The objective of our study is to look at whether (and how) participation at the primary level is affected by access to secondary education. We conduct our analysis using two indicators: the share of children in a household enrolled in school (*enrol*) and the share of children in a household attending school (*attend*). We restrict our analysis to children aged 6 to 10 years, who should be in class 1 to 5: the primary level. The average enrollment share, based on our measure is 89.7% and the average attendance share is 89.5%. [5] The nearest primary school is at a distance less than one kilometer for 91.88% of the households, and is more than five

[5] The estimated enrollment rate for these children at the all India level (rural) is 89%; (boys : 90.36%, girls :87.42%). Attendance rates do not differ much from the enrollment figures: overall it is 88.76% (boys: 90.18%, girls: 87.12%)

kilometers for only 0.13% of them. This indicates that a majority of the population have easy access to primary schools. Considering upper-primary or middle level, the nearest school is located within one kilometer for 61.7% of the households, and it is more than five kilometers for only 3.08% of the households. However, the access to secondary schooling is a lot more heterogenous. While 30.79% of the households have a secondary school within one kilometer, for 17.08% of the households the nearest secondary school is located at a distance of more than five kilometers.

Unlike in 64th round, the earlier NSSO education surveys did not have information on distance variables for post-primary schools. Therefore, only a static cross-sectional analysis is possible using the NSS data. Hence we extend our analysis by exploring a richer data set from a primary survey, and exploit the longitudinal nature of this data to establish more rigorously the causal connection between access to secondary schooling and primary school participation. This is a two-period panel data on 780 rural households in 43 villages from 9 districts in eastern and southern Uttar Pradesh. This is typically considered one of the more backward regions in the country.

As pointed out above, the baseline data is from the Survey of Living Conditions (SLC) undertaken in 1997-98 by the Living Standards Measurement Study (LSMS), World Bank. The second round of data was collected in 2007-08 by the authors and funded by the University of Oxford and the World Bank. [6] The survey comprises of a village questionnaire that contains information on, among other things, access to the nearest primary and post-primary schools; and a household questionnaire that has detailed information on various aspects of standard of living, including the schooling status of each child in the household.

According to the estimates from the SLC dataset, while *enrol* was 69.63% in 1997-98, it has increased to 82.1% by 2007-08. SLC has much detailed information on attendance. It reports the number of days in the last 7 days that the child attended school. [7] We define *attend* as proportion of children in the household who have attended at least 3 days in the last week. Based on our definition, while the attendance share was 64.57% in 1997-98, it had risen

[6] Both surveys are conducted at the same time of the year: December to April.

[7] Consistent with the base-line surveys, holidays and unusual attendances because of family events are factored in while asking this question. For this small sub-sample, attendance on the last normal week is asked. While such self reported data are not perfect, we are constrained, for the sake of uniformity, to not change questions.

to 77.28% by 2007-08. [8]

We use the distance to school variables from the village survey data. Distance to the nearest primary, middle and secondary schools are measured from the village center and reported in kilometers. The average distance to the nearest primary school has reduced from 0.7 kilometer to 0.1 kilometer. This is equally true for middle schools: the distance has reduced from 2.5 kilometers to 1.4 kilometers. However, this reflects that, on an average, both primary and middle schools were already very close to the villages in 1997-98 and therefore the change in them is modest. However, the change in the distance to the nearest secondary school has been the most spectacular. Figure 1 gives the distribution of distance to secondary schools. The average distance has reduced from 6.4 kilometers to 3.8 kilometers.

Pooling the data from both rounds, we find a significant and negative correlation of -0.17 between proportion of enrolled children at the primary level and distance to the nearest secondary school. However, this correlation coefficient does not take into account any other factor confounding factors that may have changed over time. Therefore in the next section, we specify econometric models and carry out multivariate analysis to look into the hypothesized causation more carefully.

3 Empirical Model

In this section, we lay out models to test whether access to secondary schools increases schooling enrolment and attendance at the primary level. For ease of presentation, we refer to only enrolment in the empirical model; however in testing, we test both enrolment and attendance. Our aim is two-folds: to begin with, we specify an empirical model apt for estimation using cross sectional data, available from the NSS. We then extend this model: by relaxing assumptions made for cross sectional analysis and by adding more variables that are available from a more detailed primary survey. Hence, we look at specifications that can be estimated using longitudinal data. To the extent possible, we try to be uniform across the two exercises.

For each household i living in geographical region j, let the proportion of children in the age group 6 to 10 who are enrolled in school be S_i. The choice of S_i as the dependent variable

[8] The corresponding enrolment rate for children in age group 6 to 10 years increased from 68.94% in 1997-98 to 82.46% in 2007-08. The rise in enrollment rate has been higher for girls (from 62.4% to 83.52%) as compared to boys (from 75.8% to 81.56%). Attendance rates show a similar trend.

(in contrast to, for example, a child level binary enrolment status) is again motivated by an attempt at uniformity. We elaborate on this later.

The co-variates for explaining enrolment can usually be categorized into four groups: individual child level factors: gender, age; household characteristics: wealth, household education, social group, religion; school characteristics: access to schools, quality of schooling; and geographical characterics: village/district characteristics. Since S_i is defined at the household level, we transform child level variables into their appropriate household counterparts: average age of children within the age band 6-10 (Age) and proportion of male children in the same age band ($Male$). In order to allow for non-linear impacts of age on enrolment, we include the square of age (Age^2).

Among, household level variables, we include land size to control for wealth ($land$: is taken as a dummy variable given the way its reported in the NSS data). The education level of decision makers in the household, especially that of the mother, has been found to have an important impact on educational outcome of children. In the absence of the identity of mothers in NSS data, we include dummy variables capturing whether the household head is literate (HH_Lit) and whether the household head is a woman (HH_Fem). To allow for differential impacts across different castes, we also include dummy variables that represent the social group the household belongs to (Soc_Group). Similary, we allow for different enrolment rates across different religious communities by including religion dummies; in particular we create a dummy variable for muslims ($Muslim$).

ecent literature on schooling has stressed both the access to schooling and its quality. While quality variables are lacking in the NSS survey, access to nearest schools are reported. In this paper, we would like to concentrate on secondary schooling, a level that yields perceptible market returns. While primary schools in the last decade are almost in every village, this is not true for post primary schools. We focus on secondary schooling which, in contrast to middle schools, are less likely to be susceptible to endogeneity. Secondary schools, because of the need for more infrastructure and relatively better trained teachers need a larger market size to be profitable; their placement depends on characteristics of a larger group of people than just a village. Hence, in our specification, we allow for access to the nearest primary ($PRIM$) and secondary schools (SEC). Again these variables are reported as distance bands (See $Appendix$); we consider appropriate dummy variables. Notice that while inclusion of $PRIM$

is very standard in primary schooling regressions, the inclusion of secondary levels of schools is less standard. We follow *Lavy* (1996) and include SEC to capture the idea that primary enrolments may be affected by the availability of higher levels of schooling. We hypothesize that, if after controlling for primary school access, access to the nearest secondary schools are found to be important, this would give credence to the hypothesis that the possibility of continuation plays a significant role in primary school enrolment. Indeed, if parents perceive that only returns to higher levels of schooling are worth the investment of sending children to school, then they would be unlikely to enrol their children in primary schools if secondary schools are far away.

