Essays on the Economics of Education

Soham Sahoo

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Essays on the Economics of Education

Soham Sahoo

November 2015

Thesis Supervisor : Dr. Abhiroop Mukhopadhyay

Thesis submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Acknowledgements

I am deeply grateful to my advisor Dr. Abhiroop Mukhopadhyay for his invaluable guidance and constant encouragement throughout the period of my PhD work. It has been a great learning opportunity for me to work with him. His exceptional support, that has often extended beyond academic discussions, has helped me to carry on through ups and downs, and complete this thesis.

I would like to thank all the faculty members of the Economics and Planning Unit of Indian Statistical Institute, Delhi Centre for their teaching and support. I am grateful to Dr. Farzana Afridi for her guidance and contribution in the first essay of this thesis. My work has also benefited from interactions with many scholars and participants at various seminars and conferences. I would like to express my gratitude towards them.

I am thankful to Indian Statistical Institute for the research fellowship and the ideal institutional support it has provided. In addition to the academic environment, my stay here has been enjoyable because of its excellent hostel facilities, the cricket playground and the jogging track. I must thank my colleagues at the PPRU lab – Athisii, Rakeshji, Rameshji, and Shyam – for helping me whenever I have approached them. Shyam deserves a special thanks for his assistance and involvement during collection and cleaning of a data-set which I have used in two of my papers.

I wish to thank all my teachers who have inspired me to continue with my academic pursuit. I would especially like to express my gratitude to Dr. Saikat Sinha Roy and Dr. Bhaswar Moitra of Jadavpur University for motivating me to pursue doctoral studies.

My stay at ISI has been fun and memorable due to interactions with several colleagues and friends. I would always cherish the friendship of Pawan, Kanika, Anuradha, Anirban, Anup, Dushyant, Ridhima, Sutirtha, Debarpan, Anindya, Dhritiman, Prachi, Ashokankur, Mridu, Soumendu, Sabya, and many more. Our long discussions on all kinds of topics have been an enriching experience.

My parents have been perennially supportive and encouraging in all my academic endeavours. I have stayed outside home and away from them for many years; their patience and endurance have made it an easy journey for me. Finally, I am thankful to Sudipa for being there, and for always believing in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis consists of three empirical essays that investigate issues related to the economics of education. The main focus of this thesis is to study schooling outcomes of children in a developing country: India. The first chapter explores the effect of mothers' labour force participation on children's educational outcomes in the context of a large employment guarantee program in India. This study finds that workfare schemes enhancing women's access to economic opportunities have implications for intra-household resource allocation which in turn lead to a positive effect on children's education. The second essay investigates the effect of better access to secondary education on primary school participation of rural children. Drawing motivation from recent literature on the convex shape of the education-income relationship in developing countries, this paper shows how developments at higher levels of education (secondary) influence decision of school participation even at much lower levels (primary). The third chapter looks into the issue of gender disparity in the choice of private versus government schools in rural India. Relating to the emerging literature on the recent growth of private schooling in India, this paper finds that households are less likely to send girls, as compared to boys, to fee-charging private schools. Moreover, the difference in schooling cost between private and government schools is found to have a significant association with the gender gap in private school enrollment.

The following sections provide a brief description outlining the research questions, empirical strategies and main findings of each chapter of the thesis.

1.1 Female Labour Force Participation and Child Education in India: Evidence from the National Rural Employment Guarantee Scheme

This chapter studies the impact of women's participation in the labour force on their children's education in the context of the National Rural Employment Guarantee Scheme (NREGS) in India. While the main objective of NREGS is to alleviate rural poverty by providing employment to households on local public works, it mandates equal wage rates across gender and targets one third of the beneficiaries to be women. Thus, it has a potential to increase women's access to labour market opportunities. We use child level longitudinal data collected by the Young Lives Study in Andhra Pradesh in 2007 and 2009-10. To identify exogenous changes in mothers' labour force participation and household income, lagged value of the amount of funds allocated for NREGS and rainfall shocks within sub-districts are used as instrumental variables. Thus, we take into account the endogeneity of mother's labour supply decision and household income in a two-stage least squares model. The empirical specification controls for child level unobserved heterogeneity by including child specific fixed effects, and allows the districts to have differential trends depending on their initial level of development. The main results of this paper suggest that after controlling for household income, mothers' participation in the labour force has a significant and positive effect on children's time spent in school, enrollment, and grade progression. We provide further evidence that mother's participation in the workforce significantly increases households' expenditure on children's education, and also leads to improvements in her decision making power within the household. Thus, this paper points out to the positive effects of women's labour supply on intra-household outcomes. It also shows that design of public programs promoting women's participation has consequences beyond those immediately intended by policy makers.

1.2 Does Access to Secondary Education Affect Primary Schooling? Evidence from India

This chapter investigates if better access to secondary education increases enrollment and attendance in primary schools among the children in the 6–10 age-group. Various recent studies in the literature show that the perceived (actual, in many cases) returns to education is convex in developing countries like India. If significant economic returns require at least high school level education, then households may find it worthwhile to educate their children only when they can reach that level. Therefore, better access to secondary education, by increasing the possibility of continuation to higher levels of schooling, can be an important determinant of primary school participation. Using household level longitudinal data from rural Uttar Pradesh, this paper estimates a household fixed effects model and finds that a reduction in the distance to the nearest secondary school significantly raises primary school enrollment and attendance. We consider the potential endogeneity of school access and investigate if the effect is driven by unobservable confounding variables. Following a method developed by Altonji et al. (2005) and recently extended by Oster (2013), we find that any effect of unobservable factors is likely to be too small to bias the estimates. Further robustness analyses show that the result is unlikely to be driven by baseline trends. We show that the finding is not explained by sibling externality; rather, the possibility of continuation to secondary education is what drives the effect. We also find that the effect is larger in smaller villages, for poorer households, for boys (who are more likely to enter the labour force) and in those households who invest more on children's health. The effect is also complemented by village access to a bus-stop. The effect is significant for enrollment of younger children and for both enrollment and attendance of older children. In addition, we find suggestive evidence from a nationally representative survey that this effect may be quite widespread in India.

1.3 Intra-Household Gender Disparity in School Choice: Evidence from Private Schooling in India

This chapter explores the gender inequality within households in the decision of private versus government school choice in India. During the last decade, India has experienced a huge surge in both demand for and supply of private schools. These fee-charging private schools have been perceived by households as a better alternative to government schools where quality of education has remained a concern. In rural India, education of girls is often considered to be less worthy than boys' education. Since private schools charge fees and are more expensive than their government counterpart, households may be less likely to send girls, as compared to boys, to private schools. Using a three period longitudinal data-set from rural Uttar Pradesh, this paper seeks to identify the intra-household gender gap in private school choice. Since the school choice decision is observed only for children who are enrolled, this study estimates a household-fixed effects model that also takes into account any non-random enrollment decision in the analysis. The results show a significant gender gap in private school choice, which is present for both younger and older children, and is rising over time. Further, the study finds that a higher village level difference in average cost of private and government schooling is associated with a larger gender gap in private school choice, even after controlling for average school quality. Among the cost components, the direct cost, school fees in particular, comes out to be the most significant factor to be correlated with the gender gap. Acknowledging that cost differences could be endogeneous, the analysis shows that the result is robust to village level confounding trends. The findings also remain unaltered in a separate exercise that inspects the sensitivity of the effects with respect to potential omitted variable bias. In an era of rising school enrollment of girls, this study shows that the gender disparity in educational investment is manifested in the choice of schools.

Chapter 2

Female Labour Force Participation and Child Education in India: Evidence from the National Rural Employment Guarantee Scheme

2.1 Introduction

The World Development Report (World Bank, 2012), focusing on gender equality, highlights the economic disadvantages faced by women in the poorer regions of the world. Although, significant progress has been made in reducing gender disparities in health and educational outcomes, economic opportunities remain limited for women. In particular, existing research suggests that variation in women's labour force participation is associated with changes in individual and household behavior on several fronts including marriage (viz., van der Klaauw, 1996), fertility (viz., Goldin and Katz, 2002), and intra-household resource allocation (viz., Luke and Munshi, 2011). Thus, the policy priority of closing the gender differences in access to economic opportunities is critical for not only reducing poverty but also for improving individual and household welfare in developing countries.

In this paper we exploit the exogenous, temporal sub-district level variation in the intensity of implementation of India's National Rural Employment Guarantee Scheme (NREGS) to identify shifts in rural women's participation in the labour market and the latter's impact on their children's educational outcomes. The scheme operationalized the National Rural Employment Guarantee Act (2005) by providing a legal guarantee for up to 100 days of annual employment at a predetermined wage rate to rural households willing to supply manual labour on local public works.[1] From a gender perspective, there are two interesting features of this program. First, the wage rate offered in the scheme is uniform across gender, and second, it gives priority to female employment on public works and mandates one-third of the program beneficiaries to be women. Thus, NREGS not only brings employment opportunities to rural women's doorstep, the equal wage rates provided in the program can reduce any gender disparity prevalent in the rural labour markets.

A change in parental labour supply could affect children's educational outcomes purely due to an income effect. However labour force participation of mothers, per se, could impact her children's education through additional channels. First, mothers are likely to have more alternative uses of their time than fathers – market work, household chores and leisure. If children's time in doing household chores substitutes for mother's time then an increase in labour force participation of mothers may lead to a *decline* in the educational attainment of her children. However, if mother's and children's time are not close substitutes and child care services in the market are either unavailable or unaffordable, then it is possible that schools substitute for child-care services and children attend school when mothers are at work. Note, however, that this mechanism is unlikely to hold if child-care services are available within the household through the presence of older members (viz. grandparents).

Second, a working mother's say in household resource allocation decisions may be greater due to her higher earned income. Existing research suggests that this is likely to have a *positive* effect on her children's schooling. If an increase in mother's earned income translates into greater weight being attached to her preferences in resource allocation decisions of the household and mothers prefer to invest more in their children's health and education relative

[1] `http://nrega.nic.in/rajaswa.pdf`

to fathers (Blumberg, 1988; Thomas, 1990; Hoddinott and Haddad, 1995; Quisumbing and Maluccio, 2003), then we should see an improvement in child outcomes.

In this paper, we focus on the latter two channels of impact of mother's labour supply decisions on her children's education. We, therefore, control for changes in household level income to account for any income effect. The net impact of mother's participation in the labour force on her children's time allocation and, thereby, schooling, would then depend on which of the two effects dominate – the negative substitution effect or the positive bargaining power effect.[2]

There exists relatively little empirical research on the impact of parental labour supply on children's time allocation, particularly in a developing country context. Skoufias (1993) shows that an increase in female wages (and thereby female labour supply) in rural India reduces the time in school significantly but only for girls. Similar results are found by Grootaert and Patrinos (1999) in a cross-country study. However, Ilahi (1999) does not find any impact of female wages on children's time use in Peru.

In contrast to the sparse literature on time allocation effects, there is considerable empirical evidence suggesting that households' resource allocation decisions are made in a 'collective' (Chiappori, 1988) or a bargaining framework (McElroy and Horney, 1981) where the final allocation usually depends on the bargaining power or weights attached to the preferences of the members of the household. The importance of labour income as a determinant of women's bargaining power within the household has been highlighted by Anderson and Eswaran (2009). Using data from Bangladesh, the authors show that the effect of earned income on female autonomy is far greater than that of unearned income. Women who work on the household farm have no more autonomy than those who are housewives, while those who earn independent income have considerably greater autonomy. Luke and Munshi (2011) exploit data from tea plantations in South India where women are employed in permanent wage labour, to find that a relative increase in female income has a positive effect on their children's education.

[2]We are abstracting from any long term effects of changes in fertility due to increased labour force participation of women since we are looking at these changes over a short period of 2 to 3 years.

Qian (2008) shows that a change in agricultural pricing policy in post-Mao China, which increased female labour income, raised the educational attainment of all children within the household. However, when the policy increased male labour income, educational attainment of girls decreased but had no effect on boys' attainment.

We utilize child and household level panel data for 2007 and 2009-10 from the Young Lives Study in the state of Andhra Pradesh to assess the effect of mother's participation in the labour force on three educational outcomes of her children – time spent in school, enrollment, and grade attainment. To identify exogenous shifts in mothers' labour force participation, we take advantage of the temporal, sub-district level variation in the implementation of the NREGS. Our identification strategy utilizes the lagged amount of funds allocated for the implementation of the NREGS as well as lagged rainfall shocks within sub-districts as exogenous shifters of labour force participation of women in rural areas. The analysis accounts for unobservable child characteristics and for differences in time trends by initial economic development, both at the district as well as the sub-district level.

Our results suggest that participation of mothers in the workforce results in more time spent in school by their children. Almost half of the rise in children's time in school can be accounted by an increase in the probability of a child being enrolled in school when the mother works. Further, this increase in school participation is reflected in higher grade attainment of children. Labour force participation by the mother reduces the gap between the child's actual and ideal grade by more than 40 percentage points.

In order to understand the mechanisms through which these effects occur, we exploit household level data on education expenditures and on household members' say in decision-making and control over income from various sources. The analysis of the household level panel data on education expenditures show that mother's participation in the workforce significantly increased household's education expenditures on books, uniform and school fees in less landed households. Moreover, cross-sectional 2SLS analysis suggests that the probability that mothers have a say or control over utilization of household earnings from different sources increases when they participate in the labour force.

These results suggest that greater weightage on working mothers' preferences in household decision making could be among the primary drivers of the improvements in educational attainment of their children. While we cannot completely rule out the possibility that schools serve as child care centers while mothers work, this mechanism cannot account for all of our results: the impact of mother's labour force participation on children's time spent in school is robust to the presence of older members or potential child care givers in the household and is equally significant for older children who are relatively less likely to require child care services than the young. Our findings can, hence, be explained within the framework of a bargaining model of household resource allocation.

This study not only informs us about the relevance of women's labour supply to intra-household outcomes but it also addresses the broader policy issue of the role of the design of public programs in improving household outcomes in developing countries. Specifically, our paper extends the debate on the effect of workfare programs on household and individual welfare (Jalan and Ravallion, 2003; Azam, 2012; Dasgupta, 2013; Zimmermann, 2014; Klonner and Oldigies, 2014; Imbert and Papp, 2015) and finds evidence which suggests that mandating women's participation in public programs has consequences beyond those immediately intended by policy makers.

The remainder of the paper is organized as follows. Section 2.2 describes the data, and the methodology used in this paper. Section 2.3 discusses the results and section 2.4 concludes.

2.2 Data and Methodology

Our study is focused on Andhra Pradesh (AP) – India's fifth largest state in terms of population. We utilize data from the Young Lives Study (YLS) – a child and household level panel from six districts of AP. To date, there have been three rounds of YLS surveys. We conduct our empirical analysis at the level of the child using the two comparable waves of the YLS surveys - 2007 and 2009-10.[3]

The sample is restricted to children who were 5 to 14 years old, the school going age group, in 2007. In order to construct our data set we use the following exclusion rules: first, we only include households living in rural areas in both periods. Thus children in households which moved from rural to urban areas between the two rounds (less than 1 percent of our sample) were dropped. Of the remaining sample, 2.8 percent of the children in 2007 were not present in the subsequent round and were, therefore, dropped. Attrition is, therefore, negligible in the sample. Finally, we exclude children for whom there is some missing information on relevant covariates in either of the years. Our data set, after these exclusions, contains information on 3275 children for both years.

Table 2.1 describes the summary statistics for 2007 and 2009-10. The time spent in school by children in the reference period (a typical day in the last week) went up from 5.78 hours in 2007 to almost 7 hours in 2009-10. This increase is not driven by the changing age composition of children over the two years – a comparison of children by age cohorts shows that the time spent in school is significantly higher in 2009-10 relative to 2007 (*Figure* 2.1), reflective of more regular school attendance. Children in the survey, who reported attending school regularly, spent almost two hours more in school than those who reported going to school irregularly, on a typical day. We can, therefore, interpret greater time spent in school by a child as an indicator of greater number of days of school attendance. The rise in time spent in school was accompanied by a rise in the highest grade completed as well as the average grade progression during this period.[4] Enrollment rates also rose by 8 percentage points, largely a result of several 5 year olds joining school by 2009-10.

During the same period, there was a substantial rise in the proportion of children whose mothers were part of the labour force, i.e., self-employed (farm or non-farm), wage employed

[3]Data for 2002 (round 1of YLS) are not comparable to the 2007 and 2009-10 data (rounds 2 and 3 of YLS, respectively) because of differing methods of measuring labour supply. Moreover, there are no data on income earned by the household in 2002.

[4]We measure grade progression as the ratio of actual grade completed to the ideal grade for age. *Figure* 2.3 shows the distribution of this variable by gender and for each year. More details are given in section 2.3.

(farm or non-farm) or in salaried employment. Mothers' labour force participation rate increased from 69 percent in 2007 to 88 percent in 2009-10 as shown in *Table* 2.1. This was accompanied by a 34 percentage point increase in the proportion of children whose mothers' were participating in projects under the NREGS. A significant part of the rise in mothers' NREGS participation is due to the phasing of the program. To elaborate, the NREGS was first rolled out in 2006 in the 200 "poorest" districts of India (February 2006). Thereafter, it was extended to 130 additional districts in May 2007 and to all districts in the country by 1st April, 2008. Our study period coincides with the initial implementation of the NREGS (four YLS districts in Phase 1), followed by nation-wide coverage by 2008 (an additional YLS district each in Phases 2 and 3).[5] Given that the increase in mothers' overall work participation rate was accompanied by an increase in their participation in the NREGS, it suggests a potential impact of the scheme on mothers' labour force participation. We discuss this in detail later in our identification strategy.

The rise in mothers' labour force participation correlates positively with the change in time spent in school by their children, as shown in *Figure* 2.2. When we classify mothers into those who worked in 2007 but dropped out of the labour force in 2009-10, those who did not change their workforce participation status between 2007 and 2009-10 and those who did not work in 2007 but entered the labour force in 2009-10, we find that while the change in children's time spent in school is positive for all three categories, it has been largest for the last group. In contrast to the rising labour force participation of mothers, father's participation in paid labour (approximately 98 percent) was largely unchanged during this period.[6] The real, average annual household income (in 2009 rupees) also increased during this period, primarily due to a rise in non-agricultural income.

[5]Anantapur, Cuddapah, Karimnagar and Mahbubnagar implemented the NREGS in 2006. Srikakulam and West Godavari were the two districts that came under NREGS in 2007 and 2008, respectively.

[6]There was a rise in fathers' participation rate in NREGS during this period as well. Since overall labour force participation rate of fathers did not change, this suggests that fathers may have taken up NREGS work as an additional activity.

2.2.1 Methodology

We estimate the following specification:

$$
\begin{aligned}
TSS_{chmdt} = \alpha_{chmd} &+ \varphi_1 MOTHER\_WORK_{chmdt} + \varphi_2 INC_{hmdt} \\
&+ \delta_0 \mathbf{Z}_{hmdt} + \delta_1 \mathbf{X}_{chmd} * t + \delta_2 \mathbf{V}_{hmd} * t \\
&+ \delta_3 DEV_{md} * t + \delta_4 \mathbf{DEV\_DIST}_d * t + \varepsilon_{chmdt}
\end{aligned}
\tag{2.1}
$$

TSS_{chmdt} is the time spent in school by child c in household h in a sub-district or mandal m in district d at time t.[7] $MOTHER\_WORK_{chmdt}$ is a dummy variable that takes the value 1 if the mother participates in the labour market and 0 otherwise. INC_{hmdt} is the total annual household income from all sources. It includes parental income from participation in any paid work, including the NREGS.[8] Since time allocation decisions are a function of household demographics, we include the number and the average age of household members in $\mathbf{Z}_{hmdt}$. Schooling decisions depend on the age and gender of the child. For instance, older children are more likely to spend time working for wages or taking care of their siblings. We allow for this effect to be non-linear in age by including dummy variables for each age cohort (5 to 14 years) in 2007.[9] We also address any time trends in girl's education by including a dummy for female child. Since both these variables – the age of the child in 2007 and gender – are time invariant, they are included in $\mathbf{X}_{chmd}$ and interacted with the time trend t, which takes the value 0 for the year 2007 and 1 for the period 2009-10. $\mathbf{V}_{hmd}$ is a vector of households' wealth

[7] The time spent in school is recorded as hours spent in school on a typical day in the previous week. The total time spent on education on a typical day consists of time spent in school and time spent on studying outside school (private tuition and at home). The average time spent on education outside the school in the sample is less than 20 percent of the total time spent on education on a typical day. While we focus on time spent in school (TSS) as the dependent variable here, the same specification is adopted for other child education outcomes.

[8] Whether the income effect is significant or not is a function of the cost of schooling as well. If physical access to schooling is relatively easy and costs of schooling are subsidized (as is the case for public primary schools), any effect of an increase in household income may be muted for the age group under study here.

[9] Note that we do not include the child's age as a control variable because the change in age over our study period is the same for all children. Further, we do not include the secular time trend as it is a linear combination of the age cohort trends which are included in the specification.

– asset and land ownership in 2007 – interacted with t to allow for trends in the households' economic status affecting investment in child's education.[10]

We allow regions with different initial levels of economic development to have different trends: first, we include the total night lights in 2006 in a mandal, DEV_{md}, and interact it with t.[11] Second, districts where the NREGS was implemented earlier (Phase 1) were less developed than districts where the program was implemented later (Phases 2 and 3). We include a dummy which equals 1 if the child belongs to a phase 1 district, otherwise 0, in $\mathbf{DEV\_DIST}_d$.[12] In addition, to account for differences in the initial level of school participation, we also include the average primary school enrollment rate (for children aged 5–10 years) at the district level, in 2004-05 in $\mathbf{DEV\_DIST}_d$.[13] This vector is then interacted with t.

Given this specification, and using data on a balanced panel of children over the two time periods, we estimate a child fixed effects model. In doing so, we eliminate α_{chmd} or any unobservable, time invariant child characteristics.[14] If we assume that the deviations of the observed variables from their mean values are not correlated with the deviation of the error term from its mean, this specification yields a consistent estimate of our main coefficient of interest, φ_1.

The main concern with our estimation strategy is that household income and mothers'

[10] Asset index was constructed using principal component analysis of binary variables indicating ownership of durable consumer goods by the household viz., television, radio, car, motorbike, bicycle, telephone, mobile phone, refrigerator, fan, electric oven, table and chair, sofa and bedstead.

[11] Data on night lights is from the National Oceanic and Atmospheric Administration (NOAA) website: `http://www.noaa.gov`.

[12] Note that we do not use the phasing of the NREGS to identify the effect of mother's labour force participation on her children's schooling. The inclusion of the program phase trend as a control in our empirical specification invalidates district level phasing as the source of variation for identification.

[13] We source this information from the question on the "status of current attendance (enrollment) in educational institutions" in the 2004-05 employment and unemployment round of the National Sample Survey. One of the factors that could affect temporal changes in participation in schooling is a change in the quality of schools, specifically a shift from public to private schools. The YLS contains information on the type of school the child is enrolled in for only a subset of our sample. Community level data on the type of schools are not comparable between the two survey rounds. Our results are, however, robust to the inclusion of a dummy variable for whether a private school at any level existed in the locality in 2007 and 2009-10.

[14] The child fixed effects specification also accounts for any unobservables at the geographic (district, mandal and village) level that may be correlated with regressors on the right hand-side and that may also affect children's time spent in school.

labour supply decisions are likely to be determined simultaneously with investments in children's education. To address this simultaneity issue, we adopt a 2SLS estimation procedure using temporal variation in rainfall shocks in the previous agriculture year and allocation of funds to the NREGS program at the beginning of a financial year, both at the mandal level, as instruments. We describe our instruments and discuss their validity next.

2.2.2 Validity of instruments

Using the YLS data we find that the crop which the largest proportion of rural households cultivate (almost 36 percent across rounds 2 and 3) is rice. The cultivation of rice is highly water-intensive. Rice seedlings are grown in nurseries which are then manually transplanted into the flooded fields. It is therefore expected that rainfall will promote the development of rice seedlings enabling farmers to increase their cultivation of rice which in turn could raise agricultural incomes. Furthermore, studies suggest that rural women's labour force participation in India may be higher when households face adverse shocks. Women, for instance, could be expected to contribute to household income when agricultural employment opportunities and earnings are lower for men during a drought (Himanshu, 2011). Weather shocks could, therefore, carry implications for womens' labour force participation as well. Hence we use mandal level rainfall shock as one of the instruments.

We define rainfall shock as the deviation of actual rainfall from the long term average rainfall, divided by the long term standard deviation, at the mandal level.[15] Corresponding to the reference period for agricultural income in each YLS survey round, we calculate rainfall shocks during June 2005 to May 2006 (for the 2007 YLS round) and during June 2008 to May 2009 (for the 2009-10 YLS round). Note that the reference period for children's time spent

[15] The variable capturing rainfall shocks is constructed from the precipitation data available from the Centre for Climatic Research at the University of Delaware. The data include monthly precipitation values at 0.5 degree intervals in latitude and longitude. To match this data at the mandal level, the nearest latitude-longitude to each mandal headquarter is taken. To construct the rainfall shock at the mandal level, we calculate the long term (1990-91 to 2008-09) average mandal level rainfall in the months of an agriculture year. Standard deviation of rainfall for the same period is also calculated at the mandal level. Then rainfall shock is defined as the deviation of actual rainfall in the reference period from the long term average, divided by the standard deviation.

in school is the previous week – January to July, 2007 for the 2007 round of the YLS and August 2009 to March 2010 for the 2009-10 round. Thus rainfall shock is lagged with respect to children's educational outcomes (hence we refer to it as 'lagged rainfall shock' from now). This obviates any direct effect of contemporaneous rainfall shocks on children's time in school.