We allow for geographical heterogeneity by running NSS state region fixed effects α_j[9]. In addition to allow for intra-cluster correlation among residents of a village, we cluster our standard errors at the village level[10]. Therefore in Model 1, we estimate the following specification:

$$S_{ij} = \alpha + \beta' I_{ij} + \gamma' H_{ij} + \sum \delta_{1k} PRIM_{jk} + \sum \delta_{2k} SEC_{jk} + \alpha_j + \epsilon_{ij} \quad (1)$$

where I refers to the individual characterstics that have been transformed to the household level variables, H refers to the household level socio-economic characterstics, k is an index which refers to the various distance categories. The hypothesis that we want to test is if $\delta_{2k} \geq 0$.

While results from estimation of Model 1 using NSS data have the obvious advantage that the conclusions reached are borne out of estimation of a country wide data set (and in that sense robust), there are some obvious disadvantages. Notice for example, that we cannot control for the quality of primary schools. Quality of primary schools have been found to be significant in papers that investigate schooling outcomes. If secondary schools are present in villages where the quality of primary schools are good, then our estimators for the impact of secondary schooling would be inconsistent. Moreover, given the small number of observations per village, it is infeasible to account for village level heterogeneity through village fixed effects. It is plausible then that the presence or nearness of secondary schools is confounded by unobserved village heterogeneity. The use of cross sectional data also omits the role of individual household level heterogeneity. While it is hard to argue that household level taste for schooling

[9] NSS divides states in relatively homogenous state regions. There are 87 such regions in India.

[10] We could have conducted a village level fixed effects regression but these would lead to a very large number of dummy variables and not many observations within each cluster.

affects distance to secondary schooling directly, in so far as such tastes may be geographically clustered, they may still be correlated with where secondary schools are located.

In order to address these concerns, we turn next to Model 2; where we take advantage of a larger number of variables available through the SLC longitudinal survey. While this survey has a smaller sample size, it follows households and villages over time. We wish to exploit these facets of the survey for a more robust analysis. In Model 2, we begin relaxing assumptions and adding variables. We conduct a pooled cross sectional analysis controlling for village level fixed effects.

In this specification, we control for some omitted variables that may bias our previous results. We include quality of the village primary school, an important omitted variable in Model 1. We include four quartiles of quality (D_l^{QUART}), constructed by principle components over various components of infrastructure. At the village level, we also control for the distance to nearest district headquarters ($Dist^{HQ}$). It is plausible that villages closer to district head quarters have a better preception about the returns to education. Another channel through which the distance to district head quarters may affect primary schooling is through quality of teachers that come to the nearby schools. It can be argued that villages near the district head quarters may have better qualified teachers; those who reside in district head quarters and commute to the village schools on a daily basis. We also allow for differential road access by defining dummy variables for the quality of roads (D_h^{ROAD} : See *Appendix*).

In the SLC, the identity of the mother is known. Hence we control for the education of the mothers by including the proportion of literate mothers (LIT_MOM)[11]. Another variable included in this specification is the proportion of adult household members that are engaged in off farm activities (OFF_FARM). Apart from a probable income effect through these activities, these activities may also need some level of education. Therefore we posit that greater exposure of the houshold to off farm jobs may inform households about the benefits of education.

Due to better data availability, the variables measuring the distance to nearest school are measured as continuous variables and we include squared terms to allow for non linearities (With some abuse of notation, let us now refer to the vector of the linear and the squared distance to the nearest relevant school as $PRIM$ and SEC). Thus in Model 2, we estimate:

[11] Consistent with our analysis, we look at mother of children aged 6 to 10.

$$
\begin{aligned}
S_{ijt} &= \alpha + \beta' I_{ijt} + \gamma' H_{ijt} + \ pi_1' PRIM_{jt} + \pi_2' SEC_{jt} + \sum \delta_l D_{ljt}^{QUART} \\
&\quad + \sum \theta_h D_{hjt}^{ROAD} + \gamma Dist^{HQ}{}_{jt} + \alpha_j + \sigma t + + \sum_{d \in \{Districts\}} \sigma_d t + \\
&\quad + \gamma Dist_{jt}^{HQ} * t + \epsilon_{ijt} \qquad (2)
\end{aligned}
$$

Here, j refers the village and t refers to 1997 and 2007. We include t as a regressor to eliminate time trends.

As pointed out above, if we fail to control for unobserved household heterogeneity, we may get spurious results. Apart from the possibility that there may be clustering of people with taste for education, there may also be instances where-in the composition of the village itself may have changed, with migration of households. While it is hard to argue that in rural India, education is a motive for migration, it is true that households enter and exit villages, especially given the time interval between our surveys. This may change the nature of the village over time. We eliminate these possibilities by running a household level fixed effects regression. In order to be more robust and to eliminate spurious trends, we allow for district level trends and a trend that depends on how far the village is from district headquarters. Model 3 therefore specifies:

$$
\begin{aligned}
S_{ijt} &= \alpha + \beta' I_{ijt} + \gamma' H_{ijt} + \pi_1' PRIM_{jt} + \pi_2' SEC_{jt} + \sum \delta_l D_{ljt}^{QUART} \\
&\quad + \sum \theta_h D_{hjt}^{ROAD} + \gamma Dist_{jt}^{HQ} + \alpha_i + \alpha_j + \sigma t + \sum_{d \in \{Districts\}} \sigma_d t \\
&\quad + \gamma Dist_{jt}^{HQ} * t + \epsilon_{ijt} \qquad (3)
\end{aligned}
$$

As before, we are interested in the sign and statistical significance of the coefficients in the vector π_2'. Apriori, if our hypothesis were true, one would expect a negative co-efficient for the linear term and a positive coefficient for the squared term, implying that marginal changes in distances closer to the village have a greater impact on primary school enrolment.

Even with individual and village fixed effects, one could contend that secondary schools come up endogenously. Since secondary schools need bigger markets, they may be more likely to come up in bigger villages. Indeed, our sample has villages that have grown over time. It could well be the case that some of the bigger villages were able to attract secondary schools.