However, if rainfall shocks affect children's health outcomes contemporaneously and the latter persist into subsequent years, then previous rainfall shocks could directly impact current schooling outcomes. We use data on school attendance in the last 12 months for a subset of 'indexed' children to investigate this link.[16] We find that conditional on being enrolled, approximately 11 percent of these children reported missing school due to illness or injury in each round. To check whether missing school due to current morbidity is correlated with past rainfall shocks, we run a pooled OLS regression of a dummy variable for whether a child missed school due to an illness on lagged rainfall shock (results are available on request). The coefficient on lagged rainfall shock is insignificant. This provides suggestive evidence that in our sample past rainfall shock is unlikely to directly impact children's current time in school but should affect it via the impact on past household income.

Our second instrument is lagged NREGS funds *sanctioned* at the beginning of each financial year (April) for a mandal in 2009 rupees.[17] Note that the NREGS is envisaged as a demand-driven program: households are expected to apply for work to the village council (or gram panchayat, GP) and once a critical mass of demand is generated in a GP (a collection of 1 to 3 villages) in a mandal, a project has to be selected from the approved list of works and sanctioned by the district administration.[18] Thus to avoid any reverse causality, i.e. current NREGS funds being determined by current demand for work, we use lagged NREGS funds as an instrument. This is preferred to using actual expenditure on NREGS which is more likely

[16]The YLS has been collecting more detailed information on these children since the first round of the study in 2002. Data from the 2002 survey round is, however, not comparable to the later rounds used in our study as discussed in footnote 3.

[17]Data on the sanctioned funds at the mandal level were obtained from the Department of Rural Development, Government of Andhra Pradesh.

[18]Although the NREGA envisages a demand driven program, the reality is quite different according to several recent studies. Research on Andhra Pradesh (Afridi et al., 2013) indicates that the program is supply rather than demand driven.

to be driven by the demand for the program. Since the NREGS was initiated only in 2006-07, we use the 2006-07 financial year sanctioned program funds as an instrument for the 2007 YLS round.[19] For the 2009-10 survey, we take a one year lag with respect to the households' reference period for work activities and use the sanctioned funds in the financial year 2007-08 as the instrument.[20]

Our identification strategy, thus, uses the change in sanctioned program funds within mandals irrespective of the NREGS phase the district belongs to. But what determines differential changes in NREGS fund allocation? A plausible explanation is that sanctioned program funds change to reflect demand that varies by the level of economic development of the mandal. However, we have already controlled for trends by baseline economic prosperity (i.e. $DEV_{md} * t$ and $\mathbf{DEV\_DIST}_d * t$) in our specification. Another possible determinant of systematic changes in funds sanctioned at the mandal level could be electoral exigencies during the 2009 state elections in AP. However, Sheahan et al. (2014) do not find any evidence of vote buying through NREGS expenditures before the 2009 legislative council elections at the constituency level (which roughly correspond to mandals). Thus, controlling for the above trends, the residual variation in changes in NREGS fund allocation (which we use in our identification strategy) is likely to be quasi-random.[21]

A concern that remains is temporal learning by the local administration. To elaborate, say

[19]Our second stage results are similar when we use sanctioned NREGS funds in 2005-06 as an instrument for the 2007 round of the YLS. However, since the NREGS did not exist in 2005-06, the instrument takes the value 0 for all mandals, which makes our first stage weak. We, therefore, use sanctioned program funds in 2006-07 instead. Since NREGS has just been initiated in 2006, the assumption that the sanctioned funds were not demand driven is reasonable.

[20]Note that an overwhelming proportion of NREGS funds have been utilized for irrigation and water conservation projects since the program's inception in AP: soil and water conservation; drought proofing and afforestation; micro and minor irrigation works; rehabilitation of tanks and traditional water bodies; land levelling and bush and jungle clearance. It is, unlikely, therefore that the program could directly affect access to schools or children's allocation of time to household chores, viz. fetching drinking water. `http://nrega.ap.gov.in/Nregs/FrontServlet?requestType=Common_Ajax_engRH&actionVal=Display&page=WorkCatog_eng`.

[21]To test this claim we first regress the change in the mandal level sanctioned NREGS funds (between 2006-07 and 2007-08) on the district's NREGS phase and the mandal's total night lights in 2006. We then regress the residual of this regression on various observable characteristics of the mandal that may determine NREGS fund allocation (viz. total number of households, the number of people belonging to disadvantaged communities such as scheduled castes and tribes). We do not find a significant effect of any of these variables on NREGS fund allocation. Moreover, the point estimates do not suggest any systematic determinant of NREGS fund allocation. See *Table* 2.*A*1 in Appendix for details.

that the administration is learning how to implement NREGS, which improves between 2006-09 along with the quantum of sanctioned funds and this learning spills over to the provision of the public good of interest to us – education. In this case, our IV will not meet the exclusion restriction as it would have a direct effect on educational outcomes. However in Andhra Pradesh, school participation is near universal.[22] According to the Annual Survey of Education Report (ASER, 2006), the percentage of out of school rural children in the 6-14 age group was between 0 to 5 percent in all the YLS districts except West Godavari where it was between 5 to 10 percent in 2006. Further, the Right to Education (RTE) Act of India, which guarantees schooling for 6-14 year olds, came into effect in April, 2010, *after* our study period. Thus any administrative 'learning' with respective to public schooling would be minimal, if at all. Second, while it is quite likely that administrative capacity and NREGS implementation improved over time, it is unlikely that this was accompanied by administrative improvements in public schooling. The administrative machinery that has been created for the NREGS implementation at the grass roots level and which helps expand capacity for the program is different and delinked from that required for public schooling. Third, elections to village councils for a five year term were held in 2006. Since there were no changes in local governments during the period of our study there are unlikely to have been significant changes in local political will for implementation of public programs during 2007-10.

Finally, our third instrument is the interaction of lagged rainfall shock with lagged NREGS funds. This allows for the effect of rainfall shock on household income and mothers labour force participation to vary with NREGS funds. For instance, the effect on household income of a drought may be lower if there are more NREGS funds allocated to provide local employment in a mandal.

[22]Enrollment of children in the 6-10 years age group was almost 93 percent in both 2007 and 2009-10 while enrollment in the 11-14 age group was almost 81 percent in 2007 and 86 percent in 2009-10 in our sample.

2.3 Results

2.3.1 Impact of mother's work status on child's education

Table 2.2 shows the results for children's time spent in school. The coefficient on 'mother is working' is positive and significant in column (1). If the mother works, her child's time spent in school goes up by 0.293 hours in a day. To account for the possible endogeneity of labour force participation of mothers and household income we conduct the 2SLS analysis. But before discussing the second stage, we show the first stage results in *Table* 2.3. For the endogenous variable 'mother is working' the coefficient on lagged rainfall shock is significantly negative in column (1). This result lines up with the existing literature which suggests that in India women are more likely to work during periods of economic distress such as droughts. While the coefficient on lagged funds is insignificant, the interaction term is positive and significant in column (1), implying that the effect of NREGS on mother's work force participation depends on the level of lagged rainfall shock. To elaborate, at the mean lagged rainfall shock the total effect of NREGS funds on mother's labour force participation is significantly positive. The coefficient on lagged rainfall shock is positive and significant for annual household income in column (2). An increase in the lagged funds sanctioned for NREGS projects in a mandal increases the household income significantly. In times of good rainfall, however, the sanctioned funds have a lower marginal effect on the total income (as indicated by the negative coefficient of the interaction term). Our instruments are good predictors of the endogenous variables as indicated by the F statistics reported in the last row in *Table* 2.3.

Moving back to the second stage results in column (2) of *Table* 2.2, we find that the coefficient on the dummy for mother working continues to be positive, significant and has a higher coefficient than in the OLS-FE specification. When a mother works it leads to her children attending school 6.506 hours a day more. This effect is over and above the income effect from working, the point estimate of which is positive but not significant.

The coefficient of the 2SLS estimator measures the Local Average Treatment Effect (LATE) – the impact on time spent in school among children whose mothers changed their labour force participation status – in contrast to the OLS estimator which measures the impact of mothers' labour force participation for the full sample.[23] The increase in the magnitude of the 2SLS coefficient, as compared to the OLS, typically suggests that the impact of the treatment on those who changed their behaviour (referred to as compliers) is much larger than on the rest of the population (Angrist and Pischke (2009), Chapter 4.4-4.5).[24] Further, the instruments pass the overidentification test and the weak identification critical value cut-offs.[25]

The significant coefficient on the age dummies, interacted with time, declines at higher age groups. Thus, the older the child, the smaller is the point estimate on the increase in time spent in school. This reflects the higher opportunity cost of time in school for older children. The negative coefficient on the initial district level average enrollment rate in both the OLS-FE as well as 2SLS-FE specifications, together with the positive coefficient on age dummies interacted with time, suggests that in districts where school participation was high prior to 2007 there was a smaller increase in time spent in school between 2007 and 2009-10.[26]

The large increase in the number of hours spent in school suggests that mothers' work may play an important role in the school enrollment decisions of children. We, therefore, estimate the same specification used for time spent in school but with the dependent variable 'enrollment status'. The results are shown in *Table* 2.4. While the impact of a mother working on her child's school enrollment status is insignificant in OLS, it is positive and significant in the 2SLS estimate – probability of a child being enrolled in school rises by 47.2 percentage

[23]In particular, of the 1019 mothers who did not work in 2007, almost 72 percent entered the labour force in 2009-10. Of these mothers, more than 66 percent worked on NREGS projects in 2009-10.

[24]We recognize that any additional time spent in school could be substituted by less time spent studying outside school leading to an insignificant effect of mother's work on total time spent on education on a typical day. In an alternate specification, therefore, we consider the total time spent on education (including time spent studying outside the school) as the dependent variable. Our results are unchanged.

[25]Critical values for the Stock and Yogo (2005) weak-instrument test (5-percent significance) reported in the table, based on 2SLS size with 2 endogenous variables and 3 instruments, are 13.43, 8.18, 6.40 and 5.45 for the 10-percent, 15-percent, 20-percent, and 25-percent sizes, respectively.

[26]Lagged rainfall shock may force children to drop out of school in the year previous to our reference period. To check that our results are robust to this possibility we include a dummy for whether a child is enrolled in school as an endogenous regressor in our main specification (2SLS-FE). Our results do not change.

points. There are two factors that may drive this result. First, those children who are 5 years old and not enrolled in school at the baseline may enrol in a primary school as they get older if their mothers work. Second, typically children in older cohort at the baseline tend to drop out of school over time. Mothers' entering the work force may result in lower drop outs. This 47.2 percentage point increase in the probability of enrollment can potentially account for almost half of the increase in the time spent in school.[27]

The impact of current household income is positive and significant (0.003) when the dependent variable is enrollment status of the child. This suggests that any income effect of changes in labour force participation status of mother (or parents and other household members) predominantly influences enrollment decisions. This contrasts with the insignificance of household income on the child's time spent in school ($Table$ 2.2), an outcome that measures both the extensive (enrollment) as well as the intensive margins (more regular attendance) of children's school participation.

While we find that the time spent in school has risen due to mother's participation in the workforce, a pertinent question to ask is whether there was a concomitant improvement in the educational attainment of a child. Therefore, we conduct our analysis for grade attainment by a child with the dependent variable 'actual grade attainment of a child divided by grade the child should have completed at her age' with the same specification. The sample for this outcome, therefore, consists of children aged 6–14 years. We find a significant effect of mother's work participation on her child's grade progression – as reported in $Table$ 2.5 – the gap between a child's actual and ideal grade declines by 40.6 percentage points.[28] We do not find any heterogeneity in the average impact for any of our measures of educational attainment.[29] To elaborate, the impact of mother's work is significant and of comparable

[27]In our sample the average hours spent in school by children who are enrolled in school is approximately 7.15 hours. The 47.2 percentage point increase in enrollment rates, hence accounts for an average increase of 3.37 hours in school - almost half (or 47.13 percent) of the increase in time spent in school.

[28]This result also indicates that the effect of mothers' working on schooling is not driven purely by the enrollment of those who were 5 years old in 2007.

[29]There are certain caveats to interpreting the effect of mothers' working status on children's grade progression. First, the highest grade completed is right censored for the sub-sample of children who are still enrolled in school. This is not the case, however, for children who have completed schooling (17 year olds in 2009-10) or

magnitude across both genders and age groups (younger children or 5 to 10 year olds and older children or 11 to 14 year olds in 2007). The significant impact on the older age group of children, in particular, cannot be explained away by the argument that the mothers use schools as day care centres while they work.

2.3.2 Robustness checks

In this section we report further robustness tests of our results for time spent in school by accounting for alternative confounding mechanisms in *Table* 2.6.[30]

While the Right to Education Act, which aims to increase school participation rates of disadvantaged children, was implemented in Andhra Pradesh (and nation-wide) after our study period in April, 2010, it may have been preceded by construction of additional schools which may in turn be correlated with NREGS intensity at the mandal level.[31] Using school administrative data from the District Information System for Education (DISE) at the mandal level for the years 2006 and 2009, we include the number of schools as a control variable to address this concern. We do not find any significant change on the coefficient of 'mother working' in column (2) of *Table* 2.6.[32]

In our analysis, so far, we have focussed on the working status of the mother. However, if mothers' and fathers' decisions to work are correlated and the NREGS program affects fathers' work status as well, then our analysis may be identifying the impact of fathers' work status. However, inclusion of this variable does not affect the coefficient of the mothers' workforce participation in column (3).

Past rainfall shocks as well as income earned from participation in the NREGS in previous periods may affect current schooling through an effect on current household wealth. For ex-

have dropped out by the time of the survey interview. Second, the effect of parental labour market activities may not be reflected completely in grade attainment for those households which are interviewed before April (March is the last month of an academic year) since the highest grade attained by children in these households would be right censored. Finally, the highest grade completed is a stock variable that may be determined not just by current labour force participation of mothers but also their participation between 2007 and 2009-10.

[30] We report similar robustness checks for the enrollment outcome in *Table* 2.*A*2 in the Appendix.

[31] The reference period for the 2009-10 round of the YLS was August 2009 – March 2010.

[32] Our results also hold up if we include the total number of private and public schools separately.

ample, households may have bought or sold assets in response to the shocks and NREGS work opportunities in the previous periods. However, inclusion of current wealth may be endogenous as decisions to accumulate assets may correlate with schooling as well as labour supply decisions. Hence we have included baseline household wealth trends in our main specification. However, in order to see whether our estimated coefficient is affected by lagged effects of shocks and program fund allocation on household wealth, we include current household wealth as a regressor. Results in column (4) suggest that the inclusion of this variable does not have any significant impact on our estimated coefficient on mother's work status. Similarly, past NREGS work may affect migration in the household and that may impact schooling outcomes directly. To see whether this is the case, we control directly for the number of household members who were present in the previous YLS round but are absent in the current round to approximate migration.[33] Again, the inclusion of this variable does not affect our results (column (5)).

It has been contended that NREGS may affect wages for private work in the local labour market. If any increase in past wages influences current wages (Kaur, 2014) it may affect current school outcomes. The inclusion of current income in our main specification accounts for this possibility. However, to show that our results are robust to any remaining concerns on this count, we control for the highest current daily wage for males and females in the local community (columns (6) and (7)).[34] Again, the coefficient on mother's work remains unchanged.

The NREGS mandates the conduct of regular audits of program projects by local stakeholders.[35] Since one of the objectives of the audits is to make beneficiary households aware of their rights and entitlements, it is possible that a higher frequency of these audits leads to

[33] For this calculation we also use data from the 2002 round of YLS. We exclude household members who died and women who left the household due to marriage between 2007 and 2009-10.

[34] We obtained this information from the community level questionnaire in the YLS.

[35] A novel feature of the NREGS is the introduction of compulsory 'social' audits of projects carried out under the program by beneficiary households (and therefore referred to as 'social') at regular intervals. The AP government has institutionalised the conduct of these audits in the state since the inception of the NREGS in 2006.

a greater demand for access to better quality schools by parents as well as a rise in women's participation in the NREGS and in the workforce. Hence any observed relationship between mother's labour market participation and children's time in school could be driven by a rise in households' awareness levels. To control for this, we allow the trend to depend on the number of social audits that have taken place in the mandal between the two survey rounds interacted with time in column (8). Our results are unaffected.

Finally, to test for the possibility that schools substitute for day care for working mothers we control for the demographic composition of the household in column (9) in *Table* 2.6. The effect of mothers working on children's time in school should be insignificant if there are older siblings or grandparents in the household to take care of the younger ones. But the interpretation of our results is unchanged when we control for the presence of older siblings and of household members in the 60+ age group.

2.3.3 Discussion of results

Our results establish that when a mother works, there is a significant, positive impact on her children's educational attainment. Since we have accounted for any income effects in our specification, can an improvement in mothers' say in household decision-making when she participates in the labour market explain our results? If so, we should see a positive effect of mother's labour force participation on other schooling indicators that provide more direct evidence on investments in children. We, therefore, use household level data on education expenditures to test our hypothesis further.[36]

Our specification is now run at the household level (since these data are not available at the child level) with additional controls for the number of children in the 5-17 age group, and

[36] An alternative mechanism that could explain our results is the mandatory provision of child care facilities at NREGS work sites. Mothers who participated in NREGS work may have had better access to child care facilities. This would free up the time of school-going age siblings, particularly girls, who could then attend school more regularly. The 2007 YLS survey respondents were asked whether the NREGS participant had "benefited from child care facilities at the worksite". In the 2009-10 survey, the respondents were asked: "Were there child care facilities in the last (NREGS) worksite?" Only 1 percent of households report using on-site child care facilities in 2007 while more than 80 percent of households report absence of child care facilities at the last work-site in 2009-10. We, therefore, do not consider this as a valid explanation of our findings.

the average age and gender composition of this group in the household. Our main coefficient of interest is working status of the mother of the indexed child and his/her siblings in the household. Thus the dependent variable captures the aggregate expenditure on all the children who belong to the age group of 5-17 years in the household. Note that indexed children and their direct siblings constitute almost 94 percent of the sample of all children in 5-17 age group.

We find no significant result for the overall sample. Stratifying the households by baseline land ownership, we find significant results only for households with less than median landownership. We report only these significant results in *Table* 2.7 (other results are available on request). The analysis suggests that mothers' participation in the labour force increases total schooling expenses and those related to more regular attendance (i.e. books and uniform, column (2)) for the less landed households. This indicates that mother's labour force participation is indeed leading to more investment in education in less landed households.

To explore the bargaining power mechanism further we use data available in the second round of the YLS to analyse whether participation in the labour market led to improvements in mothers' decision-making abilities within households.[37] Our dependent variable is the binary response to two questions, each, for three sources of household income:

a. "Is the caregiver responsible for making the key decisions about any of the plots (*Land*) / work for wages activities (*Wage activities*) / business and self-employment activities (*Business and self-employment*)?"

b. "Does the caregiver control the use of the earnings from the sale of goods or rent from any of these plots (*Earnings from land*) / from any work for wages activities (*Earnings from wage activities*) / from any business and self-employment activities (*Earnings from business and self-employment*)?"

The sample is restricted to caregivers who are mothers in age group 16-60 years. Our main variable of interest is whether the mother works. Results for the 2SLS specification are

[37]These data were not collected by the YLS in 2009-10. Our analysis, therefore, is cross-sectional.

reported in *Table* 2.8. The positive and significant coefficient on 'mother is working' across all outcomes, suggests that greater participation of mothers in the labour market does increase the say and control these women have on important decisions being made within the household. In rural areas earnings from land, wages and business and self-employment activities are likely to be the most important sources of income for households. This result, therefore, bolsters our claim that an increase in work opportunities for women is likely to have a positive effect on their decision-making abilities within the household.[38]

2.4 Conclusion

We utilize the sub-district level temporal variation in the intensity of the National Rural Employment Guarantee Scheme (NREGS) and in rainfall shocks in Andhra Pradesh (AP) to determine the effect of exogenous changes in the demand for labour on women's labour force participation and thereby their children's educational outcomes.

Using panel data from the Young Lives Study for 2007 and 2009-10, we find that participation of mothers in the work force has a positive effect on the probability of her children being enrolled in school and her children's time in school. This carries implications for the latter's educational attainment as well. Specifically, children whose mothers work move closer to their ideal grade for age.

We find evidence that suggests that the positive impact of mothers' participation in work could be due to her improved position in household decision-making. Our assertion is supported by recent qualitative evidence on the empowering effects of NREGS on rural women (Pankaj and Tankha, 2010; Khera and Nayak, 2009).

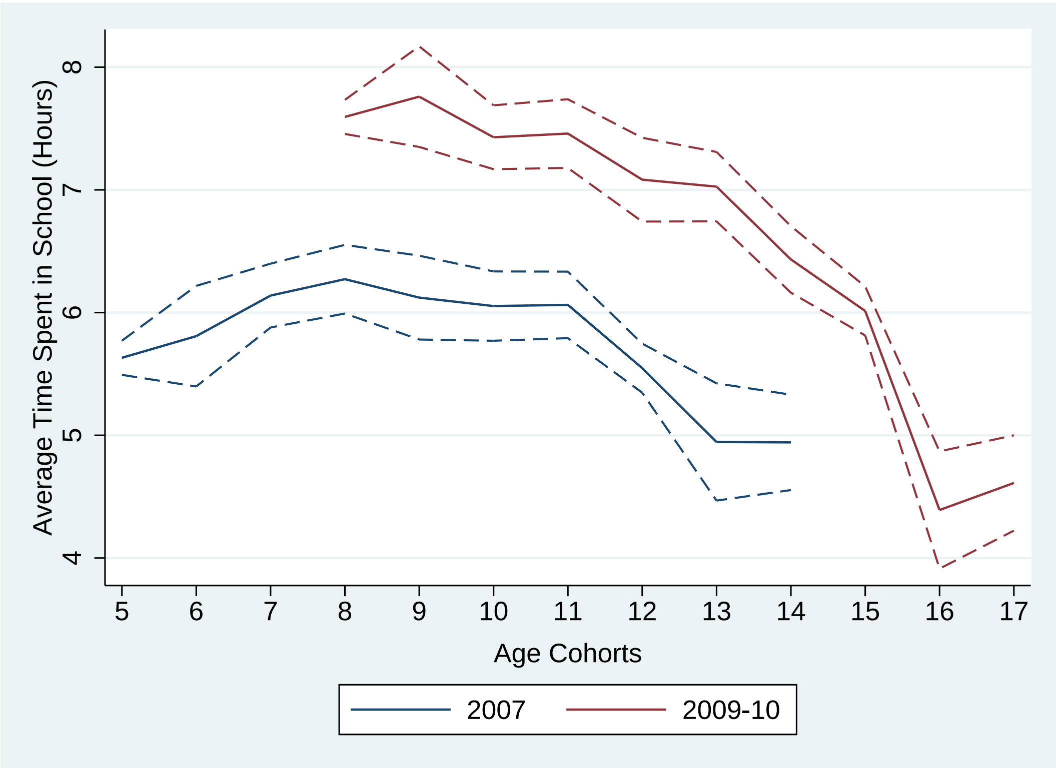
Although our results are contextual and specific to AP – a state which has traditionally exhibited high rates of women's participation in work relative to the national average and has also been among the best implementers of the NREGS since its inception – they provide

[38]Alternative definitions of mother's workforce participation – share of mother's NREGS income in total household income and total number of days mother worked in the reference period – lead us to similar conclusions. See *Table* 2.*A*3 in Appendix for details.

strong evidence of the beneficial effects of increasing women's participation in the labour force. Furthermore, the findings suggest that the design of public programs matter and have consequences beyond those intended by policy makers.