However, controlling for this effect, such expansion should have no additional impact on the decisions to send children to school. We therefore posit that village population can be used as an instrument for distance to secondary schools. Since our specification contains both distance to secondary education and its square, we use village population and its square as instruments. We also provide evidence on exogeneity of our instruments.

In the empirical models stated above, we have assumed that a significant coefficient for secondary schools would provide evidence that secondary education access has impact on primary schooling enrollment. However, many secondary schools have classes 1 to 5. Since these schools are bigger and have better infrastructure, we could be capturing the fact that children go to nearby secondary schools for their primary education. Though this would be an interesting finding, we would like to separate the effect of better quality from the impact of better access to continuation. We do not have the quality of secondary schools in our sample. Hence we follow an indirect method. We define as our dependent variable the proportion of children enrolled for primary school within the village. Futher-more, we consider villages where there is no secondary school within 2 kms in both periods. Using the same set of independent variables as in Model 3, if our dependent variable is negatively correlated to the distance to secondary school for this sub-sample, then the effect can only be through the prospect of continuation.

4 Main Results

We provide two sets of results in each table. The first column corresponds to the proportion of children in the age group 6-10 who are enrolled in school; the second column corresponds to the proportion of children in the age group 6-10 who attend school. We begin by looking at the all India evidence on covariates of primary enrollment and attendance in $Table1$. Most results explaining enrollment are qualitatively similar to results using attendance. We discuss the results in the context of enrollment, pointing out, where relevant, results that differ across the two columns.

We find that age (Age) has the usual increasing and concave relationship with enrollment. Since some children start school late, at initial induction ages, as one raises the age, more children are found to be enrolled in school. However, after a certain age, they are more likely to drop out. The proportion of males among children ($Male$) is positive and significant, hinting

that the gap between boys and girls in primary schooling still exists.[12] Among household level variables, we find that the presence of a literate head increases S by 11 percentage points while the presence of a female head increases the proportion by 2.4 percentage points. Households with greater land ownership have higher enrollment pointing out that, in-spite of $free$ education, richer households are more likely to send children to school. Households with land ownership of 0.05 to 1 acre have 3 percentage points higher proportion of children enrolled than households with less than 0.05 acre of land (the reference category). Households with more than 1 acre of land have 5 percentage points higher proportion of children enrolled in school as compared to the reference category. A greater household size, reflecting a squeeze on household resources, causes a lower share of enrollment. Muslims households have 7 percentage points lesser children enrolled in households than other religious groups. Schedule tribes are the least likely to be enrolled in school, with the smallest S, followed by Schedule Caste who have 3.9 percentage points lesser S than the reference category (General caste).

Let us move next to the variables representing access to nearest school. We find that distance to primary schools is still an important covariate of enrollment and attendance. In-spite of schemes that target primary schooling, including massive investment in school infrastructure, this result is depressing in itself pointing out to the need for more primary schools. We find that if the primary school is between 1 to 5 km away, the proportion of children enrolled falls by 1.6 percentage points as compared to when primary schools are at less than 1 km distance. This proportion falls by 17 percentage points if primary schools are more than 5 km away. The affect on attendance, in this case, is even more with a 19.2 percentage fall as compared to the reference access category.

Next we come to what we focus in this paper: Distance to Secondary schools. We find that if secondary schools are at a distance of 1 to 5 km, they have no discernible differential impact on primary schooling as compared to when the school is within 1 km radius (the reference category). However, if secondary schools are more than 5 km away from the household, then this is associated with a 2.5 percent point lower share of enrolled children. This seems to indicate that there is a link between secondary schooling and primary education. However, as pointed out before, while encouraging, there can be many confounding factors that we cannot

[12] We do not discuss the magnitude of the marginal effects of the modified child level variables as they are difficult to interpret.

take care of while estimating Model 1 with the cross sectional data collected by the NSS. For example, does higher distance to schools indicate that the households live in far off distant places where they may not be aware of benefits of education. We have indicated many such arguments in the previous section. Suffice it to say, that the negative correlation between distance to secondary schools and enrollment and attendance is a strong indication that our hypothesis, if true holds for most of rural India.

To address the problem that we cannot control for many factors that may be confounding our results, we discuss, next, estimation results of Model 2. Since this is a different sample, we do not expect results to have same quantitative effect. However, if they are qualitatively same, especially in the context of the variable of interest, i.e. access to secondary schooling, one would get confidence in the all India results discussed above. Given very small distances of primary schooling from the villages, we omit the square of distance to primary schooling from this specification.

Results in $Table2$ show that, similar to our previous results, a literate head raises S by 14 percentage points, which is not very different, even quantitatively from the All India result. A female head raises the proportion of enrolled children by around 10 percentage points, the difference in the quantitative value reflecting more the dearth of female headed households in Uttar Pradesh. Perhaps more significant is the effect of (LIT_MOM), a variable not present in Model 1 specification. A one percentage point increase in proportion of mothers literate raises the proportion of enrolment by 6.9 percentage points. The marginal effect on attendance is even larger at 8.2 percentage points. Wealth, as measured by land holding has an equally strong impact with an increase in one standard deviation of land (5.4 acre) increasing S by 3.8 percentage points. The result on (OFF_FARM) are less robust. We find that while it has an impact on enrollment, it has no impact on attendance.

Since we control for village fixed effects by village dummies, it is not surprising that some of the village level variables, like quality of primary schools and road access turn out insignificant. However, even though insignificant, the variables have, in most cases, the right signs. The insignificance of caste and religion is because our sample has people predominantly from backward and Scheduled Castes. What is however heartening to see is that the results that correspond to distance to nearest secondary schooling still survive, in-spite of village dummies. The marginal effect, as hypothesized, is negative and decreasing in distance. Thus, a unit de-

crease in distance causes a larger effect on primary enrollment when its in close distance range, rather than when the school is far away. The partial effect on enrollment of a kilometer drop in distance to nearest secondary school is 0.034 (at the mean value of distance). The marginal impact of distance of primary school turns out to be insignificant. However, this is because almost all the villages in the sample have a primary school in the village. Therefore the effect of the marginal changes in distance to primary schooling is not discernible from these results.