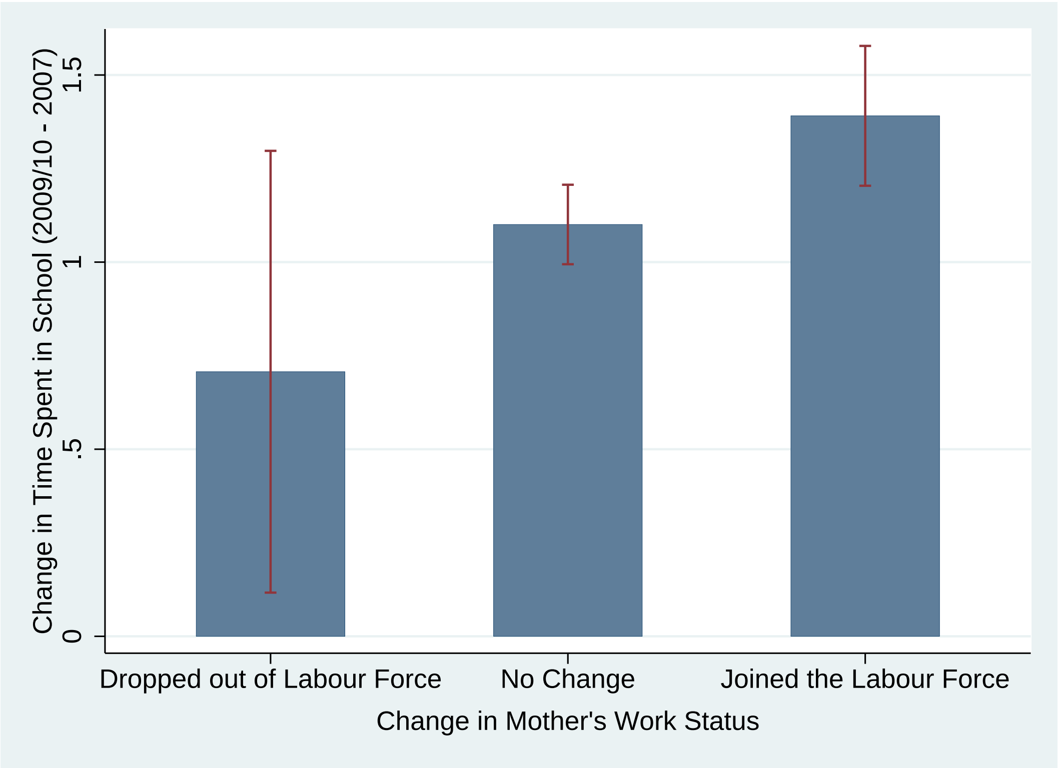
Figures and Tables for Chapter 2

Figure 2.1: Child's time spent in school



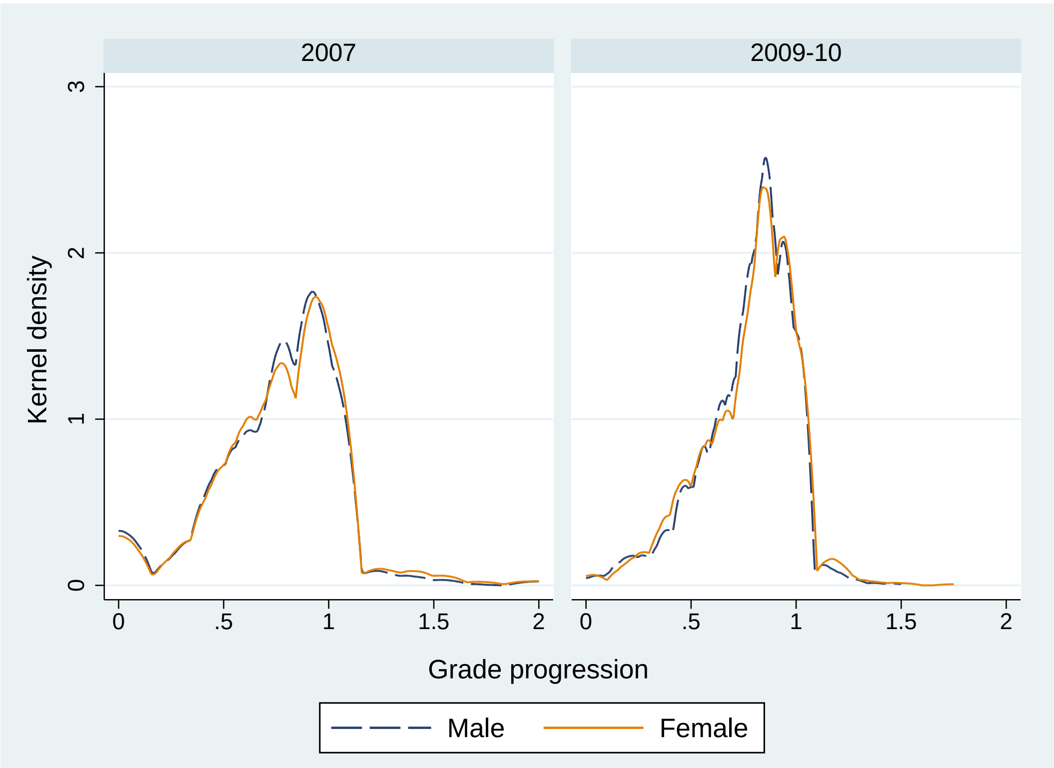
Source: Authors' calculations from YLS data.
Notes: Dashed lines indicate the 95% confidence interval.

Figure 2.2: Change in mother's labour force participation and child's time spent in school (2007 – 2009-10)



Source: Authors' calculations from YLS data.
Notes: Confidence intervals indicated in the figure are at 95%.

Figure 2.3: Kernel density of grade progression by gender and year



Source: Authors' calculations from YLS data.

Table 2.1: Summary statistics

	2007			2009-10		
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Child Characteristics						
Sex (female=1, male=0)	3275	0.51	0.50	3275	0.51	0.50
Age (years)	3275	8.35	3.01	3275	11.35	3.01
Enrollment	3275	0.79	0.41	3275	0.87	0.34
Time spent in school	3275	5.78	2.20	3275	6.93	2.64
Highest grade completed	2165	3.92	2.33	2165	6.29	2.58
Grade progression	2165	0.76	0.33	2165	0.78	0.22
Mother's Characteristics						
Mother's age (years)	3275	30.56	5.64	3275	33.52	5.59
Mother's education (highest grade completed)	3270	1.86	3.28	3270	1.86	3.28
Whether mother is working	3275	0.69	0.46	3275	0.88	0.32
Whether mother has worked in NREGS	3272	0.28	0.45	3228	0.62	0.49
Father's Characteristics						
Father's age (years)	3127	36.34	6.36	3101	39.26	6.29
Father's education (highest grade completed)	3126	3.91	4.51	3126	3.91	4.51
Whether father is working	3121	0.99	0.10	3095	0.98	0.14
Whether father has worked in NREGS	3121	0.25	0.43	3073	0.48	0.50
Household Characteristics						
Annual non-agricultural income (Rs.)	3275	28349	30452	3275	41404	46015
Annual agricultural income (Rs.)	3275	4100	21489	3275	8258	38656
Total income (Rs.)	3275	32449	36312	3275	49662	58843
Household size	3275	5.70	2.10	3275	5.71	2.19
Number of males in the age-group 16-60	3275	1.31	0.81	3275	1.40	0.85
Average age of household members	3275	22.14	5.72	3275	24.56	5.62
Land owned by household (acres)	3275	2.11	3.21	3275	3.38	38.59
Asset index	3275	-0.60	1.49	3275	0.55	1.81
Region Characteristics						
Mandal level total night lights in 2006	3275	860.49	335.78	3275	860.49	335.78
Baseline enrollment rate in the district (2004-05)	3275	0.92	0.07	3275	0.92	0.07
Instruments						
Lagged Rainfall shock	3275	1.71	0.78	3275	-0.02	0.65
Lagged amount of sanctioned NREGS funds in crores (Rs.)	3275	8.53	7.80	3275	10.59	5.56

Source: Young Lives data for all variables except Mandal level total night lights in 2006 (NOAA website), Baseline enrollment rate in the district (National Sample Survey 2004-05), Lagged Rainfall shock (Centre for Climatic Research, University of Delaware), and Lagged amount of sanctioned NREGS funds (Department of Rural Development, Government of Andhra Pradesh).
Note: 1 crore = 10 million

Table 2.2: Effect of mother's work status on child's time spent in school

Variable	Time spent in school	
	OLS-FE	2SLS-FE
	(1)	(2)
Mother is working	0.293***	6.506***
	(0.101)	(1.102)
Annual household income in thousands (Rs.)	0.001	0.013
	(0.001)	(0.013)
Age 5 in 2007 x Time	8.768***	14.127***
	(0.695)	(1.275)
Age 6 in 2007 x Time	8.674***	13.948***
	(0.702)	(1.280)
Age 7 in 2007 x Time	8.114***	13.310***
	(0.699)	(1.264)
Age 8 in 2007 x Time	8.030***	13.542***
	(0.719)	(1.336)
Age 9 in 2007 x Time	7.700***	12.987***
	(0.708)	(1.265)
Age 10 in 2007 x Time	7.743***	13.218***
	(0.696)	(1.297)
Age 11 in 2007 x Time	7.233***	12.542***
	(0.714)	(1.301)
Age 12 in 2007 x Time	7.220***	12.827***
	(0.683)	(1.303)
Age 13 in 2007 x Time	6.363***	11.784***
	(0.762)	(1.344)
Age 14 in 2007 x Time	6.458***	12.054***
	(0.728)	(1.347)
Female child x Time	-0.207**	-0.306**
	(0.088)	(0.135)
Household size	-0.023	-0.050
	(0.042)	(0.083)
Number of males in the age-group 16-60	0.006	0.114
	(0.097)	(0.127)
Average age of household members	0.017	0.029
	(0.012)	(0.019)
Land owned (acres) by household in 2007 x Time	0.015	0.043*
	(0.013)	(0.022)
Household's asset (index) in 2007 x Time	0.060**	0.149**
	(0.028)	(0.065)
Mandal level total night lights in 2006 x Time	0.0001	-0.0005*
	(0.0002)	(0.0002)
NREGS phase 1 district x Time	-0.936***	0.803**
	(0.126)	(0.336)
Baseline enrollment rate in the district x Time	-7.822***	-14.074***
	(0.766)	(1.532)
Child Fixed Effects	Yes	Yes
Observations	6,550	6,550
Number of children	3,275	3,275
R-squared	0.256	
Weak-identification Stat (Kleibergen-Paap rk Wald F)		7.750
Stock-Yogo weak-identification test critical values		6.40 (20%)
Over-identification Stat (Hansen J)		1.701
Over-identification p-value		0.192

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%.

Table 2.3: First stage regressions (for 2SLS-FE regression in *Table* 2.2, column 2)

Variable	Mother is working	Annual household income in thousands (Rs.)
	(1)	(2)
Lagged rainfall shock	-0.207***	7.607**
	(0.021)	(3.069)
Lagged amount of sanctioned NREGS funds in crores (Rs.)	0.001	0.470*
	(0.001)	(0.240)
Lagged rainfall shock x Lagged amount of sanctioned NREGS funds in crores (Rs.)	0.005***	-0.593***
	(0.001)	(0.095)
Age 5 in 2007 x Time	-2.032***	29.050
	(0.171)	(26.647)
Age 6 in 2007 x Time	-2.026***	29.953
	(0.173)	(28.678)
Age 7 in 2007 x Time	-2.011***	30.703
	(0.172)	(26.904)
Age 8 in 2007 x Time	-2.034***	22.197
	(0.172)	(26.122)
Age 9 in 2007 x Time	-2.020***	27.439
	(0.171)	(26.265)
Age 10 in 2007 x Time	-2.045***	27.600
	(0.171)	(26.407)
Age 11 in 2007 x Time	-2.007***	23.858
	(0.172)	(26.676)
Age 12 in 2007 x Time	-2.071***	25.730
	(0.170)	(26.437)
Age 13 in 2007 x Time	-2.051***	35.460
	(0.176)	(28.385)
Age 14 in 2007 x Time	-2.065***	25.810
	(0.171)	(27.004)
Female child x Time	0.011	1.936
	(0.015)	(1.822)
Household size	-0.005	4.903***
	(0.007)	(1.499)
Number of males in the age-group 16-60	-0.018	1.104
	(0.014)	(2.697)
Average age of household members	-0.001	-0.327
	(0.002)	(0.305)
Land owned (acres) by household in 2007 x Time	-0.007**	0.445
	(0.003)	(0.554)
Household's asset (index) in 2007 x Time	-0.016***	3.713***
	(0.006)	(1.047)
Mandal level total night lights in 2006 x Time	0.000***	0.001
	(0.000)	(0.003)
NREGS phase 1 district x Time	-0.223***	9.933***
	(0.028)	(3.695)
Baseline enrollment rate in the district x Time	1.945***	-6.986
	(0.164)	(25.301)
Child Fixed Effects	Yes	Yes
Observations	6,550	6,550
Number of children	3,275	3,275
R-squared	0.297	0.128
Angrist-Pischke multivariate F stat of excluded instruments	33.75	14.27

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%.

Table 2.4: Effect of mother's work status on child's school enrollment

Variable	Enrollment	
	OLS-FE	2SLS-FE
	(1)	(2)
Mother is working	-0.009	0.472***
	(0.017)	(0.138)
Annual household income in thousands (Rs.)	0.0002	0.003**
	(0.0001)	(0.002)
Age 5 in 2007 x Time	0.539***	0.898***
	(0.116)	(0.177)
Age 6 in 2007 x Time	0.308***	0.661***
	(0.119)	(0.183)
Age 7 in 2007 x Time	0.168	0.512***
	(0.116)	(0.177)
Age 8 in 2007 x Time	0.136	0.522***
	(0.117)	(0.183)
Age 9 in 2007 x Time	0.102	0.457***
	(0.115)	(0.174)
Age 10 in 2007 x Time	0.096	0.467***
	(0.114)	(0.178)
Age 11 in 2007 x Time	0.063	0.431**
	(0.116)	(0.178)
Age 12 in 2007 x Time	0.046	0.433**
	(0.113)	(0.179)
Age 13 in 2007 x Time	-0.036	0.312*
	(0.123)	(0.184)
Age 14 in 2007 x Time	-0.004	0.382**
	(0.118)	(0.185)
Female child x Time	-0.010	-0.023
	(0.014)	(0.018)
Household size	-0.017**	-0.031**
	(0.008)	(0.012)
Number of males in the age-group 16-60	0.0001	0.006
	(0.015)	(0.019)
Average age of household members	0.004**	0.006**
	(0.002)	(0.003)
Land owned (acres) by household in 2007 x Time	-0.001	0.000
	(0.002)	(0.004)
Household's asset (index) in 2007 x Time	0.013***	0.011
	(0.005)	(0.008)
Mandal level total night lights in 2006 x Time	0.00004*	0.00001
	(0.00002)	(0.00003)
NREGS phase 1 district x Time	-0.033*	0.074*
	(0.019)	(0.042)
Baseline enrollment rate in the district x Time	-0.230*	-0.720***
	(0.126)	(0.208)
Child Fixed Effects	Yes	Yes
Observations	6,550	6,550
Number of children	3,275	3,275
R-squared	0.249	
Weak-identification Stat (Kleibergen-Paap rk Wald F)		7.750
Stock-Yogo weak-identification test critical values		6.40 (20%)
Over-identification Stat (Hansen J)		1.439
Over-identification p-value		0.230

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%.

Table 2.5: Effect of mother's work status on child's grade progression

Variable	Grade progression	
	OLS-FE	2SLS-FE
	(1)	(2)
Mother is working	0.060***	0.406***
	(0.014)	(0.101)
Annual household income in thousands (Rs.)	0.00001	0.0002
	(0.0001)	(0.001)
Other control variables	Yes	Yes
Child Fixed Effects	Yes	Yes
Observations	4,218	4,218
Number of children	2,109	2,109
R-squared	0.046	
Weak-identification Stat (Kleibergen-Paap rk Wald F)		7.387
Stock-Yogo weak-identification test critical values		6.40 (20%)
Over-identification Stat (Hansen J)		0.384
Over-identification p-value		0.535

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%. Additional control variables included are: time interacted with age dummies, time interacted with gender, household size, number of males in the age-group 16-60, average age of household members, land owned by household in 2007 interacted with time, household's asset (index) in 2007 interacted with time, time trends depending on whether the district is NREGS Phase 1 district, mandal level total night lights in 2006, and baseline enrollment rate in the district. The same set of instruments (lagged rainfall shock, lagged amount fund sanctioned in NREGS, and their interaction) are used for annual household income and mother's working status.

Table 2.6: Robustness checks for the effect of mother's work status on child's time spent in school (2SLS-FE)

Variable	Time Spent in school								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mother is working	6.506***	7.047***	6.472***	6.556***	6.453***	6.421***	6.611***	6.151***	6.498***
	(1.102)	(1.299)	(1.106)	(1.104)	(1.095)	(1.074)	(1.164)	(1.018)	(1.102)
Number of schools in the mandal		Yes							
Father is working			Yes						
Asset quartiles				Yes					
Number of household members who migrated					Yes				
Maximum daily male wage rate in the community						Yes			
Maximum daily female wage rate in the community							Yes		
Number of NREGS social audits in mandal x Time								Yes	
Number of household members aged above 60 years									Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,550	6,550	6,550	6,550	6,550	6,182	6,182	6,550	6,550
Number of children	3,275	3,275	3,275	3,275	3,275	3,091	3,091	3,275	3,275
Weak-identification Stat (Kleibergen-Paap rk Wald F)	7.750	4.361	7.733	7.913	7.758	6.736	6.501	7.025	7.778
Stock-Yogo weak-identification test critical values	6.40 (20%)	5.45 (25%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)
Over-identification Stat (Hansen J)	1.701	0.191	1.649	1.809	1.730	0.932	1.905	1.619	1.695
Over-identification p-value	0.192	0.662	0.199	0.179	0.188	0.334	0.168	0.203	0.193

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%. Additional control variables included are: household income, time interacted with age dummies, time interacted with gender, household size, number of males in the age-group 16-60, average age of household members, land owned by household in 2007 interacted with time, household's asset (index) in 2007 interacted with time, time trends depending on whether the district is NREGS Phase 1 district, mandal level total night lights in 2006, and baseline enrollment rate in the district. Instruments are as mentioned in *Table* 2.5.

Table 2.7: Effect of mother's work status on household's education expenditure (2SLS-FE)

Variable	Baseline land ≤ Median		
	Total education expenditure	School fees, books and uniform	Transport and private tuition
	(1)	(2)	(3)
Mother is working	1,822.808*	1,616.609*	201.725
	(1,099.478)	(960.939)	(554.601)
Other control variables	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes
Observations	1,998	1,998	2,004
Number of households	999	999	1,002
Weak-identification Stat (Kleibergen-Paap rk Wald F)	3.256	3.256	3.204
Stock-Yogo weak-identification test critical values	5.45 (25%)	5.45 (25%)	5.45 (25%)
Over-identification Stat (Hansen J)	0.190	0.647	0.416
Over-identification p-value	0.663	0.421	0.519

Notes: Robust standard errors clustered at the household level in parentheses. * significant at 10%, ** 5% and *** 1%. Additional control variables: average age of children in the school going age (5-17 years), number of boys in school going age, number of girls in school going age, household income, household size, number of males in the age-group 16-60, average age of household members, land owned by household in 2007 interacted with time, household's asset (index) in 2007 interacted with time, time trends depending on whether the district is NREGS Phase 1 district, mandal level total night lights in 2006, and baseline enrollment rate in the district. Instruments as mentioned in *Table* 2.5.

Table 2.8: Effect of mother's work status on her decision-making within household (2SLS)

Variable	Land	Earnings from Land	Wage Activities	Earnings from Wage Activities	Business & Self-employment	Earnings from Business & Self-employment
	(1)	(2)	(3)	(4)	(5)	(6)
Mother is working	0.840***	1.046***	0.908***	1.404***	0.733**	1.063**
	(0.248)	(0.262)	(0.287)	(0.384)	(0.347)	(0.424)
Observations	1,881	1,908	1,498	1,472	452	450
Weak-identification Stat (Kleibergen-Paap rk Wald F)	8.282	9.707	4.151	3.370	3.560	7.340
Stock-Yogo weak-identification test critical values	8.18 (15%)	8.18 (15%)	5.45 (25%)	5.45 (25%)	5.45 (25%)	6.40 (20%)
Over-identification Stat (Hansen J)	0.00957	0.304	1.220	0.731	0.0469	0.0605
Over-identification p-value	0.922	0.581	0.269	0.393	0.829	0.806

Notes: Robust standard errors clustered at the community level in parentheses. * significant at 10%, ** 5% and ***1%. Other control variables include: mother's age, mother's age squared, mother's highest grade passed, number of sons and daughters in the age-group of 0-5, 6-14, and 15 or above (separately), household size, number of males in the age-group 16-60, average age of household members, household income, land owned by household, household's asset index, religion, caste, whether the district is NREGS Phase 1 district, and mandal level total night lights in 2006. Instruments as mentioned in *Table* 2.5.

Table 2.$A$1: Determinants of change in the amount of allocated NREGS Funds

Variable	Residual change in allocated funds between 2006-07 and 2007-08 in a mandal			
	(1)	(2)	(3)	(4)
Number of rural households	0.000793			
	(0.000594)			
Scheduled caste population		-0.000585		
		(0.000666)		
Scheduled tribe population		0.000171		
		(0.000169)		
Literate population			0.000337	
			(0.000268)	
Number of agricultural labourers				0.000508
				(0.000733)
Constant	-7.637	4.179	-6.265	-2.967
	(6.706)	(6.313)	(6.052)	(6.000)
Observations	15	15	15	15
R-squared	0.121	0.250	0.108	0.036

Notes: Standard errors in parentheses. * significant at 10%, ** 5% and *** 1%.

Table 2.$A$2: Robustness checks for the effect of mother's work status on child's school enrollment (2SLS-FE)

Variable	Enrollment								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mother is working	0.642***	0.399***	0.471***	0.452***	0.455***	0.437***	0.494***	0.472***	0.471***
	(0.166)	(0.146)	(0.139)	(0.135)	(0.137)	(0.136)	(0.151)	(0.130)	(0.138)
Number of schools in the mandal		Yes							
Father is working			Yes						
Asset quartiles				Yes					
Number of household members who migrated					Yes				
Maximum daily male wage rate in the community						Yes			
Maximum daily female wage rate in the community							Yes		
Number of NREGS social audits in mandal x Time								Yes	
Number of household members aged above 60 years									Yes
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,550	6,550	6,550	6,550	6,550	6,182	6,182	6,550	6,550
Number of children	3,275	3,275	3,275	3,275	3,275	3,091	3,091	3,275	3,275
Weak-identification Stat (Kleibergen-Paap rk Wald F)	7.750	4.361	7.733	7.913	7.758	6.736	6.501	7.025	7.778
Stock-Yogo weak-identification test critical values	6.40 (20%)	5.45 (25%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)	6.40 (20%)
Over-identification Stat (Hansen J)	1.749	2.624	1.452	1.660	1.423	2.318	1.301	1.441	1.448
Over-identification p-value	0.186	0.105	0.228	0.198	0.233	0.128	0.254	0.230	0.229

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%. Additional control variables included are: household income, time interacted with age dummies, time interacted with gender, household size, number of males in the age-group 16-60, average age of household members, land owned by household in 2007 interacted with time, household's asset (index) in 2007 interacted with time, time trends depending on whether the district is NREGS Phase 1 district, mandal level total night lights in 2006, and baseline enrollment rate in the district. The same set of instruments (lagged rainfall shock, lagged amount fund sanctioned in NREGS, and their interaction) are used for annual household income and mother's working status.

Table 2.$A3$: Effect of mother's NREGS income as a share of total household income, and mother's number of days worked, on child's time spent in school and enrollment

Variable	Time spent in school	Enrollment	Time spent in school	Enrollment
	(1)	(2)	(3)	(4)
Share of mother's NREGS income in total income of household	24.501***	2.244***		
	(6.160)	(0.684)		
Number of days mother worked			0.044***	0.003**
			(0.011)	(0.001)
Other control variables	Yes	Yes	Yes	Yes
Child Fixed Effects	Yes	Yes	Yes	Yes
Observations	6,374	6,374	6,546	6,546
Number of children	3,187	3,187	3,273	3,273
Weak-identification Stat (Kleibergen-Paap rk Wald F)	9.973	9.973	4.131	4.131
Stock-Yogo weak-identification test critical values	8.18 (15%)	8.18 (15%)	5.45 (25%)	5.45 (25%)
Over-identification Stat (Hansen J)	3.166	0.00005	0.572	5.001
Over-identification p-value	0.0752	0.994	0.449	0.0253

Notes: Robust standard errors clustered at the child level in parentheses. * significant at 10%, ** 5% and *** 1%. Additional control variables included are: household income, time interacted with age dummies, time interacted with gender, household size, number of males in the age-group 16-60, average age of household members, land owned by household in 2007 interacted with time, household's asset (index) in 2007 interacted with time, time trends depending on whether the district is NREGS Phase 1 district, mandal level total night lights in 2006, and baseline enrollment rate in the district. The same set of instruments (lagged rainfall shock, lagged amount fund sanctioned in NREGS, and their interaction) are used for annual household income and mother's working status.

Chapter 3

Does Access to Secondary Education Affect Primary Schooling? Evidence from India

3.1 Introduction

The second Millennium Development Goal (MDG) of the United Nations aims at universalization of primary education by 2015. Despite best attempts and significant rises in enrollment in most developing countries, recent findings suggest that the target is unlikely to be met.[1] While access to primary schooling has improved substantially, high dropout rates are still a critical problem.[2] In India, the major public policy initiatives like *Sarva Shiksha Abhiyan* (Education for All), the provision of a midday meal, free textbooks, uniforms etc. aim to universalize elementary education and reduce disparity across regions, gender and social groups. In this context, two suggestions have been popular: to reduce the access barrier through the provision of community-based primary schools; and to improve the quality of schools in terms of physical infrastructure, teacher quality etc. This paper raises a third issue: Does the pos-

[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 percent in 2000 to 89 percent 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 the proportion of pupils in India starting grade 1 who reach last grade of primary was only 68.5 percent 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 was 9.4 percent.

sibility 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 the human capital of children is crucially linked to the perceived economic returns to education (Manski, 1993; Nguyen, 2008; and Jensen, 2010). The received wisdom from studies conducted in the past two decades is that the returns to education are concave, i.e., the marginal effect of an increment in the number of years of primary schooling is larger than the effect at higher levels (Psacharopoulos, 1994 and Psacharopoulos and Patrinos, 2004). However, several recent studies show that the private returns to each extra year of education, in fact, rise with the level of education. A recent paper by Colclough et al. (2010) refers 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 have also shown similar results (Kingdon, 1998; Duraisamy, 2002 and Kingdon and Unni, 2001). On the other hand, some other studies show that while the actual rate of return to primary schooling is high, parents believe that the first few years of schooling have lower returns than in the later years (Banerjee and Duflo, 2005 and 2011). These observations suggest that households may perceive education investment as lumpy; that significant economic returns require continuation to at least high school - 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 reduction 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.

This paper aims to empirically test the hypothesis that access to secondary schools affects primary schooling outcomes. Although the importance of access to schooling in determining educational outcomes has been well recognized in the literature (Duflo, 2001; Glick and Sahn, 2006; Filmer, 2007; Orazem and King, 2007), most of it is on access to primary schooling and its effect on children. Therefore a major supply side intervention for policy makers has been to increase the availability of primary schools to the community to encourage more children to go to school. One commonly used measure of access is distance to school. Different studies have found that a reduction in distance to school improves enrollment, reduces dropout and improves test scores (Lavy, 1996; Bommier and Lambart, 2000; Brown and Park, 2002; Handa, 2002; and Burde and Linden, 2012).

Most of the literature addressing access examines the linkage between schooling outcomes at a particular level with access to that level of schooling. This is true even when studies look at access to secondary schooling. For example, Muralidharan and Prakash (2013) look at the impact of an incentive to provide cycles to girls going to secondary schools on their enrollment. Similarly, Pitt et al. (1993) look at the effects of secondary schools on enrollment of children aged 10-14. This literature interprets the provision of schools as a decrease in 'distance cost'. However they focus on the current cost of schooling. Nonetheless, difficult access to post-primary schooling that reflects the future cost of schooling hinders the possibility of continuation to a higher level of education and can hence adversely affect schooling decisions even at the primary level. Only a few studies acknowledge the importance of access to post-primary schooling in determining outcomes at the primary level.

Lavy (1996) uses a cross-sectional data-set for rural Ghana and shows that distance to post-primary schools negatively affects primary school enrollment. He suggests that the effective fees for post-primary schooling should be reduced to induce more participation at the primary level. Results similar to Lavy (1996) have been found in other cross sectional 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-sectional data-set 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 are low, as is the case for girls in Pakistan, access to it has no effect on primary school outcomes. Almost all these studies base their analysis on cross-sectional data, hence they are unable to control for time-invariant unobserved heterogeneity at the household level. Moreover, apart from Lavy (1996), there is no discussion on endogenous placement of schools, which can be a potential source of bias in the estimates. Lavy (1996), however, does not discuss placement of secondary schools as they are considered to be very far away from the sampled households. Andrabi et al. (2013) also point out to another mechanism that may link secondary schooling access and primary schooling. They posit that secondary schools may create private school teachers in the future, hence affecting primary schooling outcomes in the future. This is inherently a long run mechanism: their paper looks at data 17 years apart. In this paper the mechanism we look at is much more short term and hence completely different.

This paper uses a household level longitudinal survey of 43 villages in Uttar Pradesh, a state in India, to test the hypothesis that better access to secondary education increases enrollment and attendance among children in the primary school going age group. The main issue that potentially confounds such an analysis is when secondary schools open up as a response to increasing demand for schooling. An exploratory analysis of what determines better access to secondary schools finds that baseline village level demand signals correlate weakly with access to secondary schools. The general conclusion of our analysis is that poorer villages benefit from the development of a larger region which attracts new secondary school. The most significant determinant of opening up of new secondary schools is the baseline distance to secondary school itself: new schools tend to locate near places which did not have a secondary school before. Thus, there is a convergence in secondary school access. This paper therefore argues that given this environment and results, endogeneity of secondary school location based on village level demand is less of a concern in this context.

We use a fixed effects regression to determine the impact of better access of secondary schools on primary school enrollment and attendance. The paper finds that there is indeed a positive effect of better access to secondary education on primary school participation. This result is put to robustness checks. To provide further proof that endogeneity of schooling access does not drive our results, we use methods developed by Altonji et al. (2005) and Oster (2013) to show that our estimates, if anything, are an underestimate of the true effect. The results also do not change even after we account for additional trends based on baseline characteristics. We also show that the impact of secondary schooling does not affect primary enrollment through sibling externality. To elaborate, we show that the results are not different when one takes into account elder siblings who go to secondary or higher schools.

We find that the effect hypothesized is heterogeneous: the marginal effect is larger for smaller villages and for villages closer to bus stops. The effect is robust for poorer households and boys (who are more likely to enter the labour force). We also find a larger marginal effect for households that invest in health through immunization. Stratifying the sample by age cohorts, the paper finds significant effects on the enrollment of younger children and on both enrollment and attendance of older children. Using a nationally representative survey for India (National Sample Survey 2007-08), we also provide some suggestive evidence that this effect may be quite widespread.

This study contributes to the existing literature by extensively examining how developments at higher levels of education influence decisions at much lower levels. We point out to the lumpiness of human capital investment: that education may be attractive when children can have a less costly way of acquiring education all the way to secondary school. This backward linkage from secondary school access to primary schooling outcomes is very different from the question that most of the literature addresses: the static view of whether primary (secondary) school access affects primary (secondary) schooling. Among papers that do address a question similar to one asked in this paper, our paper provides a tighter research design employing household (hence, village) fixed effects techniques. Moreover, we attempt to validate our result by providing a battery of robustness checks that are consistent with our hypothesis that it is the continuation possibility that seems to drive our results and not other confounding factors.[3]

[3]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

The paper is organized as follows: Section 3.2 gives a background on Indian schooling system. Section 3.3 describes the data used for our analysis. Section 3.4 explores the relation between changes in secondary schooling access and baseline characteristics. Section 3.5 explains the empirical methodology. Section 3.6 reports the results obtained from the main regressions. In section 3.7, we show that our results are robust. Section 3.8 investigates if there are other pathways that explain our results. In section 3.9, results are provided that show that the impact of secondary school access is heterogeneous. Section 3.10 uses the NSS data to check whether our hypothesis can be extended at the all India level. The final section discusses the conclusions of the paper.

3.2 Background

The Indian school administration classifies grades I to V as primary schooling, VI to VIII as middle schooling (also referred to as upper primary schooling) and grades IX and X as secondary schooling. The final two grades of schooling, XI and XII, are called higher secondary though in some states, this differs with children going to intermediate college instead of class XII. Over the period starting from 1990s, India has taken great strides in enrollment in primary schooling. Primary enrollment rates among children aged 6 to 10 years have risen from a low base of 53 percent in 1986-87 to 72 percent in 1995-96 and then further to around 90 percent in 2007-08 (authors' calculation from mutliple rounds of the National Sample Survey data on Education). This growth has been most substantial for the rural society with 20 percentage point rural-urban difference in 1986-87 reducing to only a 4 percentage point difference in 2007-08. Attendance rates, from the same data source, suggest a similar rise over time and convergence between rural and urban areas. However, while students join primary school, often they don't complete school and this problem is especially acute in rural areas. To elaborate further only 22 percent of children aged 16 and above were enrolled in senior secondary schools.

at the secondary stage. Following *Sarva Shiksha Abhiyan*, a national mission on *Rashtriya Madhyamik Shiksha Abhiyan* (RMSA), or universalization of secondary education, has been set up.

The rise in primary enrollment has gone hand in hand with an increase in number of primary schools. The number of primary schools in India have grown from 590421 in 1995-96 to 787827 in 2007-08, an annual growth rate of 2.4 percentage points.[4] However, at the same time secondary schools have grown even faster, increasing from 71065 to 113824 over the same period, an annualized growth rate of 4 percentage points. This fast rise in the number of secondary schools, from a low base, lays the background for the question asked in this paper.

3.3 Data

To test our hypothesis, we provide evidence using data from a longitudinal follow up of households first surveyed as a part of the World Bank's Living Standards Measurement Study (LSMS) in Uttar Pradesh, a state in India (this survey is also called the Survey of Living Conditions, or SLC). This is a two-period panel data on rural households in 43 villages selected from nine districts in eastern and southern Uttar Pradesh.[5] The baseline data was collected in 1997-98 under LSMS. The second round of data was collected in 2007-08 by the authors; the data collection was funded by the University of Oxford and the World Bank.[6] The survey comprised a village questionnaire that contained information on, among other things, access to the nearest primary and secondary school and a household questionnaire that had detailed information on various aspects of standard of living, including the schooling status of each child in the household.

For the purpose of our study, we use information on households with children in the 6-10 age group. There were 705 children from 402 such households in 1997-98 and 566 children from 346 households in 2007-08. Considering both 1997-98 and 2007-08, we have observations on 1271 children from 531 households in our analysis.[7]

[4] `http://mhrd.gov.in/sites/upload_files/mhrd/files/statistics/SSE1112.pdf`

[5] Uttar Pradesh is usually considered one of the most backward regions in the country. Our sample includes 9 districts.The number of villages in each district vary from 3 to 6 per district.

[6] Both surveys were conducted during the same time of the year - December to April. The data collected in 1997 was verified to the extent possible during the second survey. Household attrition rate for the sample was 12.16 percent.

[7] *Table* 3.*A*1 summarizes the main variables in each time point. We discuss some of the key changes in the

Our two variables of interest are children's school enrollment and attendance rates. The SLC has detailed information on attendance. It reports the number of days in the past 7 days that the child attended school.[8] Attendance rate is defined according to whether a child has attended school for at least three days in the last week. According to the summary statistics, while enrollment rate was 69 percent in 1997-98, it had increased to 83 percent by 2007-08. Based on our definition, while attendance rate was 64 percent in 1997-98, it had risen to 81 percent by 2007-08.[9]

We use information on distance-to-schools that have been collected during the village survey. Distance to the nearest primary, middle and secondary schools is measured from the village centre and reported in kilometers. The average distance to the nearest primary school has reduced from 0.69 km to 0.08 km and to the nearest middle school from 2.54 km to 1.36 km. However, because both primary and middle schools were already very close on average to the villages in 1997-98, the change in them is relatively modest. The change in the distance to the nearest secondary school has been spectacular, though, the average distance has reduced from 6.25 km to 3.72 km (*Figure* 3.1).[10]

Pooling the data from both rounds, we find a significant and negative correlation of -0.15 between enrollment rate at the primary level and distance to the nearest secondary school. However, this correlation coefficient does not take into account any other confounding factors that may have changed over time. Moreover the placement of secondary schools could be endogenous. Therefore, in the next section, we explore further what characteristics correlate with the change in secondary schooling access.

---

discussion that follows.

[8]Consistent with the baseline survey, 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 not to change questions for the sake of uniformity.

[9]The rise in enrollment rate has been higher for girls (from 64 percent to 84 percent) as compared to boys (from 75 percent to 82 percent). Attendance rates show a similar trend.

[10]*Figure* 3.2 shows the histogram and kernel density of the change in distance to secondary school (distance in 1997-98 minus the distance in 2007-08).

3.4 Determinants of Secondary School Access

The placement of secondary schools is not random. Since we are interested in identifying the causal effect of distance to secondary school on primary level enrollment, the non-random placement of secondary schools can be a matter of concern. There can be specific factors which affect both the placement of secondary schools and primary level enrollment. If such confounding factors are not taken into account, then the estimated effect can be biased and inconsistent. For example, places which are educationally advanced have higher demand for education, thus higher enrollment rates. In response to demand, new secondary schools may also open up in these places. Hence, the effect can be overestimated. Consider another situation where a motivated regional leader drives the ambition of people, resulting in higher participation in education, and also promotes the supply of education by opening new schools. We try to capture such effects by including district level trends in the regression. The foundation of an empirical model that can be employed in this context relies on understanding the determinants of secondary school access. Therefore, we explore next how the change in distance to the nearest secondary school is correlated with baseline characteristics. We explore these correlations using a set of regressions in *Table* 3.1. For each village, we define the change in distance to the nearest secondary school as the distance in 1997-98 minus the distance in 2007-08. Thus positive changes reflect improvements in access. Since distance to the nearest school is defined at the village level, therefore we use observations on 43 villages for this analysis. Given the small sample size, we choose a parsimonious specification where we include the baseline distance to secondary school as a regressor and add different village/region level baseline covariates one at a time.

We find evidence of convergence in secondary school access: villages where the baseline distance to secondary school was higher, have experienced greater change in the following years. After controlling for the distance to secondary school, the coefficient of distance to the nearest primary school is negative and significant. Therefore, secondary schools tend to open up in villages that had a primary school close by. Since children who complete primary level are to be enrolled in secondary school, this result indicates that secondary schools may look for some coarse indicator for demand, reflected by the existence of primary school near the village. We do not find any significant correlation between change in distance to secondary school with the quality of existing public primary schools, captured by tercile-dummies based on a quality index.[11] There is also no significant correlation between change in secondary school access and distance of village to infrastructure facilities.[12] Secondary schools often cater to an area much larger than a village. They are often 3-4 km away from the centre of villages. To take this into account we consider a community larger than just the village. Using satellite data on light density at night, we calculate the average luminosity of the region with a radius of 10 km around the centroid of the village.[13] Night lights reflect local economic development, hence this measure can be used to proxy for the income of a region. Result in column (5) indicates that larger positive changes in distance to secondary schools have happened where the baseline income was higher. These indicate that more secondary schools have opened up in "richer" areas. However, when we include instead the average land-ownership of villagers, we find a negative and significant correlation. Results from column (5) and (6) together suggest that secondary schools respond to the demand from a larger area instead of individual village level demand. Hence poorer villages benefit from the prosperity of a larger area and tend to catch up in terms of access to secondary school.

The state under study had also initiated measures to increase secondary school enrollment. During 1997–2001, a state level policy aimed at giving monetary incentives to open girls only secondary schools in blocks where there was no such secondary school. Under the scheme, any

[11]The features of school quality considered in the analysis are type of structure, main flooring material, whether the school has classrooms, number of classrooms, whether the classes are held inside classrooms, whether the school has usable blackboards, whether desks are provided to the students, whether mid-day meal is provided and the proportion of teachers present on the day of survey.

[12]We construct an index variable by principal component analysis considering the distance to the nearest telephone booth, police station, public distribution shop and bank.

[13]The data on night-time luminosity is recorded worldwide for every one square kilometre area (approximately) by the Operational Linescan System (OLS) flown on the Defense Meteorological Satellite Program (DMSP) satellites. This data has been downloaded from the website of the National Oceanic and Atmospheric Administration (NOAA) of the USA (`http://ngdc.noaa.gov/eog/dmsp/download_radcal.html`).

new private girls only secondary school opened in a hitherto uncovered block headquarter, was made eligible for a one-time infrastructure grant of Rs. 1 million (Jha and Subrahmanian, 2005). Once this led to the opening of girls secondary schools, boys were allowed into these girls schools in order to make them cost-effective. Thus, this policy may have led to opening of new secondary schools in places which were lagging behind in the past. To investigate if the incentive scheme introduced by the government indeed correlates with the changing access to secondary schools, we include a dummy variable: whether the block had any single sex girls secondary school at the baseline.[14] We do not find any significant correlation between this variable and the change in distance to secondary school. To further explore whether secondary schools respond to aggregate demand, we include logarithm of village population. From the perspective of secondary schools, another important source of demand could be the number of children who have passed primary school, and children who are enrolled in primary school. Besides, the number of children below 5 years of age are currently not in any school but they would be among the future clients for new secondary schools. Therefore, we consider the logarithm of these variables in the regression.[15] It is found that the change in distance to secondary school is not significantly correlated with any of these variables.

From this exercise we find limited evidence for the impact of baseline village level demand on change in access to secondary school. The baseline distance to secondary school itself turns out to be the most robust and significant correlate of the changes in later years. Following this result, in our main regression, we also explore how the effect of changing distance to secondary school on primary school enrollment may vary with the baseline distance.

An important caveat of the analysis presented in this section is that we have observations on only 43 villages; therefore the regressions may lack enough power to identify the effect of some of the variables. We address this issue by explicity taking these factors into account in our regressions. Moreover, in a section on robustness, we use the literature that studies the

[14]Data for Girls Secondary Schools at the block level are sourced from The 7th All India Education Survey (NCERT) conducted in 2002.

[15]We find similar results when we consider the variables in levels.

sensitivity of coefficients in the face of omitted variables, as developed by Altonji et al. (2005) and Oster (2013) to test for the robustness of our result. We also check for robustness by allowing for differential trends based on some of the baseline village and community factors that may potentially confound the coefficients of interest. We turn next to the empirical model.

3.5 Empirical Model

This section discusses an empirical model to test whether access to secondary schools increases schooling enrollment and attendance at the primary level. For ease of presentation, we refer to only enrollment in the empirical model; however, in testing, we consider both enrollment and attendance.

For each child c from household i living in village j at time t, let the S_{cijt} be an indicator of whether the child is enrolled in school ($S_{cijt} = 1$) or not ($S_{cijt} = 0$). In our description below, the subscripts are implicit.

The co-variates for explaining enrollment can usually be categorized into four groups: individual child level factors, household characteristics, school characteristics, and geographical characteristics. The gender of the child is captured by a dummy variable indicating whether the child is male ($Male$). In addition to the age of child, to allow for non-linear impacts of age on enrollment, we include the square of age. Birth order of the child, which often determines child-level education investments within household, is also included.[16] The education level of decision makers in the household, especially of the mother, has often been found in the literature to have an important impact on the educational outcome of children. Hence we control for a dummy variable reflecting whether the mother is literate. We also include a dummy variable capturing whether the household head is literate. Moreover, previous literature has also found that the education outcomes of children are better when the household decision

[16]We include child's birth order as a continuous variable in the regression. The results remain unchanged even if we include birth-order specific dummy variables.

maker is a woman. Hence we also include a dummy variable that indicates if the household head is a woman.

Among other household level variables, we include land size to control for wealth. Investment on children may depend on the employment status of the household head. Therefore we include a binary variable indicating whether the household head is employed.[17] We also control for household size. To allow for differential impacts across different castes, we include dummy variables that represent the social group the household belongs to (SC/ST: Scheduled Caste/Tribe, OBC: Other backward Castes; the omitted category is the other less disadvantaged castes). Similarly, we allow for different enrollment rates across different religious communities by including religion dummies; in particular, we create a dummy variable for Muslim households.

This paper concentrates on access to secondary schooling, a level that yields perceptible market returns. Our study found primary schools close to most villages but not secondary schools. Hence, in our specification, we allow for access to the nearest primary and secondary schools. We include a squared term of distance to secondary schooling to allow for non-linearity of this effect.[18]

Let us refer to the distance to nearest primary school as $PRIM$. Moreover, the vector of the linear and the squared distance to the nearest relevant secondary school is referred to as SEC. While the inclusion of $PRIM$ is standard in primary schooling regressions, the inclusion of distance to secondary schools is less standard.[19] We hypothesize that if access to the nearest secondary school is found to be statistically significant after controlling for primary school access, the hypothesis that the possibility of continuation plays a significant role in primary school enrollment will be credible. 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

[17]We find similar results when we include dummies which reflect various occupation categories of the household head.

[18]Given very small distances of primary schooling from the villages, we omit the square of distance to primary schooling from our specification.

[19]It must be pointed out here that inclusion of the quadratic term in distance is less common. However, we include them because we posit that marginal changes in distances matter more when the distances are less.

unlikely to enrol their children in primary schools if secondary schools are far away.

Recent literature on schooling has stressed the importance of the quality of schools. The quality of primary schools has been found to be significant in papers that investigate schooling outcomes. More crucially for our analysis, if secondary schools were present closer to villages where the quality of primary schools is good, then our estimators for the impact of access to secondary school would be inconsistent. Thus, we include quality of the village primary school (D_l^{TERC}) as a regressor.

We also control for the distance of the village to the district headquarter ($Dist^{HQ}$). It is plausible that villages closer to district headquarters have a better perception about the returns to education. Another channel through which the distance to district headquarters may affect primary schooling is through the quality of teachers that come to the nearby schools. It can be argued that villages near the district headquarters may have better qualified teachers - those who reside in district headquarters 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}). We account for the size of the village by controlling for logarithm of village population (Pop). Another variable included in our specification is the proportion of adult village members who are engaged in off farm activities (OFF_FARM). We use this variable, collected as a part of the village survey, to prevent any concerns of endogeneity that may emanate from the simultaneous choice of schooling and work at the household level. Apart from having a probable income effect, these activities may also need some level of education. Therefore we posit that greater exposure of the households in a village to off farm jobs may inform households about the benefits of education. Similarly, the average density of night lights ($Night\_light$) is included to consider that school participation may depend on the intensity of economic activity of a region. Demand for education may also depend on the remoteness of the village from various other institutions of development. Hence we include the index of distance to infrastructure facilities as another regressor ($Distance\_Infrastructure$). In spite of including a rich set of observable control variables, it is plausible that the presence or nearness of secondary schools is confounded by unobserved village heterogeneity. We therefore allow for village level fixed effects α_j.

To eliminate other confounding temporal trends, we include time (t) as a regressor. Moreover, we allow trends to vary by district ($\sigma_d * t$) as well as by nearness to district headquarters ($Dist^{HQ} * t$). These trend terms control for, among other things, changes in returns to education over time for the district as a whole, as well as for the village depending on how far it is from district centres.

Next, we take care of household level unobserved heterogeneity by including household fixed effects (α_{ij}).

Let I refer to the individual characteristics and H refer to the household level socio-economic characteristics. We estimate the following model:

$$\begin{aligned} S_{cijt} = {} & \alpha + \beta' I_{cijt} + \gamma' H_{ijt} + \pi_1 PRIM_{jt} + \pi_2' SEC_{jt} + \sum_l \zeta_l D_{ljt}^{TERC} + \rho Dist_{jt}^{HQ} \\ & + \sum_h \theta_h D_{hjt}^{ROAD} + \mu Pop_{jt} + \eta OFF\_FARM_{jt} + \varphi Distance\_Infrastructure_{jt} \\ & + \psi Night\_light_{jt} + \alpha_{ij} + \alpha_j + \sigma t + \sum_{d \in \{Districts\}} \sigma_d t + \lambda Dist_{jt}^{HQ} * t + \epsilon_{cijt} \end{aligned} \tag{3.1}$$

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 coefficient 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 enrollment. In all the reported regressions, we use robust standard errors that are clustered at the village level.

Given that secondary schools usually cater to bigger populations, the use of village dummies alone should lead to consistent estimation if we feel other co-variates are not correlated to α_{ij}. Lavy (1996) uses this identification strategy in his cross sectional analysis.[20] However, household-level unobservables could still lead to inconsistent estimation for reasons pointed

[20] In addition, Lavy (1996) finds that secondary schools are very far from villages. Therefore, he postulates that the distance to secondary school is not endogenous to individual village level demand.

out above. Therefore, we estimate the model including household fixed effects.

3.6 Main Results

Results from estimating the household fixed effects model are presented in *Table* 3.2. Column (1) and (3) report results for enrollment and attendance from a regression where the linear and quadratic terms for distance to secondary school are included. These specifications point out that while distance to primary school has no effect, distance to secondary school does. The relationship between the distance to secondary school and enrollment is negative but the marginal impact of a drop in distance is less at large distances. This implies that the marginal effect of a one kilometer decrease in distance to secondary school on enrollment, evaluated at the mean distance in 2007-08 (3.72 km), is 0.075. The marginal effect on attendance is slightly lower at 0.07. The insignificance of distance to primary school for both variables for this sample is not surprising because most villages already had a primary school nearby even in 1997-98, as shown earlier.[21]

Further, to investigate whether the marginal effect of access to secondary education varies with the initial level of access, we interact the distance to secondary school with dummies reflecting the initial distances in 1997-98. These regressions are presented in columns (2) and (4). Intuitively, one can posit that a unit drop in distance would not matter if the existing school is either very close to, or very far away from the village. Therefore, we categorize the baseline distance to secondary school into four groups: less than 3 km, between 3 to 6 km, between 6 to 10 km, and greater than 10 km. We interact the distance to secondary school with four dummy variables representing these categories. Results from column (2) and (4) show that the marginal effect of distance to secondary school is significant and highest when the initial distance was between 3 to 6 km. If the baseline distance was between 6 to 10 km, we find significant effect only on enrollment. The marginal effect disappears if the baseline

[21] The impact of addition of variables in the specification can be seen in *Table* 3.*A*2. The coefficient of distance to secondary school is always negative and significant, mostly increasing in magnitude as we add more controls.

distance is below 3 km or above 10 km.[22]

Among the other variables, being male has a positive effect on enrollment, but no effect on attendance. Age of the child has a positive effect on enrollment although this effect is mitigated at a higher age. Everything else turns out insignificant, barring the quality of primary school, the road access variables and village population. We find that a primary school of the highest quality (relatively speaking) in a village increases the probability of enrollment by 0.25. Its effect on attendance is not significant. We find that a better road increases the probability of both enrollment and attendance, though this increase is higher for katcha and paved roads.[23] Village population has a positive effect only on enrollment. The insignificance of the rest of the variables could be driven by low temporal variation. Given that the model controls for household fixed effects as well as district specific time trends, the remaining variation in these covariates may not be enough to yield significant coefficients.