Our next estimation exercise, Model 3, takes into account household level unobservables. The advantage of doing so has been discussed in the previous section. *Table*3 shows that the effect of distance to secondary schools is even stronger. The average partial effect of a km decrease in distance to nearest secondary school is 0.046. Everything else turns out insignificant, barring the second quartile of primary school quality (in the case of attendance) and road access variables, which both increase S. This insignificance is not surprising given that most households and villages don't change greatly over time. However, what has changed more dramatically is the reach of secondary schooling and it is precisely the change in this one aspect, rather than every aspect which makes this estimation exercise worthwhile.[13]

5 Robustness

In this section, we subject our specifications to robustness checks. First we want to find out whether any endogeneity in secondary school placements is driving our result. We address this possibility in two ways. First we use village population as an instrument. Second, we consider only small villages which are not large enough to be attractive to secondary schools.

5.1 Instrumental Variables

We make the crucial assumption that private secondary schools will locate near villages only if the market size is big. Hence the size of the village may be an important determinant of nearness to secondary schools. However, once we have controlled for the size of the village, the regressor has no additional power for predicting household level children's enrollment proportion. Due to the presence of a linear and a square term of distance to secondary schools, we

[13] A pure pooled OLS with no village level fixed effects yields many more significant variables. However, in so far as our results on access to secondary schooling survive through all specifications, we do not forsake the fixed effects estimation models.

have two endogenous variable. We therefore posit two instruments: the village population and its square.

Results in $Table4$ show that, in the case of attendance, the results are the same. The partial effect at the mean is now 0.27. In the case of enrollment, the linear term is not significant but has the same negative sign. Thus the negative impact of distance to schooling goes through we use instruments.

The appropriateness of instruments are hard to check given we have a just identified system. At the very least, we need to check whether, at the first stage regressions, the proposed instruments are correlated with the endogenous variable. An F test to check joint significance of the instruments variable point out that they are indeed significant in both first stage regressions (the over all F of the two first stage regressions are 12 and 20). A more stringent test is one where the null of weak instruments is checked using the Cragg-Donald F statistic. The Cragg-Donald F statistic are much above the critical cut offs for 20 percent size distortion and are close to the 15 percent cut offs. The main problem emanates from the collinearity between village population and its square. However, all specifications point out that the quadratic term is important; so it is not possible to drop one of the variables to get a more robust result.

5.2 Small Villages

In the face of collinearity problems, we propose another strategy to check if endogeneity problem plague our results. We consider small villages, where the number of households are less than 300. While it is possible that small villages have a secondary school close by, it is unlikely that the school opened only to service their demand. $Table5$ reports the results corresponding to only small villages. We find that impact of the secondary schools goes through. Because of collinearity, we drop the quadratic term in distance, but the linear term is negative and significant for both enrollment and attendance. A unit drop in distance to secondary schools raises enrollment rate by 0.034. We also present results with villages with less than 100 households for robustness. Results go through.

5.3 Measurement Error and Outliers

While we have made special efforts to verify distances to nearest schools, distances could still be reported with error[14]. To check whether measurement error in distances cause inconsistency, we estimate Model 3 with distance dummies instead of distances. We define distance dummies for distance to secondary schools analogous to the exercise done using NSS data. We find that the results are robust to smoothing of distance using dummies. If the nearest secondary school is at a distance of 1 to 5 km then it reduces enrollment by 63 percentage points as compared to when it is within 1 km. The impact is around -0.49 for distances beyond 5 km (relative to there being a secondary school within 1 km of the village)[15].

In small data-sets, large outliers often tend to drive results. In order to remove the impact of outliers, based on Model 3, we remove observations that have residuals greater than one standard deviation of the estimated residuals. Results stay robust to removal of these observations.

5.4 Secondary Schools or Secondary Education

In this paper, we would like to check the hypothesis that the possibility of completing secondary education affects the decision to enrol and attend school at the primary level. So far, we have shown that access to secondary schools affect primary school enrollment and attendance. But, as pointed out earlier, many secondary schools have classes 1 to 5. Secondary schools are usually better quality, at least in terms of infrastructure, than stand-alone primary schools. It may be that children go to a secondary schools as they open up nearer. In this case, our results would give further evidence, indirectly, to the importance of better quality. There may still be the incentive due to possibility of continuation; however it is not possible in this case to unbundle the two effects.

In order to extricate the impact of secondary education, we follow another strategy. We define another dependent variable: the proportion of household children aged 6 to 10, who go to a school within the village. Further, we use a sub-sample of villages that do not have

[14] We have verified distances using a village survey and by putting the same question to households. In most cases, the distances reported in the village survey are the same as the modal value of the distances from the household surveys for the village.

[15] While the coefficient falls in magnitude beyond 5 km, statistical testing of the coefficients of these two dummies reveals that one cannot reject the hypothesis that they are of the same magnitude.

a secondary school within 2 kms in both periods.[16] In this scenario, if larger proportion of children go to schools within the village, this cannot be the outcome of their going to a secondary school, because there are none in the village (or even within a 2 km radius of the village).[17] In this case, the impact of lower distance to secondary schools has to because of the perception that students can continue to secondary levels. This is, then, the impact of secondary education rather than secondary schools.

Results in *Table*7 show that this is indeed the case. The marginal impact of a reduction in distance to secondary school turns out to be 0.97 (the mean distance to nearest secondary schools is 6.2 for such villages) Thus, for villages which do not have secondary schools, there has indeed been a considerable impact of secondary schools opening up near the village. And given that we only look at children studying inside the village, this effect can only be due to an increased awareness of the possibility to get secondary education.

5.5 Other Pathways

It is possible that improvement in primary school participation is because of the fact that now these children in primary school going age group have elder siblings who are better educated due to better access to secondary schooling. The elder siblings may be teaching their younger brothers and sisters and motivating them to attend primary school. If this is true, then it establishes another pathway through which access to secondary school may be important for primary schooling. However, our objective is to check if, despite its effect through this channel, nearness of secondary school has any direct impact on primary schooling decision. Therefore, we run the regression specified in Model 3, introducing a new household level covariate which captures the number of children in the age group of 14 to 18 years who are enrolled in secondary or higher level of schooling. The results show that even after we control for the elder sibling effect through this new variable, the distance to secondary school and its square term both are still significant, even their magnitudes remain almost unchanged. We accept the possibility that decisions regarding primary and secondary schooling of children can be determined simultaneously in a household, and in that sense this new covariate may not be purely exogeneous. However, our objective is not to interpret the result in terms of the sibling

[16] This selection is based on distance, an independent variable. There-fore our estimators are still consistent.

[17] Since this information is based on self reported responses and households are not always sure about the boundary of villages, we choose the 2 km radius to reduce chances of error.

effect, rather to show that our main results remain unperturbed even if we account for this pathway.