3.7 Robustness

In this section, we subject our specification to robustness checks. First, we assess the extent of potential omitted variable bias due to unobservable factors in the model. Second, we consider the distance to middle school and check if it affects our results. Thirdly, we want to know if the effect of secondary school is sensitive to inclusion of additional trends. Finally, we show that our results are robust to alternate assumptions about the age-grade profile.

3.7.1 Assessment of Potential Bias from Unobservables

Our main regression model controls for a large set of observable explanatory variables including household fixed effects and district specific time trends. Nevertheless, if there is any unobserved

[22]Instead of the distance to secondary school, we also estimate the model including a dummy variable indicating whether there is a secondary school in the village or not. We find significant and positive effect of presence of secondary school in the village on primary level enrollment and attendance. These results are available on request.

[23]Presence of all weather roads (Pucca) does not come out to be significant in the regressions. This is somewhat puzzling; it could be caused by a slight misclassification between paved and "pucca".

time-variant factor which affects both the placement of secondary schools and primary school enrollment, then the estimated effect can be biased. In this section, we investigate the extent of such omitted variable bias following a strategy developed by Altonji et al. (2005). This methodology is based on the idea that selection on observables can provide a useful guide to assess the selection based on unobservables. Oster (2013) derives a more generalized version of this method which relies on the "proportional selection assumption" (PSA), i.e., selection on observables is proportional to the selection on unobservables. In the context of our model, PSA implies that the relationship between distance to secondary school and the observable explanatory variables is informative about the relationship between distance to secondary school and the unobservable factors. This link is represented by a degree of proportionality (δ). Furthermore, Oster (2013) incorporates the change in R-squared value in analyzing the movements of the coefficient of interest as more control variables are included in the regression. Thus, under the proportional selection assumption, she identifies the omitted variable bias and derives a consistent estimator of the causal effect. According to this method, we calculate the true effect of distance to secondary school on primary level enrollment and the potential bias due to unobservables. The estimate depends on the value of two parameters, δ and R_{max}, where R_{max} is the R-squared of the hypothetical regression which includes the complete set of controls including the unobservable variables.

In our analysis, distance to secondary school is defined at the village level in each year. Hence the possibility of endogeneity arises from the fact that this variable may be correlated with some unobservable factors which vary across villages and over time. Inclusion of village specific time trends would completely subsume the effect of such unobservable factors. Therefore, we estimate the model without all village level covariates but including village level trends and consider the R-squared from this regression (0.69) as a reasonable value of R_{max}. Similarly, any household level time-varying unobservable characteristics which could be correlated with both access to secondary school and enrollment, will be accounted for if we include household specific time trends in the regression. R-squared from this regression (0.81) is considered as another appropriate value for R_{max}.[24] Next, we need to assume the degree of proportionality, δ. It is plausible to conceptualize that $\delta \in [0, 1]$ if we consider that the observable variables are at least as important as the unobservables. A value of $\delta = 1$ will be in accordance with the assumption of equal selection on observables and unobservables (Altonji et al., 2005). In addition to the value of $\delta = 1$, we also relax this equality assumption and present results for $\delta = 0.5$ and $\delta = 1.5$ to check robustness of the effect.

The method developed by Oster (2013) allows for a single potentially endogenous variable, therefore, we include only the linear term of distance to secondary school (SEC) in the model specified in *Equation* (3.1). Even after excluding the quadratic term, we still find a significant negative effect of SEC on both enrollment and attendance, though the magnitudes are lower.[25] Results presented in *Table* 3.3 suggest that the effect estimated from our main specification is unlikely to be driven by unobservable factors. With the evidence of a very small bias, there is negligible change in the magnitude of the coefficient once we take care of the selection based on unobservables according to this method. Moreover, the direction of the bias indicates that our estimate of the negative effect of the distance to secondary school is a lower bound to the true negative effect. Therefore, this robustness exercise adds more credibility to our finding that a reduction in distance to the secondary school increases enrollment and attendance at the primary level. While we present the results with only the linear term for distance to secondary school, conceptually, this exercise holds true even if we include both the linear and the quadratic terms.[26] Inclusion of the quadratic term makes the model richer by taking into account the non-linearity of the effect. Hence we stick to our main model that considers both

[24]When we include village (household) level time trends, the model cannot identify the effect of distance to secondary school or any other variable which does not vary within village (household). However, our objective is only to estimate the R-squared of these two regressions so that they can be used as suitable values for R_{max}.

[25]When only the linear term for SEC is considered, we find that a one kilometre reduction in distance to the nearest secondary school results in a significant 3.6 (3.7) percentage point rise in the probability of enrollment (attendance) in primary schooling.

[26]The estimated bias is calculated using the covariance between the potentially endogenous variable and the added control variables. In our model, inclusion of the control variables moves the coefficient away from zero due to a positive covariance, suggesting that our estimated effect is a lower bound of the true effect. This phenomenon remains true even in the case where we consider both the linear and the quadratic term of the distance variable.

the linear and the quadratic terms for the rest of the paper.

3.7.2 Distance to Middle School

In this section, we include distance to the nearest middle school as an additional control variable to check whether it confounds our results. A priori, there are two reasons why we do not include this variable in the main regression. Firstly, middle schools are placed on average at a distance of 2 km from the villages. While secondary schools are located farther away and their placement is likely to be affected by the characteristics of a region larger than individual villages, placement of middle schools may be driven by village specific factors. Hence the distance variable for middle school is more susceptible to endogeneity concerns. Secondly, from the household's perspective, continuation to middle school may not be as meaningful as secondary level of education. Given the convexity of education-earning relationship, educating a child at the secondary level is likely to yield higher returns than middle level. Hence, access to secondary school may be more relevant in household's decision of sending a child to a primary school.

Results of the regressions including the distance to middle school are presented in column (1) and (2) of *Table* 3.4. We find that the coefficients of both linear and quadratic terms for distance to secondary school are not affected by the inclusion of distance to middle school. Moreover, distance to middle school has no significant effect on primary school enrollment or attendance. This supports the hypothesis that households consider continuation to at least secondary level of education for getting significant economic returns.

3.7.3 Additional Trends

In this section, we check the robustness of our results by taking into account additional trends that may be correlated with both the demand for schooling and placement of schools. There may be differential trends in demand for schooling depending on the initial level of educational attainment. The effect of proximity of secondary schools on primary school enrollment can be overestimated if villages with lower base-level education experience higher growth in both demand for and supply of schooling. Therefore, we consider the village level total number of children in the age group of 11-15 years who have completed primary education. Since this cohort should have already finished primary schooling at the time of the baseline, it is a predetermined variable reflecting the baseline demand for education in the village. We interact the logarithm of this variable with time and include it in the regression. Thus, the villages are allowed to follow different trajectories depending on their baseline demand for schooling. Similarly, one can argue that the growth in enrollment and attendance rate over time may vary with caste with the less privileged castes showing a greater increase in demand for education. We take this into account by including additional time trends which vary with the caste and religion of the households.

Furthermore, we consider another trend variable that depends on whether the block had a single sex secondary school for girls in the baseline. As discussed earlier, promoted by a government policy, more number of secondary schools may have opened up in blocks which did not previously have a girls' secondary school. Blocks which had a girl's secondary school in the baseline can be educationally more developed. Hence the growth in the demand for education can be different in these blocks in comparison with their counterpart. The trend variable we include will control for any such differential effect.

After controlling for these additional trends in the regression, we still find significant negative effect of distance to secondary school on primary level enrollment and attendance. The magnitude of the effects in this specification are in line with our main specification. The results for household fixed effects regressions are reported in column (3) and (4) of *Table* 3.4.[27] The blocks where a girls' secondary school existed before experienced a higher rate of primary school enrollment, but there is no effect on attendance. Other trends are not significant.

In addition to the regressions presented here, we also checked the robustness of our results

[27]The number of observations for these regressions is 1194 because for three blocks we could not find information on whether there was a girls' secondary school in the baseline.

by including trends based on logarithm of number of children in 0-5 years age, logarithm of number of enrolled children in 6-10 years age, and village level average land-ownership. We find that the inclusion of additional trends does not affect, significantly, the coefficient of either the linear or the quadratic term for distance to secondary school. Also, none of these trend variables is itself significant, indicating that our main model is likely to have already taken care of any endogeneity problem related to the placement of secondary schools. These results are available on request.

3.7.4 Age Gradient

Our analysis is based on children in the age-group of 6-10 years, since this age-group corresponds to primary level of schooling. Nevertheless, if some children enter school at an early (or late) age, or there is grade retention, then the age-band considered here may not be entirely appropriate for primary schooling. In this context, $Figure$ 3.3 shows the age-grade profile by gender. It indicates that both boys and girls in 6-10 years age-group attend primary level grades on an average. Also, we test the sensitivity of our results by considering the following age-bands: 5-10 (inclusion of 5 year olds), 6-12 (inclusion of slightly older children who may be attending primary schools), and 6-14 (as many studies on elementary schooling use this age group). Results from $Table$ 3.5 indicate that the effects are significant and similar for 5-10 and 6-12 age-groups. The effects decrease with inclusion of older children (6-14 age-group) as nearness to secondary schools will have limited impact if the older children did not enroll in primary schools to begin with.

3.8 Other Pathways

Next we conduct two exercises to rule out the possibility of other explanations for the results obtained.

3.8.1 Secondary Schools or Secondary Education

In this paper, we 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 school affects primary school enrollment and attendance. However, many secondary schools have classes 1 to 5. Secondary schools usually have better infrastructure than stand-alone primary schools. It may be that children go to a secondary school for their primary education as one comes up closer. In this case, our results would give further evidence, indirectly, to the importance of better quality.

To extricate the impact of the possibility of continuation to secondary education, we follow another strategy. We define another dependent variable to reflect whether a child in the 6-10 years age-group is enrolled in (attending) a school within the village. Further, we use a sub-sample of villages that do not have a secondary school within two kilometers in both periods.[28] In this scenario, if there is an increase in enrollment of children in schools within the village, this cannot be the outcome of their going to a secondary school because there is none in the village (or even within a two kilometer radius of the village).[29] In this case, the marginal effect on primary school enrollment must be driven by the perception that students can continue to secondary level. This is indeed the case (*Table* 3.6).

The marginal impact on the enrollment in the village primary school from a reduction in distance to secondary school turns out to be 0.14 (the mean distance to nearest secondary schools is 6.09 for such villages in 2007-08). In the case of attendance in the local village school, this impact is 0.168. Thus, for villages which do not have secondary schools, there has indeed been a considerable impact of secondary schools opening 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 of availing secondary education.

[28] This selection is based on distance, an independent variable. Therefore our estimators are still consistent.

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

3.8.2 Sibling Externality

Participation in primary school may have improved because children have elder siblings who have had better access to secondary schooling and are better educated. Their elder siblings may be teaching them 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 the proximity of secondary school has any direct impact on the primary schooling decision despite its effect through this channel. Therefore, we run the regression model introducing a new household level co-variate that captures the number of children in the 14-18 age group enrolled in secondary or higher school. 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 remain significant ($Table$ 3.7). The magnitude of the effects remains almost unchanged. It is important to note here that we recognize the possibility that decisions regarding primary and secondary schooling of children are determined simultaneously in a household. However, our objective is not to estimate the causal effect of siblings; rather, we seek to show that our main results remain unperturbed even if we account for a possible correlation between the education of primary school age children and their older counterparts.

3.9 Heterogeneity of Effects

In this section, we investigate the heterogeneity of impact of access to secondary schools. To begin with, we investigate the heterogeneity for child level variables. We decompose the sample into two age groups, 6-7 years and 8-10 years, to see whether the effect of secondary schooling has been more on the younger children or the older ones. Access to secondary schooling affects the two groups differently and the impacts on attendance and enrollment are different ($Table$ 3.8). The effect of a one kilometer decrease in distance to secondary school on enrollment is larger for the younger age group (marginal effect of 0.172) than for the older age group (marginal effect: 0.049). For the younger children, both the linear and square terms are not significant for attendance. However, for the older children, the effect on attendance is found to be significant: the marginal effect of a unit drop in distance to secondary school is 0.049. These results are consistent with our stated hypothesis. Households making decisions on whether to send their six year olds to primary school may not find it optimal to do so if they perceive that it is unlikely that the child will be able to go to secondary school, a level where there are economic returns. Once the child is not sent to school in his initial years, it is less likely that he/she will be enrolled in a primary school at a later age (8-10 years) in response to change in access to secondary school. Hence, the effect on enrollment is larger for the younger age group. The results on attendance suggest that the older age groups start to lose interest and attend school less if there is no possibility of future continuation. This may explain their subsequent dropping out and why many children stop going to school when they get older, despite higher enrollment rates for younger age groups.

Next, we examine if the effects vary by gender. While distance to secondary schools affects enrollment for boys, there are no effects for girls (*Table* 3.8). This differential impact has two explanations.

First, recall that we hypothesize that the effect of secondary schooling comes from the possibility to earn economic returns after schooling. However, in rural areas, it is less likely for educated women to seek jobs.[30] Therefore, the effect of continuation seems to be restricted to boys. Second, it is possible that distances to secondary schools are still far, although they have gone down. It is well known that parents do not send girls too far from the village. Therefore, it may be the case that the distances are such that this effect has not kicked in for girls.

Next we consider heterogeneity based on household and village specific characteristics. For this part, we report results only for enrollment in *Table* 3.9, since the effects on attendance are very similar.

[30]This was also pointed out by Hazarika (2001).

Distance to school has implication on both the direct and indirect cost of schooling. A greater distance implies higher travel cost. The opportunity cost is also higher if it requires more time to travel to school. When schooling is costly, wealthier households may still be able to afford it. However, poorer households may be constrained and unable to pay for a costly education. Therefore, when the distance barrier is reduced, it can be more beneficial for the poorer households. In other words, it is likely that a change in access cost will have a larger effect on poorer households. 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 for the ones above it. The results suggest that while we get significant effects for both groups, the magnitude is higher for households below the median level of landholding. The marginal effect of a one kilometer decrease in distance - on enrollment - for poorer households (evaluated at 3.18 km, the mean of the sample in 2007-08) is 0.116 (for attendance: 0.092). The corresponding marginal effect for richer households (evaluated at the sample mean of 4.03 km) is lesser at 0.021 (0.039 for attendance). This result is intuitive as the cost of continued schooling binds most for poorer households. Hence the impact of reduction in distance must be higher for them.[31]

Empirical studies have often cited a strong connection between health and education (Miguel and Kremer, 2004). In our data set, we do not have anthropometric information on children for 1997-98. In the absence of that, we use the immunization record of households for children aged 0-5 collected in 1997-98.[32] We divide the households into two groups: those where all the children in the relevant age group were immunized and those where some of the children were not immunized. We find that in households where all children were immunized, the marginal effect of a unit decrease in distance on enrollment (evaluated at the sample mean

[31]It would have been interesting to look at the impact of decreasing distance of secondary schools for various social groups. However, because of the small sample size, we are not able to run a household fixed effects regression for the General Category. The households in our sample are primarily from Scheduled Castes and Other Backward Castes in rural Uttar Pradesh (representative of the composition of the region). When we run our regression on this sub-sample, 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 0.083 (results available on request).

[32]The survey asks if the children received any immunization, for example, for Polio, DPT, Measles.

of 3.8 km) is 0.082 whereas that for households where not all children were immunized (evaluated at the sample mean of 4.01 km) was 0.04. The difference in the effect on attendance is even larger. The marginal effect is 0.082 for the households where all children were immunized where as it is 0.017 in the case of the other group. This difference may have two linked explanations. First, households where all children were immunized may have a higher preference for health. And since preference for health and education are often linked, the higher marginal effect for these households may reflect that. A second plausible reason could be that the health of children in households which have traditionally immunized children is better. Therefore, they respond to changes in education opportunities better than those whose health is not as good. Although without further information it is not possible to idenfity which of the potential channels is at play here, yet this finding indicates an interesting link between health and educational investment.

Next we investigate if the impact of better access to secondary schools varies by the size of the village. We define small villages as ones where the number of households are less than 200 in the base year and large villages where the households are more than 200.[33] We find that the marginal effect of a decrease in the distance to secondary school is positive for both small and large villages. However, the impact of closer secondary schools is more prominent on small villages than on large ones. A unit drop in distance to secondary schools raises enrollment rate by 0.172 and the attendance rate by 0.123. But the marginal effects in large villages are relatively smaller: 0.088 and 0.081 respectively for enrollment and attendance. These larger marginal effects for smaller villages reassure us that our results are not driven by the opening of secondary schools in response to absolute size of demand. Rather, smaller villages were further off from secondary schools in the base year and, therefore, have the most to gain from changing proximity to secondary schools.

We investigate whether the impact of access to secondary schools varies with transport infrastructure by using a formulation where we interact a dummy variable that measures

[33]We run separate regressions for the two classes of villages as all coefficients may be very different across the two classes.

if there is a bus stop within 1.5 km radius of the village with the distance to the nearest secondary school.[34] We find that the effect of a decrease in distance is increasing if there is a bus stop (last column of *Table* 3.9). The marginal impact of a reduction in distance if there is a bus stop is 0.104 on enrollment (0.112 for attendance) and 0.072 (0.065 for attendance) when there is no bus stop. This result makes sense since the distance to the nearest secondary school measures the perception of how difficult it is to get to the next level of schooling. Insofar as most children would need a bus to go to these schools, our results point out that the impact of continuation kicks in when there is a complementary bus service. It therefore points out the need for developing infrastructure to reap these benefits.

3.10 Extending Hypothesis to a Nationally Representative Survey

It may be contended that our results are specific to the small part of India that our sample represents. Therefore, in this section, we provide additional evidence that suggests that our (qualitative) results are general. We do so by analyzing the same effect using a cross sectional sample for rural India collected by the National Sample Survey. The National Sample Survey Organization (NSSO) conducts nationally representative household surveys on "participation and expenditure in education" once in 10 years. We use the education survey (64th round) by NSSO conducted in 2007-08.[35] Consistent with the analysis above, we restrict our analysis to children aged 6-10 years. The average enrollment rate is 87.2 percent while the average attendance rate is 86.8 percent.[36] The access to secondary schooling is heterogeneous across

[34]In most regressions, we stratify our regressions by variables defined at the base year, However, in the case of access to bus stop, we use an interaction term. The logic for this is that while village size classifications are sticky over time (small villages remain small in both periods), access to bus stop is not and changes over time. Hence, conditioning on the access to bus stop in the base year may not correctly reflect reality in the latter year.

[35]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.

[36]The enrollment (attendance) figures for boys and girls are 88.7 (88.5) and 85.4 (85) percent respectively.

rural India.[37] While 27 percent of the children have access to a secondary school within 1 km, for 21 percent of them the nearest secondary school is located at a distance of more than 5 km.[38] Conditioning for other standard covariates for which data is available[39], an Ordinary Least Square regression of enrollment on distance to secondary school finds that secondary schools have no discernible differential impact on primary schooling enrollment if they are at a distance of one to five kilometers than when the school is within a one kilometer radius (the reference category). However, if secondary schools are more than five kilometers away from the household, then this is associated with a 3.3 percentage-point lower probability of enrollment in primary level. These results are reported in *Table* 3.10. While these cross sectional results are only suggestive, they seem to indicate that the negative partial correlation between distance to secondary schools and primary enrollment/ attendance may hold for most of rural India.

3.11 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 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 literature on the economics of education that shows that the perceived (real, in many cases) 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

[37]NSS data validate our claim that a majority of the population have easy access to primary schools. The nearest primary school is at a distance less than 1 km for 91 percent of the households.

[38]Distance is reported in terms of intervals in the National Sample Survey.

[39]We consider individual characteristics: age, square of age, gender; household characteristics: A dummy variable representing whether the household head is literate, a dummy variable representing whether the household head is female, household size, dummy variables representing land categories, dummy variable representing whether the household is a Muslim, dummy variables representing social groups. We control for sub-regional differences by using dummy variables.

from education, households have to invest in their child's education until they pass high school. Households take into account the cost of post-primary schooling in making decisions at even the primary level. A major component of the cost of post primary schooling is distance to secondary schools. This paper explores whether access to secondary schools affects primary schooling.

We estimate the significance of the hypothesized 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 first explore on baseline characteristics that correlate with positive changes in secondary schooling access. Our analysis points out to the importance of the development of a region much larger than the village. Further we find that new secondary schools have located in places to which a secondary school was far away in the baseline. This points out to a convergence in schooling access. Given these results, we find that the distance to the nearest secondary school is indeed a significant determinant of primary school enrollment and attendance. The marginal effect on the probability of enrollment, from a one kilometer decrease in distance to secondary school, evaluated at the mean distance in 2007-08 (3.72 km) is 0.075. The marginal effect on attendance is slightly lower at 0.07. These results are evaluated using methods developed by Altonji et al. (2005) and by Oster (2013). Results suggest that the omitted variable bias would make the hypothesized effect even stronger. Furthermore, we find that that our results are robust to inclusion of trends that depend on baseline covariates that may affect secondary schooling access.

Next, our paper also finds that the impact of secondary schools is driven by the possibility of continuation and not merely because secondary schools may provide better quality education at the primary level. In villages that do not have secondary schools, the marginal impact on enrollment in the village primary school from a reduction in distance to secondary school turns out to be 0.14 (the mean distance to nearest secondary schools is 6.09 km for such villages in 2007-08). The impact on attendance in the local primary school is 0.168. This suggests that the effect on primary schooling outcomes is driven by continuation possibilities. We also rule out the possibility of other pathways confounding the effect. For example, it is not the case that the effect disappears if we account for the fact that children studying in secondary schools mentor their younger siblings to go to primary schools.

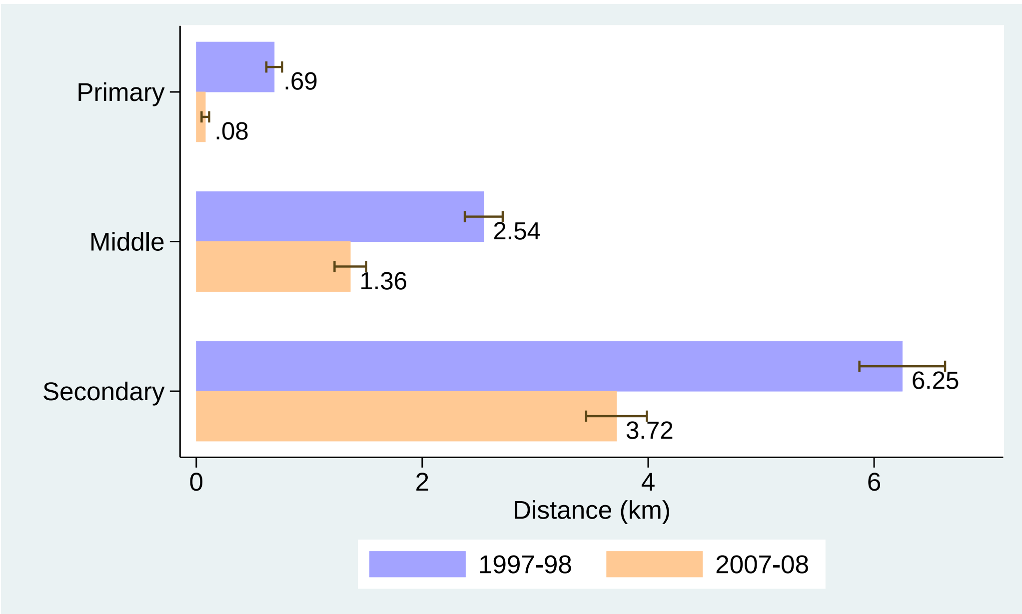
We find that the impact of secondary schools is heterogeneous. The impact is greater when there is a complementary bus stop close to the village and the smaller the villages in the baseline survey, the greater the effect. We find that households that lie in the bottom half of the baseline land distribution are affected more. We also find that households that have immunized all their children react more to the decrease in distance. This suggests that the ability to respond to better schooling opportunities depend on health outcomes of children. Interestingly, we find that the effect is larger for enrollment of children aged 6-7 years; however, for the children aged 8-10, there is a significant effect on attendance. We find the effect is significant for boys but not for girls. This is again consistent with our hypothesis since work participation rate among men is larger than for women. So men are more likely to reap economic benefits from reaching secondary schools.

While these results are obtained for a sample of 43 villages in a poor part of India, our hypothesis may be much more general. Despite the omission of some key variables in a nationally representative survey (National Sample Survey 2007-08), we run a simple regression to show that there is a negative correlation nationally between access to secondary schools and primary schooling enrollment (and attendance). Controlling for distance to primary schools, we find that if secondary schools are more than five kilometers away, there is a 3.3 percentage-point decrease in primary level enrollment.

In light of these results, our paper suggests that access to post-primary schools is 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, this critical mass of primary educated children may never develop in many parts of the developing world. In the absence of continuation possibilities, households may pull their children out of primary schools. Thus, all levels of schooling need to be developed and accessible at the same time to achieve universal education.

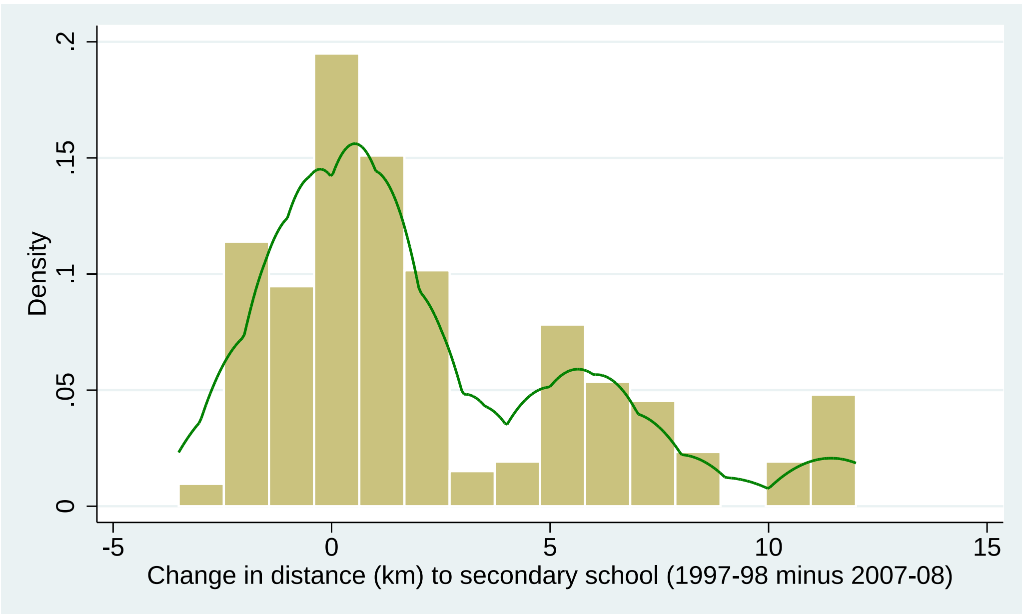
Figures and Tables for Chapter 3

Figure 3.1: Average distance to nearest school from the village



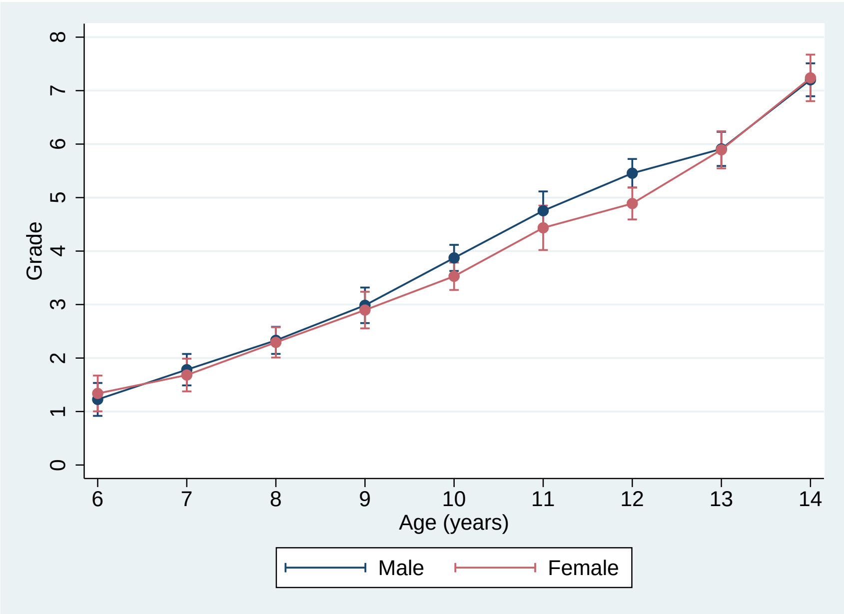
Source: SLC data.

Figure 3.2: Histogram and Kernel Density Estimates of change in distance to nearest secondary school



Source: SLC data.

Figure 3.3: Age-grade profile by gender



Source: SLC data.

Table 3.1: Baseline regression of change in distance to secondary school

	Dependent Variable: Change in distance to secondary school (1997-98 minus 2007-08)										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Distance to secondary school	0.425***	0.440***	0.437***	0.502***	0.477***	0.602***	0.433***	0.426***	0.424***	0.424***	0.440***
	(0.096)	(0.097)	(0.097)	(0.099)	(0.096)	(0.099)	(0.096)	(0.107)	(0.105)	(0.106)	(0.103)
Distance to primary school	-0.472*										
	(0.246)										
School quality tercile 2		-0.424									
		(0.786)									
School quality tercile 3			2.364								
			(1.630)								
Distance to infrastructure facilities				-0.424							
				(0.362)							
Night lights					0.445**						
					(0.207)						
Average land ownership in village						-0.682***					
						(0.211)					
Baseline girls' secondary school							-0.092				
							(0.855)				
Log of village population								0.417			
								(0.682)			
Log of total 0-5 year olds in village									0.399		
									(0.524)		
Log of total enrolled 6-10 year olds in village										0.459	
										(0.570)	
Log of total primary pass 11-15 year olds in village											0.286
											(0.366)
Constant	-0.297	-0.651	-0.943*	-1.082**	-2.019***	0.186	-0.805	-3.799	-2.934	-3.027	-2.028
	(0.585)	(0.584)	(0.516)	(0.528)	(0.664)	(0.595)	(0.789)	(4.918)	(2.808)	(2.841)	(1.672)
Observations	43	43	43	43	43	43	40	43	43	43	43
R-squared	0.454	0.435	0.463	0.451	0.497	0.566	0.441	0.438	0.441	0.441	0.440

Notes: Robust standard errors are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.2: Effect of distance to secondary school on primary school participation

	Enrollment		Attendance	
Variable	(1)	(2)	(3)	(4)
Distance to primary school	0.043	0.022	0.045	0.031
	(0.038)	(0.043)	(0.038)	(0.042)
Distance to secondary school	-0.143***		-0.127***	
	(0.024)		(0.037)	
Square of distance to secondary school	0.009***		0.008**	
	(0.002)		(0.003)	
Distance to secondary school * Base distance ≤ 3		-0.064		-0.051
		(0.044)		(0.044)
Distance to secondary school * Base distance $\in (3, 6]$		-0.091***		-0.096***
		(0.033)		(0.031)
Distance to secondary school * Base distance $\in (6, 10]$		-0.045*		-0.034
		(0.023)		(0.027)
Distance to secondary school * Base distance >10		0.032		0.009
		(0.033)		(0.036)
Male	0.064**	0.063**	0.039	0.038
	(0.031)	(0.031)	(0.036)	(0.036)
Age	0.282*	0.300**	0.259	0.278*
	(0.140)	(0.140)	(0.156)	(0.156)
Square of age	-0.015*	-0.016*	-0.013	-0.014
	(0.008)	(0.008)	(0.010)	(0.010)
Birth order	0.001	0.001	-0.013	-0.013
	(0.027)	(0.027)	(0.027)	(0.027)
Mother literate	0.031	0.037	0.039	0.040
	(0.075)	(0.077)	(0.083)	(0.085)
Head literate	0.068	0.057	0.045	0.035
	(0.130)	(0.133)	(0.132)	(0.135)
Head female	0.133	0.116	0.122	0.105
	(0.158)	(0.156)	(0.161)	(0.159)
Head employed	0.030	0.026	0.029	0.022
	(0.098)	(0.102)	(0.107)	(0.111)
Household size	-0.010	-0.009	-0.013	-0.013
	(0.011)	(0.011)	(0.011)	(0.011)
Land owned (acres)	0.002	0.002	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
School quality tercile 2	0.078	0.101	0.044	0.033
	(0.075)	(0.105)	(0.083)	(0.103)
School quality tercile 3	0.250**	0.265	0.198	0.165
	(0.121)	(0.182)	(0.120)	(0.190)
Off farm employment rate	0.003	0.002	0.003	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Road access - Katcha	0.485***	0.371**	0.504***	0.353**
	(0.107)	(0.147)	(0.125)	(0.161)
Road access - Paved	0.509***	0.306*	0.455**	0.267
	(0.154)	(0.163)	(0.173)	(0.160)
Road access - Pucca	0.187	0.094	0.170	0.061
	(0.117)	(0.130)	(0.113)	(0.122)
Log of village population	0.210*	0.206	0.106	0.089
	(0.114)	(0.147)	(0.124)	(0.141)
Distance to infrastructure facilities	-0.016	0.042	0.020	0.082
	(0.051)	(0.069)	(0.066)	(0.069)
Night lights	0.030	0.011	0.025	0.003
	(0.022)	(0.029)	(0.028)	(0.033)
Distance to district headquarter * Time	-0.001	-0.002	-0.003	-0.005
	(0.003)	(0.004)	(0.003)	(0.005)
District Specific Time Trends	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,271	1,271	1,271	1,271
R-squared	0.676	0.672	0.674	0.672

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.3: Assessment of potential bias due to unobservables

	Coefficient of Distance to secondary school							
	Uncontrolled (R^2)	Controlled (R^2)	Identified [Estimated Bias]					
			$R_{max} = 0.69$			$R_{max} = 0.81$		
			$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Enrollment	-0.0141	-0.0361	-0.0365	-0.0368	-0.0372	-0.0385	-0.0409	-0.0433
	(0.02)	(0.67)	[0.0004]	[0.0007]	[0.0011]	[0.0023]	[0.0048]	[0.0072]
Attendance	-0.0175	-0.0368	-0.0371	-0.0374	-0.0378	-0.0390	-0.0411	-0.0432
	(0.03)	(0.67)	[0.0003]	[0.0006]	[0.0009]	[0.0021]	[0.0042]	[0.0063]

Notes: The uncontrolled coefficient is from the regression of enrollment (attendance) on distance to secondary school without any other covariate. The controlled coefficient is from the regression of enrollment (attendance) on distance to secondary school with all other covariates including household fixed effects and district specific time trends. In all these regressions, only the linear term of the distance to secondary school is considered. The results are obtained through the Stata command `psacalc` (Oster, 2013).

Table 3.4: Robustness - controlling for distance to middle school and additional trends

	Enrollment	Attendance	Enrollment	Attendance
Variable	(1)	(2)	(3)	(4)
Distance to secondary school	-0.146***	-0.131***	-0.134***	-0.115***
	(0.027)	(0.041)	(0.028)	(0.041)
Square of distance to secondary school	0.009***	0.008**	0.009***	0.007**
	(0.002)	(0.003)	(0.002)	(0.003)
Distance to middle school	0.007	0.012		
	(0.016)	(0.018)		
Caste (SC/ST) * Time			0.038	0.004
			(0.164)	(0.159)
Caste (Backward) * Time			0.123	0.118
			(0.142)	(0.131)
Religion (Muslim) * Time			0.162	0.186
			(0.211)	(0.269)
Log total primary pass (age 11-15) in village * Time			-0.034	-0.031
			(0.040)	(0.043)
Girls' secondary school * Time			0.250**	0.181
			(0.122)	(0.110)
Other Covariates	Yes	Yes	Yes	Yes
District Specific Time Trends	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,271	1,271	1,194	1,194
R-squared	0.676	0.675	0.690	0.687

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.5: Sensitivity of results considering various age groups

	5–10 years		6–12 years		6–14 years	
	Enrollment	Attendance	Enrollment	Attendance	Enrollment	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
Distance to secondary school	-0.114***	-0.101***	-0.115***	-0.095***	-0.043*	-0.028
	(0.019)	(0.026)	(0.019)	(0.029)	(0.021)	(0.023)
Square of distance to secondary school	0.008***	0.007***	0.008***	0.006***	0.002	0.001
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Other Covariates	Yes	Yes	Yes	Yes	Yes	Yes
District Specific Time Trends	Yes	Yes	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,526	1,516	1,692	1,686	2,069	2,058
R-squared	0.686	0.681	0.631	0.628	0.574	0.578

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.6: Effects of change in secondary-school-distance outside village on primary-school-participation within village

	Enrollment	Attendance
	(1)	(2)
Distance to secondary school	-0.298***	-0.330***
	(0.034)	(0.041)
Square of distance to secondary school	0.013***	0.013***
	(0.002)	(0.002)
Other Covariates	Yes	Yes
District Specific Time Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	651	651
R-squared	0.733	0.729

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.7: Effects after controlling for sibling externality

	Enrollment	Attendance
	(1)	(2)
Distance to secondary school	-0.142***	-0.129***
	(0.025)	(0.038)
Square of distance to secondary school	0.009***	0.008**
	(0.002)	(0.003)
Number of children in secondary school	-0.005	0.028
	(0.050)	(0.059)
Other Covariates	Yes	Yes
District Specific Time Trends	Yes	Yes
Household Fixed Effects	Yes	Yes
Observations	1,271	1,271
R-squared	0.676	0.675

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.8: Heterogeneity of effects based on children's age and gender

	Age-group				Gender			
	6–7 years		8–10 years		Male		Female	
	Enrollment	Attendance	Enrollment	Attendance	Enrollment	Attendance	Enrollment	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance to secondary school	-0.282**	-0.225	-0.102***	-0.101**	-0.133**	-0.126	-0.133	-0.076
	(0.122)	(0.134)	(0.036)	(0.037)	(0.062)	(0.077)	(0.079)	(0.113)
Square of distance to secondary school	0.015*	0.008	0.007**	0.007**	0.008*	0.008	0.006	0.001
	(0.009)	(0.011)	(0.003)	(0.003)	(0.004)	(0.006)	(0.007)	(0.010)
Other Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Specific Time Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	495	495	776	776	647	647	624	624
R-squared	0.873	0.873	0.809	0.807	0.817	0.820	0.806	0.801

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.9: Heterogeneity of the effect on enrollment depending on households' land ownership (baseline), immunization status of children (baseline), village size (baseline) and village access to bus-stop

	Land ownership		All children immunized?		Village size		Bus-stop
	Land≤Median	Land>Median	No	Yes	Small	Large	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance to secondary school	-0.211***	-0.102***	-0.071*	-0.149***	-1.053***	-0.162***	-0.153***
	(0.049)	(0.028)	(0.038)	(0.038)	(0.283)	(0.018)	(0.020)
Square of distance to secondary school	0.015***	0.010***	0.004	0.009***	0.100***	0.011***	0.011***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.028)	(0.002)	(0.002)
Distance to secondary school * Bus stop							-0.032**
							(0.015)
Other Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Specific Time Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	563	708	473	543	449	822	1,271
R-squared	0.659	0.740	0.740	0.670	0.679	0.693	0.677

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. The median land ownership of households was 1.27 acres in 1997-98. Information on immunization was not collected for households that did not have any child in the age-group of 0-5 years in 1997-98. Therefore, we have a reduced sample for the analysis based on whether the household had all children (0-5 years) immunized or not. Small (large) villages are defined as villages where the number of households were less (more) than 200 in 1997-98.

Table 3.10: Effect of distance to secondary school on primary school participation: Household level regressions with NSS data

	Enrollment	Attendance
	(1)	(2)
Distance to primary school $\in$ [1km, 5km)	-0.021**	-0.020**
	(0.010)	(0.010)
Distance to primary school $\geq$ 5km	-0.126	-0.134
	(0.090)	(0.090)
Distance to secondary school $\in$ [1km, 5km)	-0.002	-0.003
	(0.005)	(0.005)
Distance to secondary school $\geq$ 5km	-0.033***	-0.034***
	(0.008)	(0.008)
Male	0.038***	0.039***
	(0.004)	(0.004)
Age	0.266***	0.269***
	(0.016)	(0.016)
Square of age	-0.015***	-0.016***
	(0.001)	(0.001)
Head literate	0.120***	0.121***
	(0.005)	(0.005)
Head female	0.027***	0.029***
	(0.008)	(0.008)
Household size	-0.002	-0.002*
	(0.001)	(0.001)
Land (0.05 - 1 acre)	0.037***	0.040***
	(0.007)	(0.007)
Land (more than 1 acre)	0.068***	0.070***
	(0.006)	(0.006)
Religion - Muslim	-0.074***	-0.073***
	(0.010)	(0.010)
Caste - ST	-0.070***	-0.071***
	(0.010)	(0.010)
Caste - SC	-0.037***	-0.037***
	(0.007)	(0.007)
Caste - OBC	-0.020***	-0.021***
	(0.006)	(0.006)
Constant	-0.319***	-0.332***
	(0.080)	(0.081)
State Region Fixed Effects	Yes	Yes
Observations	32,168	32,186
R-squared	0.124	0.124

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3.*A*1: Summary statistics

	1997-98			2007-08		
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Enrollment	705	0.69	0.46	566	0.83	0.38
Attendance	705	0.64	0.48	566	0.81	0.40
Distance to primary school	705	0.69	0.94	566	0.08	0.41
Distance to middle school	705	2.54	2.27	566	1.36	1.70
Distance to secondary school	705	6.25	5.13	566	3.72	3.25
Male	705	0.49	0.50	566	0.54	0.50
Age	705	8.07	1.45	566	8.04	1.46
Birth order	705	2.99	1.71	566	5.00	2.32
Mother literate	705	0.17	0.38	566	0.23	0.42
Head literate	705	0.44	0.50	566	0.42	0.49
Head female	705	0.04	0.21	566	0.08	0.28
Head employed	705	0.91	0.28	566	0.83	0.38
Caste - General	705	0.16	0.37	566	0.14	0.35
Caste - SC/ST	705	0.26	0.44	566	0.27	0.44
Caste - Backward	705	0.58	0.49	566	0.59	0.49
Religion - Hindu	705	0.95	0.22	566	0.94	0.24
Religion - Muslim	705	0.05	0.22	566	0.06	0.24
Household size	705	9.25	5.00	566	8.49	3.16
Land owned (acres)	705	3.37	6.23	566	1.72	3.09
School quality tercile 1	705	0.54	0.50	566	0.06	0.24
School quality tercile 2	705	0.37	0.48	566	0.33	0.47
School quality tercile 3	705	0.10	0.29	566	0.61	0.49
Off farm employment rate	705	46.43	28.83	566	28.34	21.79
Distance to district headquarter	705	33.95	17.46	566	31.10	15.58
Road access - None/Trail	705	0.12	0.33	566	0.10	0.31
Road access - Katcha	705	0.31	0.46	566	0.09	0.29
Road access - Paved	705	0.26	0.44	566	0.30	0.46
Road access - Pucca	705	0.31	0.46	566	0.50	0.50
Village population	705	2102	1165	566	3504	1748
Distance to infrastructure facilities	705	0.25	1.36	566	-0.33	1.08
Night lights	705	2.02	2.02	566	3.56	3.21
Baseline girls' secondary school	659	0.55	0.50	-	-	-
Baseline total primary pass (age 11-15) in village	705	114	85	-	-	-

Source: Survey of Living Condition (SLC) data

Table 3.$A2$: Results with different sets of covariates

	Enrollment						Attendance					
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance to primary school	-0.088**	-0.079*	-0.019	-0.019	0.029	0.043	-0.105**	-0.094**	-0.035	-0.036	0.014	0.045
	(0.042)	(0.042)	(0.049)	(0.048)	(0.039)	(0.038)	(0.043)	(0.042)	(0.050)	(0.050)	(0.043)	(0.038)
Distance to secondary school	-0.028*	-0.079***	-0.065***	-0.066***	-0.106***	-0.143***	-0.031*	-0.089***	-0.076***	-0.076***	-0.110***	-0.127***
	(0.014)	(0.023)	(0.022)	(0.022)	(0.017)	(0.024)	(0.016)	(0.022)	(0.020)	(0.020)	(0.021)	(0.037)
Square of distance to secondary school		0.004**	0.003**	0.003**	0.006***	0.009***		0.004***	0.004***	0.004***	0.006***	0.008**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Child Level Covariates	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Household Level Covariates	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Village Level Covariates	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Trend Variables	No	No	No	No	No	Yes	No	No	No	No	No	Yes
Household Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,271	1,271	1,271	1,271	1,271	1,271	1,271	1,271	1,271	1,271	1,271	1,271
R-squared	0.622	0.630	0.658	0.660	0.671	0.676	0.617	0.627	0.657	0.658	0.670	0.674

Notes: Robust clustered standard errors (clustered at the village level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. Child level covariates include child's gender, age, square of age, birth order, and whether mother is literate. Household level covariates are whether the household head is literate, whether the head is female, whether the head is employed, household size, and land ownership. Village level variables control for quality of public primary schools in the village, off-farm employment rate, road access, log of village population, distance to infrastructure facilities, and night lights. Trend variables capture time trends depending on distance of the village to district headquarter, and district specific time trends.

Chapter 4

Intra-Household Gender Disparity in School Choice: Evidence from Private Schooling in India

4.1 Introduction

Gender equality is one of the central issues in the discourse of development economics. Equal economic opportunity for men and women is essential not only for maintaining basic human rights, but also for its catalytic effect in propelling economic development (Duflo, 2012). Gender parity is one of the six goals of the global "Education for All" campaign led by the United Nations Educational, Scientific and Cultural Organization (UNESCO). In India, some of the major public policy initiatives like *Sarva Shiksha Abhiyan* (Education for All) have aimed to universalize elementary education and to reduce disparity across gender, regions and social-groups. While the government has concentrated on providing free education and improving enrollment rates at the elementary level, the quality of education has remained a concern.[1] During the last decade, India has experienced a huge surge in schools established by private providers. Parents who are not satisfied with the quality of government schools have

[1]Annual Status of Education Reports (ASER) from 2005 to 2013 show that enrollment rate among the children in the primary school going age group has improved substantially and remained steady at above 95 percent. In contrast, learning outcomes in reading and mathematics have been quite unsatisfactory and have not improved over time.

perceived private schools as a better alternative. In the literature, there is no consensus about the effectiveness of private schools in increasing human capital. Several studies opine that private schools provide better quality of teaching, teacher absenteeism is less and students' learning outcomes in private schools are higher than that in government schools (Kingdon, 1996; Tooley et al., 2007; Muralidharan and Kremer, 2008; Desai et al., 2008).[2] On the other hand, there are studies which find that private schooling has serious equity issues: children from poorer households, lower socio-economic backgrounds, from rural areas and girls are less likely to attend private schools (Harma, 2011; Maitra et al., 2013; Woodhead et al., 2013). One of the factors that can potentially preclude the access to private schooling for disadvantaged groups is the cost of schooling. Private schools are fee-charging, hence they are relatively more expensive than government schools. Thus, there is an ongoing debate on whether private schools are capable of contributing in the path towards the Millennium Development Goals of universalization of primary education and elimination of gender disparity in primary and secondary education by 2015.

In this context, the aim of this paper is to investigate whether households differentiate between their sons and daughters while choosing schools. In particular, are boys more likely to be sent to schools that are perceived to be of better quality, i.e., private schools? Further, the study provides evidence that the pro-male bias in private school choice is linked to the cost difference between private and government schools.

The literature on gender inequality offers plausible explanation why households may treat boys and girls differentially while deciding on their education (Alderman and King, 1998). It has been widely recognized that households tend to allocate more resource to those children who are expected to be more economically productive adults (Rosenzweig and Schultz, 1982). Prevailing cultural norms can also cause gender inequality of various forms (Jayachandran,

[2]Desai et al. (2008) reports a large inter-state variation in the relative performance of children enrolled in private vis-a-vis government schools. After controlling for family background characteristics, students in private schools perform only modestly better than those in government school, and the pattern is even reversed in some states.

2014). In rural India, females are less likely to enter the labour force.[3] Besides, the society is patrilocal, where the girl stays with her husband's family after marriage. Hence, in most cases, old age support for the parents is provided by the sons. Therefore, from the perspective of the parents, it may be rational to invest more on the education of their sons rather than daughters. Since boys will enter the labour force and compete with others, the quality of education they receive may be important. Therefore, it is plausible that boys will be sent to fee-charging private schools which are deemed to be superior in quality. The financial burden of dowry may discourage households from investing more on the girls' education, rather they might want to save the resource for later dowry payment (Subramaniam, 1996).[4] Female disadvantage in education can also be explained by supply constraints. For instance, if private schools are farther away from home, the distance can act as a deterrent for girls' participation in these schools. Thus, an amalgam of all these economic and cultural factors can lead to gender disparity in the choice of schooling.