5.6 Child Level Analysis

So far the results obtained from household level analysis have been supportive of our hypothesis that access to secondary schooling do play a significant role in determining primary school participation. To further examine the robustness of these results, we conduct child level analysis. Hence the unit of observation becomes specific to a child instead of a household. Thus, the dependent variable becomes binary enrollment or attendance, and similarly the independent variables are also at the child level now. We follow exactly the same specification as in Model 2, and run a linear probability model[18]. As reported in Table9, the results show a significant negative coefficient of distance to secondary school variable, and a significant positive coefficient of its square term. The marginal effect of one kilometer reduction in distance to secondary school is 0.021 (at average distance) on enrollment and 0.023 on attendance.

6 Group-wise Effect of Secondary Education

In this section we consider different groups and investigate if the impact of secondary schooling differs across them. We can however only do limited exercises, as sub-samples have very small sizes and while our coefficients of interest retain their sign, the models have too few degrees of freedom to be reasonable.[19]. We conduct the following four exercises:

6.1 Age Group Decomposition

We decompose the primary school going age group into two parts: a younger 6 to 7 years and an older 8 to 10 years group. The objective is to see whether the effect of secondary schooling has been more for the younger children or the older children. In order to conduct fixed effects analysis at the household level (as in Model 3) for each of these age groups, we transform the child level variables into their household level counterpart, considering the relevant age group.

[18] The marginal effects are very similar even if we run a probit model with this specification (results are not presented here).

[19] for example, often our statistical software STATA refuses to report F statistics in order to not mislead us

Results from Table 10 show that access to secondary schooling affects both the younger and the older children significantly. However, the marginal effect (evaluated at the sample mean) of distance to secondary school on enrollment is higher for younger children (-0.08) as compared to older children (-0.031). These results are obtained from fixed effects regressions which effectively estimate how the change in enrollment between two time periods is associated with the change in distance to secondary school. If households take time to respond to positive changes in the economy, then probably opening of new secondary schools was too recent for the older children to react to it. Whereas this is not the case for the younger children and hence we find that better access to secondary schooling has improved enrollment of the younger children more than the older children.

6.2 Economically Backward Communities

In this section we inspect the effect of access to secondary schooling on primary school participation for only the economically backward communities. The backward communities are consisted of Scheduled Castes, Scheduled Tribes and Other Backward Castes in rural Uttar Pradesh. So we separate out the sample of households belonging to these backward communities and estimate Model 3 using this data. Similar and significant result is found again. The average partial effect of a unit reduction in the distance to nearest secondary school on enrollment of the primary school going children in the backward communities is found to be 0.064.

6.3 Gender Decomposition

We run the regression following Model 3 separately for boys and girls by transforming the dependent variable into the proportion of enrolled boys (or girls) aged 6 to 10 in a household. Results from Table 12 show that the effect is much higher for girls as compared to boys. If secondary schools are far away, the access cost is higher for girls than boys. Therefore the effect of a change in distance has greater impact on girls than boys.

6.4 Landholding

It is likely that the a change in access cost will have differential effects on poor households as compared to the rich ones. To check if this is indeed true, we carry out the regression separately

for households whose landholding in 1997 was below the median level, and the ones above the median level. The results suggest that while we get significant effects for both the groups, the magnitude is higher for households below the median level of landholding, indicating that a reduction in the distance to secondary school will benefit the poorer households more in terms of primary schooling outcomes of the children.

7 Conclusion

Universal primary education has been a stated aim of development policy experts as well as of governments. Policies to improve outcome for primary education have largely focussed on access to primary schools. In the recent years, this emphasis has moved to quality of education: with efforts being made to improve the quality of teachers. However, a key component that drives the decision of households to send children to school is the economic returns to schooling. This paper builds on recent work in the economics of education literature that shows that the perceived (in many cases real) returns to education are convex. We posit that if this is true, it is plausible that education investment is lumpy: that to elicit profitable returns from education, households have to invest in their child's education till they graduate out of high school. In this context, in making decisions at even the primary level, household take into account costs of post primary schooling. A major component of the cost of post primary schooling is distance to secondary schools. In this paper, we explore whether access to secondary schools affects primary schooling.

Using a nationally representative cross sectional data for India, we show that distance to secondary school does affect primary schooling. Controlling for distance to primary schools, we find that if secondary schools are more than 5 kilometer away, there is a 2.5 percentage point decrease in share of primary enrolled children. However, such analyses are unable to control for unobserved heterogeneity at the village and household level and are subject to problems of endogeneity. Therefore in order to rule out confounding factors, we estimate the relation using a panel data set on households from 43 villages in Uttar Pradesh, a state in India, where primary schooling is far from universal. We find that the distance to the nearest secondary schools is a significant determinant of primary school enrollment and attendance. In order to test this hypothesis, we run a large number of robustness checks that include instrumenting for the endogenous placement of schools. We find that our results are robust. Further, we find

that this impact is larger for younger children, for girls and for households with low ownership of land in the baseline survey.

In light of these results, our paper suggests that access to post primary schools are important for meeting primary schooling objectives. While on the one hand it can be argued that secondary schools will open up privately as soon as enough children are primary educated, in many parts of the developing world, this critical mass of primary educated may never be met because they do not see themselves going on to complete a economically meaningful level of schooling. In such scenarios, it may be worthwhile to keep in mind that all levels of schooling need to be developed and accessible at the same time to achieve universal education.

References

[1] Banerjee, Abhijit and Esther Duflo (2005), "Growth Theory through the Lens of Development Economics", Handbook of Economic Growth, 1A

[2] Banerjee, Abhijeet and Esther Duflo (2011), Poor Economics

[3] Bommier, Antoine and Sylvie Lambart (2000), "Education Demand and Age at School Enrolment in Tanzania", The Journal of Human Resources, XXXV(1)

[4] Brown, Philip H. and Albert Park (2002), "Education and poverty in rural China," Economics of Education Review, 21(6)

[5] Burde, Dana and Leigh L. Linden (2010), "The Effect of Village-Based Schools: Evidence from a RCT in Afghanistan", Working Paper

[6] Burke, Kathleen and Kathleen Beegle (2004), "Why Children Aren't Attending School: The Case of Northwestern Tanzania", Journal of African Economics, 13(2)

[7] Colclough, C.; G. Kingdon and H. Patrinos (2010). "The Changing Pattern of Wage Returns to Education and its Implications", Development Policy Review, 28(6)

[8] Duflo, Esther (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", American Economic Review, 91(4)

[9] Duraisamy, P. (2002), "hanges in Returns to Education in India, 1983-94: By Gender, Age-cohort and Location", Economics of Education Review, 21(6)

[10] Filmer, Deon (2007), "If You Build It, Will They Come? School Availability and School Enrolment in 21 Poor Countries", Journal of Development Studies, 43(5)

[11] Glewwe, Paul (2002), "Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes", Journal of Economic Literature, XL

[12] Glick, Peter and David E. Sahn (2006), "The Demand for Primary Schooling in Madagascar: Price, Quality, and the Choice between Public and Private Providers", Journal of Development Economics, 79

[13] Handa, Sudhanshu (2002), "Raising Primary School Enrolment in Developing Countries: The Relative Importance of Supply and Demand", Journal of Development Economics, 69

[14] Hazarika, Gautam (2001), "The Sensitivity of Primary School Enrollment to the Costs of Post-Primary Schooling in Rural Pakistan: A Gender Perspective", Education Economics, 9(3)

[15] Jensen, Robert (2010), "The (Perceived) Returns to Education and the Demand for Schooling", The Quarterly Journal of Economics

[16] Kingdon, Geeta Gandhi and Jeemol Unni (2001), "Education and Women's Labour Market Outcomes in India", Education Economics, 9(2)

[17] Kingdon, Geeta Gandhi (2007), "The Progress of School Education in India", GPRG Working Paper 071

[18] Kingdon, G., H. Patrinos, C. Sakellariou and M. Soderbom (2008), "International Pattern of Returns to Education", mimeo

[19] Lavy, Victor (1996), "School Supply Constraints and Children's Educational Outcomes in Rural Ghana", Journal of Development Economics, 51

[20] Lincove, Jane Arnold (2009), "Determinants of Schooling for Boys and Girls in Nigeria under a Policy of Free Primary Education", Economics of Education Review, 29

[21] Manski, Charles F. (1993), "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?", Studies of Supply and Demand in Higher Education, University of Chicago Press

[22] Nguyen, Trang (2008), "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar", mimeo, MIT

[23] Orazem, Peter F. and Elizabeth M. King (2007), "Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy", Handbook of Development Economics

[24] Psacharopoulos, G. (1994), "Returns to Investment in Education: a Global Update", World Development, 22(9)

[25] Psacharopoulos, G. and H. Patrinos (2004), "Returns to Investment in Education: A Further Update", Education Economics, 12(2)

[26] Schultz, T. P. (2004), "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education", Journal of African Economies, 13(2)

[27] Vuri, D. (2008), "The Effect of Availability and Distance to School on Children's Time Allocation in Ghana and Guatemala", Working Paper

Appendix

Table A1: Descriptive Statistics from NSS 64th Round Education Survey Data (Household Level Variables)

Variable	Observations	Mean	Std. Dev.	Min	Max
Enrol	22410	0.88	0.31	0	1
Attend	22410	0.88	0.31	0	1
PRIM (1km ≤ d < 5km)	63019	0.09	0.28	0	1
PRIM (d ≥ 5km)	63019	0.00	0.04	0	1
SEC (1km ≤ d < 5km)	62996	0.52	0.50	0	1
SEC (d ≥ 5km)	62996	0.18	0.38	0	1
Male	22410	0.54	0.45	0	1
Age	22410	8.04	1.23	6	10
HH_Lit	63318	0.59	0.49	0	1
HH_Fem	63318	0.11	0.32	0	1
Household size	63318	4.58	2.19	1	30
Land (0.05 - 1 acre)	63282	0.31	0.46	0	1
Land (more than 1 acre)	63282	0.37	0.48	0	1
Muslim	63318	0.10	0.30	0	1
Soc_Group = ST	63318	0.17	0.38	0	1
Soc_Group = SC	63318	0.20	0.40	0	1
Soc_Group = OBC	63318	0.39	0.49	0	1

Table A2: Descriptive Statistics from SLC Data (Household Level Variables)

	1997-98					2007-08				
Variable	Observations	Mean	Std. Dev.	Min	Max	Observations	Mean	Std. Dev.	Min	Max
Enrol	441	0.6963719	0.42	0	1	356	0.8210674	0.36	0	1
Attend	441	0.6457834	0.44	0	1	356	0.7728464	0.39	0	1
PRIM	441	0.7414966	0.98	0	5	356	0.1067416	0.49	0	3
SEC	441	6.399093	5.09	0.5	20	356	3.780197	3.26	0	16
Male	441	0.481719	0.41	0	1	356	0.5543539	0.43	0	1
Age	441	8.042328	1.08	6	10	356	8.076639	1.12	6	10
LIT_MOM	441	0.170338	0.37	0	1	356	0.2153558	0.40	0	1
HH_Lit	440	0.4363636	0.50	0	1	334	0.4221557	0.49	0	1
HH_Fem	441	0.0430839	0.20	0	1	356	0.0730337	0.26	0	1
Household size	441	8.013605	3.82	3	29	356	8.036517	2.95	3	19
Land	387	3.579044	6.82	0	93	356	1.643817	2.95	0	33
OFF_FARM	420	44.81905	28.42	2	95	356	29.23876	22.48	2	90
D^{HQ}	441	33.78458	17.44	6	75	356	32.12921	16.42	6	75
Caste (SC/ST)	441	0.2789116	0.45	0	1	356	0.2837079	0.45	0	1
Caste (Backward)	441	0.5464853	0.50	0	1	356	0.5758427	0.49	0	1
Muslim	441	0.0453515	0.21	0	1	356	0.0589888	0.24	0	1
D_2^{QUART}	441	0.4217687	0.49	0	1	356	0.1123596	0.32	0	1
D_3^{QUART}	441	0.1587302	0.37	0	1	356	0.3820225	0.49	0	1
D_4^{QUART}	441	0.0113379	0.11	0	1	356	0.4550562	0.50	0	1
D_2^{ROAD} (Katcha)	441	0.2834467	0.45	0	1	356	0.0870787	0.28	0	1
D_3^{ROAD} (Paved)	441	0.2970522	0.46	0	1	356	0.2893258	0.45	0	1
D_4^{ROAD} (Pucca)	441	0.30839	0.46	0	1	356	0.5140449	0.50	0	1

Figure 1: Distance to Nearest School (SLC Data)