In this paper, I use household level panel data from rural India to investigate whether there is an intra-household gender gap in private school choice. A number of studies have extensively analyzed the issue of gender disparity in overall enrollment, grade progression and education expenditure in developing countries (Deolalikar, 1993; Alderman and King, 1998; Glick, 2008; Sawada and Lokshin, 2009; Azam and Kingdon, 2013). In the era of growing incidence of private schooling, the recent literature also sheds light on the determinants of private versus public school choice (Alderman et al., 2001; Glick and Sahn, 2006; Nishimura and Yamano, 2013). However, very few papers explore the boy-bias in private school choice (Maitra et al., 2013; Woodhead et al., 2013). Moreover, most of these studies on private schooling are based on cross-section data and they do not look into the intra-household decision making process. To address the issue of gender bias within household, it is imperative to take into

[3]For example, the National Sample Survey of 2009-10 shows that in rural Uttar Pradesh, the male and female labour force participation rates (in 15-59 age group) were 81.5 percent and 14.6 percent respectively.

[4]Girls may still be provided some basic education since there is assortative matching in the marriage market. However, it is likely that the matching takes place based on years of schooling rather than the quality of schooling.

account innate household preferences, i.e., the household fixed effect (Subramaniam, 1996; Jensen, 2002; Azam and Kingdon, 2013). Besides, analyzing the choice between private and government schools considering only the enrolled children may yield inconsistent estimates if the enrollment decision is not random. Therefore, this paper estimates a model proposed by Wooldridge (1995) where selection corrected estimates are obtained in a panel data set up, and unobserved household heterogeneity is allowed to be correlated with independent variables.[5]

The analysis shows that there is a significant gender gap of 5.4 percentage points in the probability of private school enrollment among the children in the school going age group of 6 to 19 years. Decomposing this effect, I find that the gender inequality is rising over time. The gender gap in private school choice was almost zero in 1997-98, it has increased to 5 in 2007-08 and to 9 percentage points in 2010-11. The gender gap is found to be significant for both younger (6 to 10 years) and older (11 to 19 years) children.

Since private schools are more expensive than their government counterpart, the cost difference between these two types of schools may have an implication on the school choice and any gender bias associated with that decision.[6] In this paper, we explore this angle and find that higher village level average differences in cost of private and government schooling is associated with a larger gender gap in private school enrollment, even after controlling for average school quality. Among the cost components, the direct cost, school fees in particular, comes out to be the most significant factor to be correlated with the gender gap. Acknowledging that cost differences could still be endogeneous, I show that the result is robust to village level confounding trends. The findings also remain unaltered in a separate exercise that inspects the sensitivity of the effects with respect to potential omitted variable bias following a method developed by Altonji et al. (2005) and Oster (2013).

While the growing literature on private schooling has mainly focused on the relative efficiency of these schools, this study contributes to the discussion by highlighting the important

[5]The paper by Maitra et al. (2013) is the closest to this study in terms of context as well as methodology.

[6]See Glick (2008) for a discussion on the existing literature regarding the effect of cost on gender gap in education in general.

aspect of gender disparity in private school participation. While private schooling is becoming more common in rural areas, the cost difference between private and government schools may have an important role to play in households' decision about sending their boys and girls to a private school.

The rest of the paper is organized as follows. Section 4.2 describes the dataset and provides a descriptive analysis of trends in private school participation. Section 4.3 lays out the empirical model. Sections 4.4 presents the results on gender gap in private school choice. Section 4.5 discusses the relationship between gender gap and the difference in cost of private and government schooling. Section 4.6 concludes.

4.2 Data and Descriptives

The data set used in this paper is a longitudinal study of households first surveyed as a part of the World Bank's Living Standards Measurement Study (LSMS) in Uttar Pradesh, a state in India (this survey is also called the Survey of Living Conditions, or SLC). This is a three period panel data on rural households in 43 villages from eastern and southern Uttar Pradesh.[7] The baseline data were collected in 1997-98 under LSMS. The same set of households were resurveyed in 2007-08 and again in 2010-11.[8] The survey comprises of a village questionnaire and a household questionnaire which contains detailed information on the demographics of each household member and schooling information for every child belonging to the age group of 6 to 19 years. For the purpose of this study, I concentrate on children who are in the school going age, 6-19 years, at each time point observed in the dataset. The analysis considers 810 households and 5175 observations on children in this age group over the three years.[9]

[7]Uttar Pradesh (UP) is considered to be one of the most backward states in India. Our sample includes 10 districts in UP. The number of villages in each district varies from 2 to 6 per district.

[8]The second round of survey in 2007-08 was funded by the University of Oxford and the World Bank. The third round in 2010-11 was funded by the Planning and Policy Research Unit of the Indian Statistical Institute, Delhi. All three surveys were conducted during the same time of the year - from December to April.

[9]The sample consists of 888 households in the baseline (1997-98). Among these households, 767 in 2007-08 and 798 in 2010-11 were surveyed. Thus, household level attrition rates with respect to the baseline data are 13.6 and 10.1 percent in 2007-08 and 2010-11 respectively.

The summary statistics of variables used in this study are presented in *Table* 4.*A*1. From descriptive analysis, we get some clear patterns of overall enrollment and private school choice for boys and girls. According to estimates from the panel data, there is a convergence in the overall enrollment rates of boys and girls over time (*Figure* 4.1). In 1997-98, 69 percent of the boys in 6-19 age group were enrolled in school, while among girls, only 50 percent were enrolled at that time. However, this gender gap has reduced significantly over time. In 2007-08, 65 percent of the girls were enrolled as against 70 percent of the boys. By 2010-11, the gap had almost vanished with 72 percent of girls in school as compared to 74 percent for boys. A t-test for mean comparison suggests that there is no statistically significant difference between the enrollment rates of boys and girls in 2010-11. On the other hand, if we look at the trends in private schooling among children who are enrolled, there is no such convergence across gender.[10] The private school enrollment rates for boys and girls have grown steadily apart over the period of study, from being similar (24 percent for boys and 23 percent for girls) in 1997-98 to 39 percent for boys and 33 percent for girls in 2007-08 to 57 percent for boys and 47 percent for girls in 2010-11 (*Figure* 4.2). The gender difference in private school enrollment is found to be statistically significant in both 2007-08 and 2010-11. Combined with the trends in overall enrollment rates, this suggests that the steep rise in private schooling over the years can be attributed more to boys than girls. Moreover, while the gender gap in overall school enrollment tends to disappear over time, the gap in terms of private school participation has become starker.

Although the descriptive analysis points out a gap in private school enrollment rates between boys and girls, it is important to take into account other factors, including households' intrinsic preferences that may affect private school choice. The following section illustrates an empirical model for this purpose.

[10]Schools attended by the children are categorized into two groups: *government* and *private*. The first category includes schools which are completely funded and managed by government, and also government-aided schools which are regulated by the state government. The second category includes private unaided schools which are fully self-financed and autonomous in management (Kingdon, 1996; Harma, 2011).

4.3 Empirical Model

This section lays out an econometric model suitable for studying intra-household gender disparity in private school choice. The model begins with the issue of identifying the effect of gender in the school choice decision within households. In a later section on the implication of cost of schooling, the model incorporates cost difference variables to examine the relationship between schooling cost and gender gap.

4.3.1 Basic Set-up

In this section, a multivariate regression model is set up to control for various explanatory factors and to investigate if households actually prefer boys over girls while deciding about whether to send their children to private school. Consider P^* to be the latent decision making process by the household to enrol a child in a private school. However, we can observe only the binary outcome, P, of this decision, that indicates whether the child goes to a private school ($P = 1$) or a government school ($P = 0$).

$$P^*_{cht} = \mathbf{X}_{cht}\beta + \alpha_h + \varepsilon_{cht} \ ; \ P_{cht} = 1[P^*_{cht} > 0]; \tag{4.1}$$

$$c = 1, ..., C(h, t); \ h = 1, ..., H(t); \ t = 1, 2, 3.$$

The choice of private versus government school is modeled by *Equation* (4.1). The subscript c refers to a child in household h at time period t. In any time period t, there are $H(t)$ households who have at least one child in the school going age group and are included in the sample, and there are $C(h, t)$ children in h-th household in t-th time period. $\mathbf{X}$ denotes the vector of explanatory variables that could affect the private school choice, and β is the corresponding coefficient vector. $\mathbf{X}$ contains our main variable of interest, the gender of the child, captured by a dummy variable indicating whether the child is female. It also includes other child, household and region specific variables. The child specific variables are the age of the child, square of age (for possible non-linearity in the effect of age), birth order of the child within the household, dummy variables representing whether the father and mother are literate. The household specific variables that are included are dummy variables indicating whether the head is literate and whether the head is female, total number of female and male children in the household, household size, household wealth captured by wealth index, religion and caste.[11] Among the region level variables, we have the proportion of private schools among all the schools in the village. Quality of education in the government schools in the village can be an important factor to determine participation in private schools. Therefore, I include an index measuring the quality of village based government primary schools.[12] Village level infrastructure, population and economic prosperity may affect the location choice of private schools. These factors can also influence the households' demand for private schooling. Therefore, the model includes variables capturing the village population, access to all weather (*pucca*) road, and facility index.[13] The light density at night reflects local economic development; it can be used as a proxy for the income of the region. Hence the model includes the average night-time luminosity of the area with a radius of 10 km around the centroid of the village.[14] To consider the secular rise in private school enrollment over time, year dummies are included. The growth in demand can also depend on the remoteness of the village. Therefore,

[11]Household wealth is captured by an index obtained from principal component analysis of binary indicators of various durable assets and land ownership of the household. The assets considered here are: radio, camera, bicycle, motorcycle, motorcar, refrigerator, washing machine, fan, heater, television, pressure lamp, telephone, sewing machine, pressure cooker, and watch.

[12]This school quality index is derived using principal component analysis of the following features: type of structure, main flooring material, whether the school has classrooms, number of classrooms, whether the classes are held inside classrooms, whether the school has usable blackboards, whether desks are provided to the students, whether mid-day meal is provided and the proportion of teachers present on the day of survey. For those villages where more than one government primary schools are present, I consider the representative school to be the one which has the maximum number of students.

[13]The facility index is constructed by principal component analysis of binary variables indicating whether a village has certain facilities, namely, bus stop, police station, telephone service, public distribution shop, bank, and primary health centre.

[14]This variable is constructed from the satellite data on lights at night. This data is recorded worldwide for every one square kilometre area (approximately) by the Operational Linescan System (OLS) flown on the Defense Meteorological Satellite Program (DMSP) satellites. This dataset has been downloaded from the website of the National Oceanic and Atmospheric Administration (NOAA) of the USA (`http://ngdc.noaa.gov/eog/dmsp/download_radcal.html`). The information pertaining to SLC villages has been extracted by matching the latitude and longitude coordinates of the villages in our sample. The night lights data pertaining to the year 1997, 2007, and 2010 were matched with the respective SLC survey rounds.

the year dummies are interacted with the distance of village from the district head-quarter. Besides, it is well-documented in the literature that various kinds of gender-specific discriminatory practices are manifested through a population sex imbalance in India (Rosenzweig and Schultz, 1982; Das Gupta, 1987; Oster, 2009). To control for such gender bias prevailing in the community, the model includes population sex ratio in the villages, interacted with year fixed effects.[15] Furthermore, the vector of explanatory variables also contains district-by-time fixed effects to allow for differential rate of growth in private schooling across districts.

In addition to the explanatory variables described above, the model also includes household specific unobserved heterogeneity by the term α_h. Since we follow the same households over time, and there are multiple children of different gender in a household, it is possible for us to identify the coefficient of female dummy even after controlling for household specific fixed effects. By including these household level fixed effects, we control for unobservable factors that are particular to each household and do not change over time. It also ensures that we focus on the decision making process within household rather than comparing outcomes across different households. It is widely observed that female children tend to end up in larger families because fertility decisions are endogenous and parents prefer to have at least one boy child. Due to this son preferring, differential stopping behaviour, the number of children as well as their birth order is often determined endogenously within the household (Yamaguchi, 1989; Clark, 2000; Basu and Jong, 2010). Therefore, it is crucial to control for these fertility preferences by including household fixed effects (Jensen, 2002; Azam and Kingdon, 2013).

4.3.2 The Selection Problem

We can estimate *Equation* (4.1) following a Linear Probability Model and obtain Ordinary Least Square (OLS) estimates of β. In this method, we can control for the household level fixed effects (α_h) either by taking a time-demeaned transformation of *Equation* (4.1), or by

[15]The village level sex ratio is defined as the number of females per 1000 males in the population. This variable has been sourced from the 2001 Census of India data. Since this variable is time invariant, therefore it is interacted with time fixed effects to consider its differential effect on private schooling over the years.

explicitly including household specific dummy variables. However, note that the choice of school type is observed only for those children who are enrolled. If the decisions of school choice and enrollment are correlated, then estimating the school choice equation considering only the selected sample of enrolled children may lead to biased and inconsistent estimates. This is similar to the standard sample selection problem (Heckman, 1979). A comparison of means of the explanatory variables show significant difference in several characteristics between the sample of enrolled and not-enrolled children (*Table* 4.*A*2). Therefore, the decision to enrol is taken into account in the econometric model by the latent variable S^*, and its observable counterpart is caputured by the binary enrollment outcome S, which is one if the child is enrolled in some school, and zero otherwise.

$$S^*_{cht} = \mathbf{Z}_{cht}\gamma + \delta_h + u_{cht} \; ; \; S_{cht} = 1[S^*_{cht} > 0]. \tag{4.2}$$

The enrollment decision modeled in *Equation* (4.2) has a form similar to the school choice model. The set of explanatory variables, $\mathbf{Z}$, contains all the variables that are present in $\mathbf{X}$; but for ease of identification, it also contains an additional variable that is validly excluded from *Equation* (4.1). After controlling for the composition of private and government schools in the village, it is plausible to think that total number of schools in the village will affect only the enrollment decision, but not the private school choice decision. Therefore, I use the total number of all kinds of schools in the village as the variable which is excluded from the main equation, but is included in the selection equation. Unobserved household heterogeneity are considered by the term δ_h. The private school choice variable P_{cht} in *Equation* (4.1) is observable only when $S_{cht} = 1$.

When the selection process is non-random, the most widely used method in the literature for correcting sample selection bias is the model proposed by Heckman (1979), also known as the "Heckit" model. However, in the presence of unobserved heterogeneity, implementation of Heckit model becomes problematic. While we can estimate the main relationship

(*Equation* (4.1)) as a linear model based on the selected sample, the Heckit model requires that we estimate the selection equation (*Equation* (4.2)) by using a probit model. But, probit being a non-linear model, it is not possible to eliminate the fixed effects by taking any within-transformation of the equation. Besides, since probit model employs maximum likelihood estimation, if we attempt to estimate the selection equation including household specific dummy variables to capture unobserved heterogeneity, we will face the "incidental parameters problem" (Neyman and Scott, 1948; Lancaster, 2000). This will lead to inconsistency in the estimates of not only δ_h, but also γ. On the other hand, failure to account for the unobserved heterogeneity, which may be correlated with other regressors in the model, may result in biased and incosistent estimates of the parameters of interest. Thus, standard Heckit model is infeasible to solve the selection problem in our context.

4.3.3 Method for Selection Correction in Panel Data

Wooldridge (1995) offers a method for correcting for sample selection bias in linear panel data models where unobserved heterogeneity is allowed to be correlated with the observable explanatory variables in both the selection equation and the equation of interest. While this method is conceptually similar to Heckman (1979), it is appropriate for panel data models such as in this paper.[16]

Since the choice of private versus government school is observed only for the sample of enrolled children, a sufficient condition for obtaining a consistent estimate of β by running a pooled OLS model on *Equation* (4.1) is given by:

$$E(\alpha_h + \varepsilon_{cht} \mid \mathbf{X}_{cht}, S_{cht} = 1) = E(\alpha_h \mid \mathbf{X}_{cht}, S_{cht} = 1) + E(\varepsilon_{cht} \mid \mathbf{X}_{cht}, S_{cht} = 1) = 0.$$

The conditional expectation specified above will not be zero if household heterogeneity is correlated with the explanatory variables or the selection process is non-random. One way to

[16]Dustmann and Rochina-Barrachina (2007) use this method in a similar set up to estimate the females' wage equations.

tackle this problem would be to parameterise these conditional expectations and add them to the main equation (Wooldridge, 1995; Dustmann and Rochina-Barrachina, 2007). To derive this estimator in the context of our analysis, based on Wooldridge (1995), I assume the following structure of our econometric model:

(i) To allow the unobserved household heterogeneity in the selection equation (*Equation* (4.2)) to be correlated with the explanatory variables, following Mundlak (1978), I assume that δ_h is a linear function of the within-household average of $\mathbf{Z}_{cht}$ over all children and all time periods. Thus,

$$\delta_h = \eta_0 + \overline{\mathbf{Z}}_h \eta + e_h, \tag{4.3}$$

where e_h is a random variable independent of other factors, and $\overline{\mathbf{Z}}_h = \frac{1}{T(h)} \sum_t (\frac{1}{C(h,t)} \sum_c \mathbf{Z}_{cht})$ is the household specific average values of the observed explanatory variables, with $C(h,t)$ being the number of children present in h-th household at t-th time period and $T(h) \in \{1,2,3\}$ being the number of time-periods when h-th household has at least one child in the relevant age group so that it is included in the sample of our analysis.[17]

(ii) Following (i), the reduced form of the selection equation becomes:

$$S^*_{cht} = \eta_0 + \overline{\mathbf{Z}}_h \eta + \mathbf{Z}_{cht}\gamma + v_{cht}; S_{cht} = 1[S^*_{cht} > 0], \tag{4.4}$$

where $v_{cht} = e_h + u_{cht}$. Assume that v_{cht} is independent of $\mathbf{Z}_{ch}$, and $v_{cht} \sim Normal(0, \sigma_t^2)$, where $\mathbf{Z}_{ch} = (\mathbf{Z}_{ch1}, ..., \mathbf{Z}_{chT})$.

(iii) In the main equation, let us assume that the household specific unobserved effects has a linear relationship with the household level averages of the explanatory variables (Mundlak, 1978). Hence, I assume that the conditional expectation of α_h given $\mathbf{Z}_{ch}$ and v_{cht} is linear.

[17]Another way of allowing for correlation between the unobserved household specific effects and the explanatory variables is similar to Chamberlain's Correlated Random Effects Model (Chamberlain, 1980; Wooldridge, 2002). This model would assume that δ_h is a linear function of the leads and lags of the explanatory variables. But our data is unbalanced in nature because not all households have children in the school-going age group in all time periods, hence Chamberlain's (1980) specification is not suitable here.

Thus, we have the following relationship:

$$E(\alpha_h \mid \mathbf{Z}_{ch}, v_{cht}) = \psi_0 + \overline{\mathbf{X}}_h \psi + \pi_t v_{cht}, \tag{4.5}$$

where $\overline{\mathbf{X}}_h = \frac{1}{T(h)} \sum_t (\frac{1}{C(h,t)} \sum_c \mathbf{X}_{cht})$. Note that under the exclusion restriction, the elements of $\mathbf{Z}$ which are not in $\mathbf{X}$ are independent of α_h and ε_{cht}, hence they do not appear in the above relationship.

(*iv*) Finally, assume that ε_{cht} is mean independent of $\mathbf{Z}_{ch}$ conditional on v_{cht}, and its conditional expectation is linear in v_{cht}:

$$E(\varepsilon_{cht} \mid \mathbf{Z}_{ch}, v_{cht}) = E(\varepsilon_{cht} \mid v_{cht}) = \rho_t v_{cht}. \tag{4.6}$$

Note that we do not observe v_{cht}, rather only the binary enrollment decision (S_{cht}) for each child. Since S_{cht} is a function of $\mathbf{Z}_{ch}$ and v_{cht}, we can apply the law of iterated expectations to *Equation* (4.5) and (4.6), and combine them to get the following relation:

$$\begin{aligned} E(\alpha_h + \varepsilon_{cht} \mid \mathbf{Z}_{ch}, S_{cht} = 1) &= \psi_0 + \overline{\mathbf{X}}_h \psi + (\pi_t + \rho_t) E(v_{cht} \mid \mathbf{Z}_{ch}, S_{cht} = 1) \\ &= \psi_0 + \overline{\mathbf{X}}_h \psi + \zeta_t \lambda_{cht}, \end{aligned} \tag{4.7}$$

where $\zeta_t = \pi_t + \rho_t$, $\lambda_{cht} = E(v_{cht} \mid \mathbf{Z}_{ch}, S_{cht} = 1)$. Finally, the main equation capturing private school choice is modified in accordance with the econometric structure above. Thus, we have:

$$E(P_{cht} \mid \mathbf{X}_{ch}, S_{cht} = 1) = \psi_0 + \overline{\mathbf{X}}_h \psi + \mathbf{X}_{cht} \beta + \zeta_t \lambda_{cht}. \tag{4.8}$$

A consistent estimate of β can be obtained through *Equation* (4.8) following a few steps. First, the reduced form sample selection equation (*Equation* (4.4)) is estimated using standard probit model separately for each time period, and λ_{cht} is estimated as the ratio of normal density to cumulative distribution function (also known as the Inverse Mills Ratio). Thus, we have separate estimates of λ_{cht}, or the Inverse Mills Ratio, for different time periods. In the next step, these Inverse Mills Ratios are included in the main regression (*Equation* (4.8)) as additional regressors. Then *Equation* (4.8) is estimated by pooled OLS method. Since the set of explanatory variables now contains generated regressors (the estimated Inverse Mills Ratios), therefore, the standard errors are bootstrapped and clustered at the household level.

It is noteworthy that this method allows for possible correlation between the unobserved household heterogeneity and the observed explanatory variables through the Mundlak formulation. For the regressors which vary within household or over time, the corresponding elements of β are identified in this method. On the contrary, due to the Mundlak formulation, it is not possible to separately identify the elements of β from the elements of ψ for those regressors which neither vary within household nor over time.[18] Nevertheless, this does not hinder us from identifying the effect of gender. Once the estimation is carried out, we can investigate if it is important to control for unobserved heterogeneity by performing a Wald test for the joint significance of the elements in ψ. Similarly, if the null hypothesis of $\zeta = 0$ is rejected, then it would imply that it is crucial to correct for sample selection bias in the regression. Thus, this method can be used as a test to determine whether the sample selection can potentially bias the estimates. If the inverse mills ratio is found to be statistically insignificant, then one may ignore the possibility of non-random enrollment decision. In that case, the usual fixed effects estimation can be performed directly on *Equation* (4.1).

4.4 Results: Gender Gap

This section discusses the results that are obtained from estimating the empirical model to identify intra-household gender gap in private school choice.

Table 4.1 contains the results from the main regression for all children in the school going age group of 6 to 19 years. The first column shows results under the model of selection

[18]These variables in our model would be religion, caste, and distance of village from district headquarter. So I do not report the coefficients of these regressors in the output table.

correction, while the second column, for the purpose of comparison, estimates the regression model without taking into account the sample selection problem (i.e., excluding the λ_{cht} terms). Being a female child reduces, on an average, the probability of enrollment in a private school by 5.4 percentage points. The Wald test shows that the selection effects are jointly significant at five percent level, indicating the importance of taking into account the non-random enrollment decision.[19] From the regression output in the second column, it is found that if we ignore the potential selection bias, then the coefficient is still significant, but it is underestimated. The household fixed effects are also jointly significant at five percent level in both the regressions.

Among the other child level variables, age and birth order have significant effect on private school enrollment. Probability of enrollment in private school increases with age, but the rate of increase reduces with age. There is a strong negative and significant birth order effect. It suggests that parents tend to invest more on education of the first born children and send them to fee-charging private schools, but under resource constraints, they are probably unable to further keep up the investment on the later born children's education. Considering the village level variables, a higher share of private schools in the village positively affects the likelihood of private school enrollment. The facility index comes out to be positive and significant, indicating that villages with better infrastructure facilities have experienced a greater demand for private schooling. Private school enrollment is higher in places nearer to the district headquarter, but this effect is significant only in 1997-98. This result suggests that in the later years, the demand for private schooling has expanded irrespective of the remoteness of the location.

[19]As mentioned before, the total number of schools has been included in the enrollment equation but not in the private school choice equation. Even if this variable is included in the private school choice equation, we find very similar result and the Inverse Mills Ratios are still significant. In this case, identification relies on the non-linearity of the Inverse Mills Ratio function in its argument. The results also remain unchanged when some other identifying variables are used. For example, after controlling for wealth index in both the equations, we can use a binary variable that indicates whether the household belongs to the bottom decile of the wealth distribution. The idea is that this variable identifies the poorest of the households for whom enrollment decision is important, but they are so poor that private school choice is not much relevant to them. In another case, I used three identifying variables: whether there is any school in the village, distance to the nearest public primary school, and their interaction. In all these cases, the selection correction terms are statistically significant in the main regression. Also, the estimate of gender gap remains almost unchanged.

Next part of the analysis looks into how the gender gap in private school enrollment varies over time. For this purpose, the female dummy is interacted with year specific dummy variables in the model. There is clearly an increasing trend in the extent of gender gap over time, as shown in *Table* 4.2. In 1997-98, when private schooling was a sparse phenomenon, the gender gap was statistically insignificant and almost zero in magnitude. In the subsequent years, with the rise in private schooling, the differential treatment between boys and girls has also increased. In 2007-08, girls were 5 percentage points behind boys in private school enrollment. More recently in 2010-11, the female disadvantage in the school choice decision has been most striking, with the gap being 8.9 percentage points.