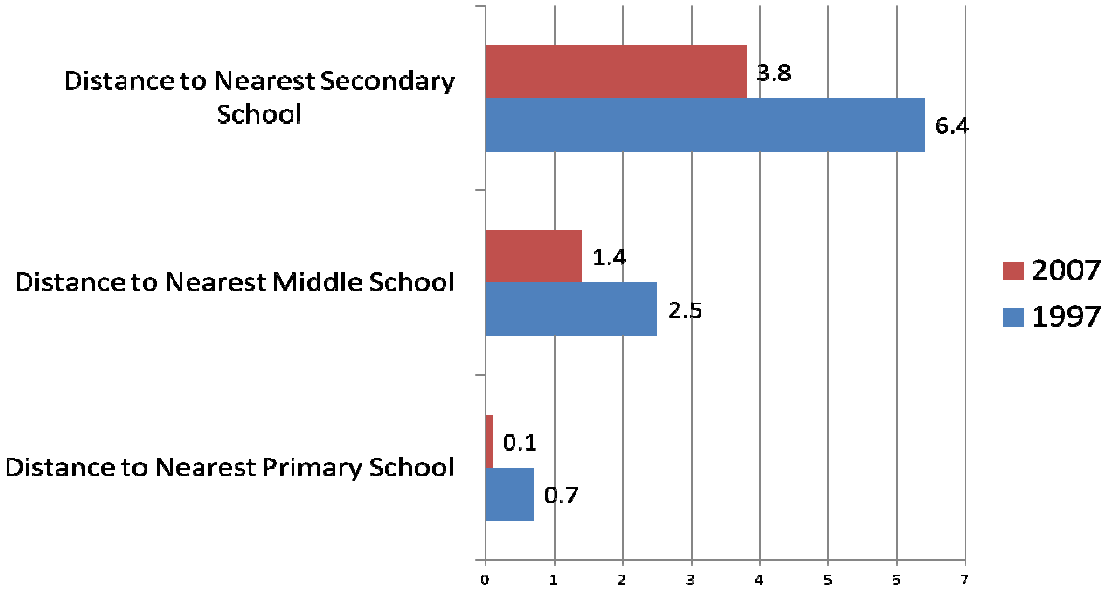

Table 1: Effect of distance to nearest secondary school on primary school enrollment and attendance (NSS Data)

	Enrollment	Attendance
PRIM (1km ≤ d < 5km)	-0.016	-0.016
	[0.067]*	[0.064]*
PRIM (d ≥ 5km)	-0.177	-0.192
	[0.080]*	[0.054]*
SEC (1km ≤ d < 5km)	0.001	0.001
	[0.816]	[0.872]
SEC (d ≥ 5km)	-0.025	-0.026
	[0.001]***	[0.000]***
Male	0.037	0.038
	[0.000]***	[0.000]***
Age	0.202	0.204
	[0.000]***	[0.000]***
Age2	-0.012	-0.012
	[0.000]***	[0.000]***
HH_Lit	0.113	0.113
	[0.000]***	[0.000]***
HH_Fem	0.024	0.025
	[0.002]***	[0.001]***
Household size	-0.002	-0.002
	[0.029]**	[0.019]**
Land (0.05 - 1 acre)	0.032	0.034
	[0.000]***	[0.000]***
Land (more than 1 acre)	0.057	0.06
	[0.000]***	[0.000]***
Muslim	-0.073	-0.072
	[0.000]***	[0.000]***
Soc_Group = ST	-0.067	-0.067
	[0.000]***	[0.000]***
Soc_Group = SC	-0.039	-0.039
	[0.000]***	[0.000]***
Soc_Group = OBC	-0.02	-0.021
	[0.000]***	[0.000]***
State Region Fixed Effects	Yes	Yes
Constant	-0.028	-0.036
	[0.771]	[0.707]
Observations	22288	22288
R-squared	0.122	0.122

Robust p values in brackets

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

Table 2: Effect of distance to nearest secondary school on primary school enrollment and attendance (Pooled Regression with SLC Data)

	Enrollment	Attendance
PRIM	0.016	-0.008
	[0.750]	[0.875]
SEC	-0.075	-0.055
	[0.000]***	[0.010]**
SEC^2	0.006	0.004
	[0.000]***	[0.008]***
Male	0.025	0.013
	[0.488]	[0.733]
Age	0.228	0.214
	[0.195]	[0.239]
Age^2	-0.012	-0.011
	[0.276]	[0.333]
LIT_MOM	0.072	0.086
	[0.031]**	[0.023]**
HH_Lit	0.138	0.138
	[0.000]***	[0.000]***
HH_Fem	0.097	0.097
	[0.097]*	[0.098]*
Caste (SC/ST)	-0.047	-0.017
	[0.346]	[0.758]
Caste (Backward)	-0.002	0.025
	[0.972]	[0.612]
Muslim	0.053	0.119
	[0.442]	[0.113]
Household size	-0.006	-0.009
	[0.125]	[0.043]**
Land	0.007	0.008
	[0.001]***	[0.000]***
D_2^{QUART}	0.071	0.069
	[0.386]	[0.419]
D_3^{QUART}	-0.057	-0.013
	[0.568]	[0.902]
D_4^{QUART}	0.078	0.024
	[0.596]	[0.875]
OFF_FARM	0.004	0.004
	[0.034]**	[0.082]*
D^{HQ}	-0.014	-0.005
	[0.345]	[0.746]
$D^{HQ} * t$	-0.003	-0.002
	[0.255]	[0.325]
D_2^{ROAD} (Katcha)	0.205	0.242
	[0.116]	[0.073]*
D_3^{ROAD} (Paved)	0.266	0.178
	[0.048]**	[0.209]
D_4^{ROAD} (Pucca)	0.125	0.161
	[0.221]	[0.129]
Time	0.171	0.097
	[0.366]	[0.610]
District dummy * time	Yes	Yes
Village Fixed Effects	Yes	Yes
Constant	-0.011	-0.341
	[0.990]	[0.696]
Observations	704	704
R-squared	0.259	0.265

Robust p values in brackets

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

Table 3: Effect of distance to nearest secondary school on primary school enrollment and attendance (Household Fixed Effects Regression with SLC Data)

	Enrollment	Attendance
PRIM	0.06	0.055
	[0.360]	[0.438]
SEC	-0.112	-0.12
	[0.000]***	[0.000]***
SEC^2	0.008	0.008
	[0.000]***	[0.000]***
Male	0.012	-0.001
	[0.851]	[0.990]
Age	-0.105	-0.082
	[0.710]	[0.784]
Age^2	0.008	0.007
	[0.630]	[0.704]
LIT_MOM	0.074	0.031
	[0.366]	[0.734]
HH_Lit	0.098	0.052
	[0.292]	[0.578]
HH_Fem	0.074	0.091
	[0.439]	[0.338]
Household size	-0.009	-0.012
	[0.403]	[0.263]
Land	0.002	0.002
	[0.585]	[0.615]
D_2^{QUART}	0.153	0.216
	[0.127]	[0.043]**
D_3^{QUART}	0.07	0.129
	[0.585]	[0.345]
D_4^{QUART}	0.185	0.228
	[0.312]	[0.243]
OFF_FARM	0.004	0.004
	[0.184]	[0.203]
$D^{HQ} * t$	0.001	0
	[0.806]	[0.966]
D_2^{ROAD} (Katcha)	0.43	0.501
	[0.004]***	[0.001]***
D_3^{ROAD} (Paved)	0.412	0.392
	[0.013]**	[0.024]**
D_4^{ROAD} (Pucca)	0.131	0.16
	[0.312]	[0.219]
Time	0.223	0.303
	[0.330]	[0.180]
District dummy * time	Yes	Yes
Household Fixed Effects	Yes	Yes
Constant	0.66	0.545
	[0.560]	[0.652]
Observations	704	704
Number of Households	537	537
R-squared	0.262	0.299

Robust p values in brackets

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

Table 4: Effect of Distance to Nearest Secondary School on Primary Schooling (Instrumental Variables Estimation with SLC Data)

	Enrollment	Attendance
PRIM	0.275	0.264
	[0.113]	[0.139]
SEC	-0.099	-0.144
	[0.219]	[0.080]*
SEC^2	0.013	0.016
	[0.078]*	[0.038]**
Other Covariates	Yes	Yes
District Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	704	704
Number of Households	537	537

Table 5: Effects in Small Villages (SLC Data)

	Villages with # households <300		Villages with # households <100	
	Enrollment	Attendance	Enrollment	Attendance
PRIM	-0.059	-0.042	0.099	0.013
	[0.385]	[0.565]	[0.209]	[0.890]
SEC	-0.091	-0.119	-0.048	-0.049
	[0.005]***	[0.001]***	[0.019]**	[0.085]*
SEC^2	0.005	0.006	-	-
	[0.045]**	[0.014]**		
Other Covariates	Yes	Yes	Yes	Yes
District Trends	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes
Observations	530	530	91	91
Number of Households	398	398	63	63
R-squared	0.232	0.271	0.641	0.547

Table 6: Distance Dummies for Distance to Secondary School (SLC Data)

	Enrollment	Attendance
PRIM	0.019	0.013
	[0.769]	[0.855]
SEC (1km ≤ d < 5km)	-0.628	-0.575
	[0.003]***	[0.005]***
SEC (d ≥ 5km)	-0.489	-0.486
	[0.000]***	[0.000]***
Other Covariates	Yes	Yes
District Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	704	704
Number of Households	537	537
R-squared	0.256	0.282

Table 7: Effects within Village of Change in Distance outside Village (SLC Data)

	Enrollment	Attendance
PRIM	-2.031	-2.062
	[0.000]***	[0.000]***
SEC	-1.177	-1.191
	[0.000]***	[0.000]***
SEC^2	0.016	0.016
	[0.000]***	[0.000]***
Other Covariates	Yes	Yes
District Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	484	484
Number of Households	395	395
R-squared	0.493	0.485

Table 8: Effects after Controlling for Sibling Effects (SLC Data)

	Enrollment	Attendance
PRIM	0.06	0.055
	[0.358]	[0.435]
SEC	-0.116	-0.126
	[0.000]***	[0.000]***
SEC^2	0.008	0.008
	[0.000]***	[0.000]***
SEC_SIBLING	0.044	0.07
	[0.263]	[0.093]*
Other Covariates	Yes	Yes
District Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	704	704
Number of Households	537	537
R-squared	0.265	0.307

Table 9: Child Level Regressions (SLC Data)

	Enrollment	Attendance	Enrollment	Attendance
PRIM	0.012	-0.025	0.026	0.009
	[0.780]	[0.586]	[0.689]	[0.900]
SEC	-0.074	-0.066	-0.107	-0.104
	[0.000]***	[0.000]***	[0.000]***	[0.000]***
SEC^2	0.005	0.004	0.007	0.006
	[0.000]***	[0.002]***	[0.001]***	[0.002]***
Other Covariates	Yes	Yes	Yes	Yes
District Trends	Yes	Yes	Yes	Yes
Village Fixed Effects	Yes	Yes	Yes	Yes
Household Fixed Effects	No	No	Yes	Yes
Observations	1154	1132	1154	1132
R-squared	0.236	0.247	0.696	0.702

Table 10: Age-group Decomposition (SLC Data)

	6-7 Age-group		8-10 Age-group	
	Enrollment	Attendance	Enrollment	Attendance
PRIM	0.231	0.186	-0.08	-0.087
	[0.117]	[0.244]	[0.227]	[0.256]
SEC	-0.157	-0.155	-0.081	-0.085
	[0.066]*	[0.048]**	[0.001]***	[0.005]***
SEC^2	0.007	0.005	0.005	0.005
	[0.244]	[0.402]	[0.017]**	[0.021]**
Other Covariates	Yes	Yes	Yes	Yes
District Trends	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes
Observations	428	428	566	566
Number of Households	365	365	462	462
R-squared	0.459	0.508	0.359	0.369

Table 11: Results for Economically Backward Communities SC/ST & OBC (SLC Data)

	Enrollment	Attendance
PRIM	-0.042	-0.029
	[0.549]	[0.711]
SEC	-0.124	-0.128
	[0.000]***	[0.000]***
SEC^2	0.006	0.006
	[0.010]***	[0.007]***
Other Covariates	Yes	Yes
District Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	586	586
Number of Households	452	452
R-squared	0.306	0.326

Table 12: Gender Decomposition (SLC Data)

	Male		Female	
	Enrollment	Attendance	Enrollment	Attendance
PRIM	0.08	0.243	-0.059	-0.051
	[0.494]	[0.096]*	[0.571]	[0.604]
SEC	-0.105	-0.092	-0.256	-0.247
	[0.009]***	[0.052]*	[0.000]***	[0.000]***
SEC^2	0.007	0.008	0.01	0.01
	[0.035]**	[0.043]**	[0.003]***	[0.005]***
Other Covariates	Yes	Yes	Yes	Yes
District Trends	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes
Observations	475	475	445	445
Number of Households	404	404	373	373
R-squared	0.358	0.45	0.419	0.453

Table 13: Decomposition Based on Landholding (SLC Data)

	Landholding ≤ Median (1.78 Acre)		Landholding > Median (1.78 Acre)	
	Enrollment	Attendance	Enrollment	Attendance
PRIM	0.082	0.086	0.132	0.096
	[0.261]	[0.291]	[0.207]	[0.383]
SEC	-0.122	-0.148	-0.085	-0.101
	[0.003]***	[0.000]***	[0.030]**	[0.014]**
SEC^2	0.009	0.011	0.007	0.007
	[0.009]***	[0.002]***	[0.006]***	[0.005]***
Other Covariates	Yes	Yes	Yes	Yes
District Trends	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes
Observations	270	270	434	434
Number of Households	188	188	349	349
R-squared	0.415	0.451	0.336	0.368