To find out whether the gender gap is different for younger and older children, I decompose the effects into two age-groups. Children who are between 6 and 10 years in age are supposed to be enrolled in the primary level. The remaining children who are relatively older are in the age-group corresponding to the post-primary level. So I create indicator variables specific to these two age-groups and include their interactions with female dummy in the model. *Table* 4.3 shows that the gender gap in private school choice is significant for both younger and older children. Although the point estimate appears to be slightly higher in magnitude for younger children, the difference between these two coefficients is not statistically significant.

4.5 Exploring the Role of Schooling Cost

This section extends the analysis further and investigates the role of schooling cost in explaining the gender gap in private school choice within households. The SLC dataset has information on various components of educational expenditure for each child who is enrolled in school. The total schooling expenditure is divided into its constituent parts. These can be categorized into *direct expenditure* and *transport and other expenditure*. Direct expenditure includes spending on school fees, books, and uniform. Thus, it can be viewed as the necessary expenditure a household has to incur on a child's education once a particular school has been chosen. The remaining part, i.e., transport and other expenses, depends on the distance to the chosen school, as well as on the mode of transport. Hence they are treated as a separate group of expenditure.[20]

By comparing the schooling expenses on children enrolled in private and government schools, we find that private schooling has been associated with a significantly higher expenditure than government schooling in all three years. The mean of total annual schooling expenditure per child considering children in the government schools has decreased from Rs. 980 in 1997-98 to Rs. 631 in 2007-08 to Rs. 498 in 2010-11.[21] This indicates that the government policies of providing free and compulsory education may have been effective in reducing the cost of government schooling.[22] In contrast, the corresponding figure for children in private schools is Rs. 1863 in 1997-98, Rs. 1845 in 2007-08 and Rs. 2318 in 2010-11. Thus, in 2010-11, the mean expenditure incurred by children in private schools is 4.7 times higher than that in government schools. The difference in terms of average school fees paid by children in these two types of schools is even starker. In 2010-11, the mean expenditures on school fees for private school going children is Rs. 1069, which is almost 9 times higher than the average fees paid by children in government schools (Rs. 119).

The divergent trend between the expenditures of children in private and government schools is further corroborated by *Figure* 4.3. This figure depicts for each year the cumulative distribution functions of direct expenditure incurred by the students in these two types of schools. In every year, the expenditure of private school going children first order stochastically dominates the expenditure of children in government schools. Besides, the gap also appears to be increasing over time. The Kolmogorov-Smirnov test reveals statistically significant difference between these two distribution functions in every year.

[20]The survey has also collected information on the payment made to private tutors, in case the child is sent to private coaching in addition to formal schooling. This component is separate from the expenditure for school education, hence it is not included in this analysis.

[21]The cost figures are deflated using the Consumer Price Index for Rural Labourers, and are expressed in 2010-11 prices. Similarly, any other variable which is in value terms is also expressed in 2010-11 price level.

[22]The Right to Education Act (2009) of India makes it a constitutional right of every children in the age group of 6 to 14 years to get free and compulsory elementary education.

If households' preferences are biased against girls' education, it is plausible that a higher relative cost of private schooling will discourage them even more from sending the girls to private schools. Therefore, it is important to investigate whether the gender gap in private school enrollment is explained by the cost difference in private and government schools. To capture the effect of schooling cost on enrollment choices, one requires variation in schooling opportunities available to the households (Alderman et al., 2001). However, the SLC survey does not have school level data on cost; rather, we observe actual expenditure incurred for each child enrolled in any school. In similar settings, some other studies in the literature have constructed community level cost figures by aggregating the child or household level data on expenditure (Gertler and Glewwe, 1990; Glick and Sahn, 2006; Nishimura and Yamano, 2013). Following this literature, I measure the average yearly expenses of schooling in the village from those children who are enrolled in private and government schools in the sample. This measure is constructed separately for children in primary (6–10 years) and post-primary (11–19 years) schooling age-groups separately. Taking the differences of the village level average costs in the two types of schools, we have a measure of the relative costliness of private schooling with respect to government schooling in the village.[23] Unlike child level expenditure, the average expenditure over all children in the village is less likely to be affected by the child and household level characteristics. *Table* 4.*A*3 presents the child level regression of actual expenditures. It shows that the expenditures for both government and private school going children significantly correlate with various individual and household level factors. On the contrary, as described in the following section, the village level differences in average cost of private and government schooling is determined mostly by the community level characteristics.[24]

[23] Glick and Sahn (2006) takes the median expenditure to reflect the schooling cost in the community. The results presented in this paper remain unchanged if median, instead of mean, is considered to construct the village level cost of schooling. The expenditure patterns and trends in private and government schooling are very similar when median expenditures are considered. For instance, the mean spending on school fees over the three years is Rs. 924 in private schools and Rs. 164 in government schools. The corresponding figures in terms of median are Rs. 600 and Rs. 51 respectively. Thus, the cost difference between private and government schooling is aptly reflected irrespective of whether mean or median is considered.

[24] Since the cost difference is defined at the village level for each age-group, a regression that aims to explore the determinants of cost difference should be based on village level observations. However, there are only 43 villages in the sample, hence a village level regression would not have enough power. Thus, the analysis of

4.5.1 Analysis of Cost Difference

The cost of schooling is not likely to be random. It can be systematically related with other location specific characteristics. This section explores if the cost difference can be attributed to various observable factors at the village and household levels. For this part of the analysis, I consider the direct cost of schooling, which is comprised of school fees, uniform, and books, and constitutes more than 90 percent of the total cost.[25] *Figure* 4.4 plots the cost difference variable against the village level average cost of private as well as government schooling. It is evident that the difference is driven mainly by the private schooling cost; the simple correlation coefficient between these two variables is found to be 0.84 and statistically significant. Conversely, the scatterplot between the difference variable and the cost of government schooling shows a weaker negative relationship. In the next step, I regress the cost difference on several village and household level variables. The results are presented in *Table* 4.4. To control for district level heterogeneity and trends, all specifications include district fixed effects and district specific time fixed effects. Column (1) shows that there is a strong effect of village level factors on the cost variable. Private schools are likely to be more expensive than government schools in places which are accessible by all weather (*pucca*) roads. Villages with better infrastructure facilities, captured by the facility index, have a higher relative cost of private schooling. Cost may also vary depending on the intensity of economic activities in the locality. This is taken into account by including the light density at night which reflects the economic prosperity of a region. We find that the cost gap is larger in richer areas. This result suggests that private schools may set their prices by observing the income level, hence purchasing power, of the local residents. After controlling for other factors, size of the village population is found to have a negative effect on the cost differential. To understand this phe-

cost difference in the following section considers child level data. It also helps to reassure us that the village level cost difference is less likely to be affected by individual level characteristics. When the regression is run on village level data including only the village characteristics as explanatory variables, the coefficients have same sign, but only a few of them are statistically significant.

[25]The results remain qualitatively unchanged even if the total cost (direct plus transport and other costs) is considered instead.

nomenon, let us consider how a private school may decide on its charges from the perspective of a firm. It is reasonable to assume that the size of the student population is larger in bigger villages. Hence, the average class size is also likely to be higher in such villages. In turn, this can have economies of scale effect on the real cost of running a private school. A major part of the revenue of a private school is spent to pay the teachers their salary. The earned revenue will increase if there are more number of students in a class. This may lead to lower school fees if the market for private schools is competitive. In fact, the results also show that the cost gap is significantly lower in places where the ratio of private schools to the total number of schools is higher. It highlights the possibility of price competition among the private school providers. Interestingly, the sex ratio in the population is found to have a significant effect on the cost. A higher proportion of males in the population raises the gap between private and government schooling cost. Considering that the villages are located in a very backward region of northern India, participation in modern economic activities is dominated by males. A community with more men can have greater exposure and aspiration, resulting in higher willingness to pay for a better quality education of their children.

While analyzing the cost of schooling, it is imperative to look into the quality angle as well. SLC data has information on all the government primary schools located inside the village. The regression shows that presence of higher quality government primary schools in the village leads to lower relative cost of private schooling. This suggests that availability of good quality government schools can pull down the demand for private schools, thus reducing the price. However, this does not shed any light on the effect of the quality of private schools. Since SLC did not cover the private schools, therefore, I extract school level information for the SLC blocks from District Information System for Education (DISE) dataset for the years 2007-08 and 2010-11.[26] I construct a school quality index by principal component analysis of several

[26]The DISE data is collected by the National University of Educational Planning and Administration (NUEPA), in collaboration with the Government of India and the UNICEF, and covers information on both government and private schools with elementary level (grade 1 to 8). It is publicly available, and has district, block and village level identifiers. Childrens from villages often go to schools outside the village. Therefore, to measure the quality of available schools, I consider all schools at the block level. Blocks are administrative units within districts; there are 33 blocks in the SLC data.

characteristics related to the school infrastructure and teachers.[27] The cumulative distribution functions of the quality index have been plotted in $Figure$ 4.5 separately for government and private schools. Private schools dominate their counterpart in terms of quality in both the years. In the next step, I calculate the difference in the average quality between private and government schools for each block, and include it as an additional regressor. Results in column (2) show a strong positive association between the block level difference in quality and the cost gap. This indicates that private schools, which are relatively superior in quality than government schools, are also likely to charge higher price.

Finally, I include a set of individual and household level covariates in addition to the village specific factors. Column (3) shows that apart from mothers' literacy and household wealth, other variables do not have any significant effect. The positive coefficients of both these variables suggest that costlier private schools are located in relatively better-off and more educated communities.

The analysis in this section points out the importance of village level attributes in determining the relative cost of private schooling. While our main focus has been on the cost differential, the findings are in line with the existing literature on the location choice of private schools.[28] The next section seeks to investigate if households are relatively more unlikely to send girls to private schools if these schools are more expensive. The village characteristics are explicitly included in the model. In the robustness section, I also take into account the possibility of omitted variable bias arising from unobservables.

[27]The school quality index computed from the DISE data considers the following variables: proportion of classrooms in good condition, proportion of classrooms having blackboards, whether there is a toilet, whether separate toilet is available for girls, whether the school has proper boundary wall, playground, drinking water facility, ramp for disabled children, any computer, if students undergo medical checkup, ratio of teachers to total classes, proportion of female teachers, proportion of graduate teachers and proportion of teachers who have a professional qualification.

[28]Previous studies have found that private schools tend to locate in villages where government schools are of lower quality, villages which are more populated, are close to district head-quarter and have better road access (Muralidharan and Kremer, 2008; Pal, 2010).

4.5.2 Results of Cost-Interactions

To analyze the impact of cost, the age-group specific *cost difference* variable is constructed separately for each of the cost components along with the total cost. Then each variable is incorporated, one at a time, in the empirical model to explore the relationship between cost differences and gender gap in private school enrollment. A separate model also considers all the cost variables to find out which component has the most significant impact. In particular, *Equation* (4.1), including the cost variables and their interaction with the female dummy, is estimated.

It is likely that private schools which are better in quality have higher school fees as well. This conjecture is supported by the findings in Section 4.5.1 where we have explored the correlates of cost difference. The cost differential has been found to be significantly and positively correlated with the quality differential. Demand for private schooling comes from the perception that these schools have better quality than their government counterpart. This perception can lead the households to have higher willingness to send boys, relative to girls, in private schools where the quality differential is also larger. If this is true, then the regression will overestimate the association between cost difference and gender gap in private school choice. Contrarily, if better quality attracts not only boys, but also girls from the households, then the negative effect of cost will be underestimated (Glick, 2008).

To separate out the effect of quality from cost, I include the block level difference in the average quality of private and government schools in the model, and interact it with the female dummy.[29] Since this variable is available only for the later two years, I exclude the data of 1997-98 from this part of the analysis.[30] Moreover, once the sample selection model following Wooldridge's method is estimated, I find that the inverse mills ratios are neither individually

[29]As described in footnote 27 in page 110, this variable is an index of school quality. The information is obtained from the DISE data for the years 2007-08 and 2010-11.

[30]Even if we estimate the model including data from 1997-98 and following Wooldridge's (1995) method, the results are found to be significant. However, it cannot address the concern that the effects found in these regressions are a mix of both cost and quality of schooling.

nor jointly significant in this case.[31] Therefore, I estimate a standard fixed effects model considering the enrolled children.

Table 4.5 summarizes the results of this analysis. All the cost variables presented in the regressions are in real terms and are expressed in hundreds of rupees. The specifications are identical to the main model, with additional variables capturing cost and quality differences, and the interaction terms. The first column shows the regression where the female dummy has not been interacted with the cost or the quality difference. It indicates that in 2007-08 and 2010-11, the average gender gap in private school enrollment has been 6.5 percentage points. The quality difference is positive and signficant, indicating that private school enrollment is higher when the quality of these schools, relative to their government counterpart, is also better. In the next columns, the regressions include the interaction terms of female dummy with cost and quality differences. While quality difference has a significant positive effect on private school enrollment of both boys and girls, its effect on the gender gap is not significant in any of the regressions. Column (2) considers the difference in the total cost of schooling, including the direct (fees, books and uniform) and transport and other cost. We find that the interaction term is negative and significant, but both the female dummy and the cost difference variable are not significant individually. The results are very similar in the third column which considers the direct cost of schooling. Next, columns (4) to (7) present the results of the interaction between female dummy and cost differences calculated separately for each of the cost components. Once the cost components are decomposed, we find that the difference in school fees is the only factor to have a significant relationship with the gender gap in private school choice. Even in the last column which considers the interaction with all these individual cost components together, the effect of school fees is significant and the magnitude is also unchanged. Except the school fees interaction, none of the other interaction terms are significant.

[31] This finding makes sense because unlike in 1997-98 when enrollment rates were relatively lower and there was a significant gender gap in overall enrollment, in 2007-08 and 2010-11 data, the enrollment rates for both boys and girls are higher and the gender gap has become insignificant.

The results on relative cost point out a significant association between the gender gap and cost of private schooling. It suggests that households may be discouraged to send their girls to private schools which are fee charging and requires higher investment than government schooling. In 2010-11, the average difference in school fees between private and government schools is Rs. 958. While the estimated gender gap in these later two years has been 6.5 percentage points, the point estimate of the coefficient of fees-interaction indicates that the average difference in school fees is associated with almost 5.7 percentage point difference in the gender gap in private school enrollment.[32]

4.5.3 Robustness

In spite of including a large number of control variables, household fixed effects and district level time fixed effects, it may still be contended that the village level cost difference in private versus government schools is endogenous. Firstly, the cost is aggregated from the actual expenditure incurred by the households, hence it is an equilibrium outcome. Secondly, as illustrated in Section 4.5.1, the cost of private schooling is dependent on several village level characteristics. The main model includes the village level factors that appear to be significant determinants of the cost differential. Yet, there can be village level unobservable effects leading to omitted variable bias. For instance, private schools may charge a higher fee in places where demand for private schooling is already higher, resulting in a reverse causation. Further, if the demand for private schooling stems more from boys than girls, then the pro-male bias in private school enrollment would lead to higher cost of private schooling in the village. In this section I address some of these concerns and find out if the relationship between the cost differential and the gender gap is robust.

[32] Since the quality difference variable is available at the block level, in an additional specification, I include the cost variables calculated on average expenditures in the block. The association between the female dummy and the difference in school fees is still found to be significant and negative in that regression.

Village Level Time-Variant Unobservables

This section presents results of regressions which include village by time fixed effects in the model. The village by time fixed effects control for all unobservable village specific characteristics which are not only fixed, but also vary over time. This specification takes into account the differential growth of prosperity and demand for schooling among the villages.

The results are presented in $Table$ 4.6. The findings are almost identical with the last set of results shown in $Table$ 4.5. Even after controlling for village level unobserved factors, we find a significant relation between the cost differences and gender gap. Particularly, the difference in school fees remains as an important factor in explaining why households are unwilling to choose private schools for their daughters as compared to their sons.

Assessing the Extent of Omitted Variable Bias

Since the cost differential is potentially endogenous, the objective of this robustness exercise is to investigate further if the estimated coefficient of the interaction term is confounded due to omitted variables. By including the village specific time fixed effects, the model takes care of all the unobserved variables at the village level. Nevertheless, the interaction between female dummy and the cost difference variable can still be endogenous. To see how, let us assume that the villagers have unobserved pro-male attitude. This attitude has differential effect on a child's school choice decision depending on the gender of the child. Besides, cultural norms and preferences are more biased against adolescent girls as compared to the younger girls. Therefore, the correlation between attitude and gender also varies depending on whether the child is in primary or post-primary age-group. In other words, the village level attitude variable has an interaction effect through the female dummy and the age-group in which the child belongs. At the same time, private schools set their fees depending on the demand. If the demand for private schooling is higher in places where the pro-male attitude is more pronounced, then the interaction between female dummy and the cost difference variable will also capture the effect of this attitude variable. Since the cost difference variable is defined at the village level for each age-group, therefore, a model that includes village-by-age-group specific time fixed effects will take into account the endogeneity of the cost variable. However, the interaction of gender and cost will still be confounded with the interaction of gender and unobserved attitude.

In this section, I examine the extent of such kind of omitted variable bias following a method prescribed by Altonji et al. (2005) and recently extended by Oster (2013). This strategy aims to draw inference about the bias due to unobservables by investigating the movements in the coefficient of interest, along with the movements in the R-squared value, as more control variables are included in the regression. The method relies on the "proportional selection assumption" (PSA), i.e., selection on observables is proportional to the selection on unobservables. This relation is expressed by the coefficient of proportionality, δ. Under PSA, Oster (2013) identifies the omitted variable bias and derives a consistent estimator of the coefficient. This estimator is a function of two parameters, δ and R_{max}, where R_{max} is the R-squared of the hypothetical regression which includes the complete set of controls involving the unobservable variables as well.

Based on some reasonable values (to be discussed later) of δ and R_{max}, I use this method to derive a "bounding set" which contains the true effect of cost differential on the gender gap in private schooling. Note that after controlling for village-by-age-group specific time fixed effects, the interaction between female dummy and the cost difference variable can be endogenous if there is some village level unobserved factor (e.g., pro-male attitude) which is correlated with cost and which also has an interaction effect through the gender of the child. If the village-by-age-group specific time fixed effects are interacted with the female dummy and included in the specification, then they would completely subsume the effect of such unobservable factor. The R-squared from the hypothetical regression controlling for the unobserved confounding factor cannot be greater than the R-squared estimated from this regression. Therefore, the R-squared calculated from this particular regression is a suitable value for R_{max}. The regression with village-by-age-group-by-gender specific time fixed effects yields an R-squared of 0.7 which is used as the value of R_{max}.[33] Next, we need to assume a reasonable value to represent the proportionality constant, δ. If we consider that the observable variables are at least as important as the unobservables, then it is plausible to conceptualize that $\delta \in [0, 1]$. A value of $\delta = 1$ will be in accordance with the assumption that the observables and the unobservables have equal effects on the coefficient of interest (Altonji et al., 2005). I present the results of this exercise in *Table* 4.7 assuming $\delta = 1$.

Each row of *Table* 4.7 corresponds to a separate regression that estimates the interaction between female dummy and the respective cost difference variable. I consider the difference in aggregate cost, direct cost, as well as each of the cost components separately. Column (1) of this table shows the result from the regression that does not control for any explanatory variable other than the interaction term. The second column reports the "controlled coefficient", which is estimated from the regression where all other control variables, including the household fixed effects and village-by-age-group specific time fixed effects are incorporated.[34] The final column identifies an interval which is bounded by the controlled coefficient on one side, and by the identified coefficient on the other side. For the given value of $R_{\max}$ and any value of $\delta \in [0, 1]$, the true coefficient lies in this identified set. For each of the regressions, the identified set suggests that even after accounting for the potential bias due to unobservables, there is very little change in the estimated coefficient from the controlled regression. This robustness exercise indicates that the point estimate of the interaction of female dummy with the difference in school fees is quite stable at -0.006. Similar inferences can be drawn for total cost and direct cost differences which also have significant effect on the gender gap. Therefore, this analysis gives more credence to the finding that the difference in direct cost of schooling, school fees in particular, is a significant determinant of intra-household gender gap in private

[33] Of course, this regression does not allow us to identify the interaction between female dummy and the cost difference variable because this interaction is also absorbed by the village-by-age-group-by-gender specific time fixed effects. However, our objective is only to estimate the R-squared of this regression.

[34] Note that in Section 4.5.3, village level time-variant unobservables are controlled by including village-by-year fixed effects. Column (2) of *Table* 4.7 refers to the specification which further controls for unobserved effects that vary with village, year and age-group too. The coefficients of the interaction of gender and cost are found to be very similar in these two specifications.

school choice.

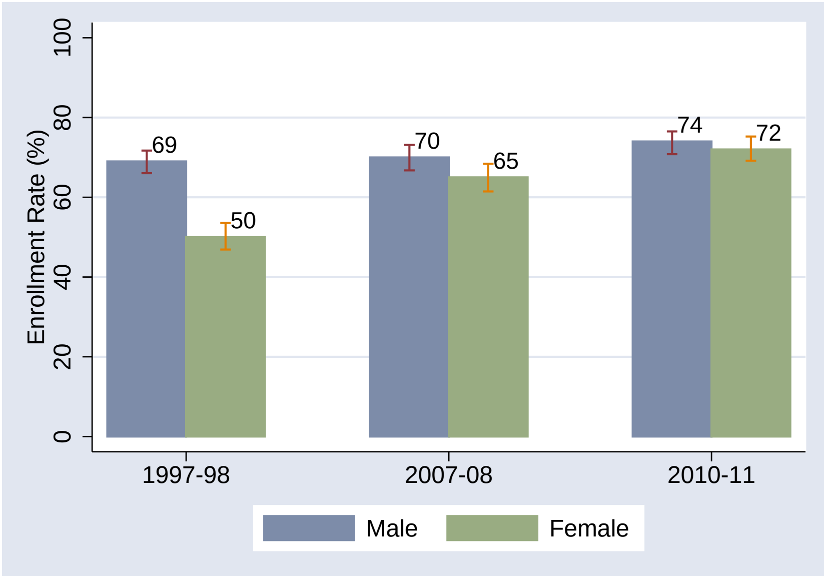
4.6 Conclusion

Using a three-period longitudinal data, this paper estimates a correlated unobserved effects model that also incorporates non-random enrollment decision for children to identify the presence of intra-household gender disparity in private school choice. The analysis suggests that there is a 5.4 percentage point gender gap in private school enrollment. Contrary to the trend in overall enrollment rates, the gender gap in private school enrollment is widening over time. This finding reveals that households choose to provide their sons rather than daughters with an education which is more expensive and which they perceive to be better in quality. Further, the study investigates if the gender inequality is driven by the cost-differential between private and government schools. The regression analysis points out that the gender gap and the cost difference are closely-knit. Female disadvantage is higher when private schools are more expensive. The effect of cost difference remains significant even after taking into account the difference in quality of private and government schools. Further robustness exercise suggests that the estimated effect of cost difference is unlikely to be confounded by unobserved factors. Although we find that the difference in quality between private and government schools induce higher demand for private schooling, unlike cost, it has no differential effect for boys and girls.

This study shows that in an era when basic education is treated as a fundamental right of every child, and government schools are made more accessible by reducing the cost of schooling, the pro-male bias in educational investment within the households has become prevalent in the choice of more expensive and perceptibly better quality schools. Behind the apparent progress in the overall enrollment rates of girls, there is a more nuanced gender story that needs careful consideration by the policy makers to achieve the goal of equality in educational opportunity.

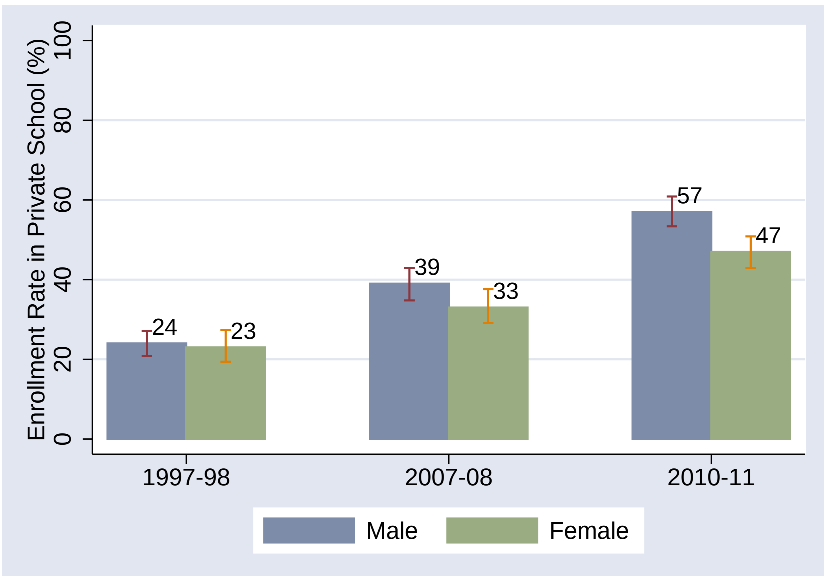
Figures and Tables for Chapter 4

Figure 4.1: Enrollment rates by gender (age 6-19 years)



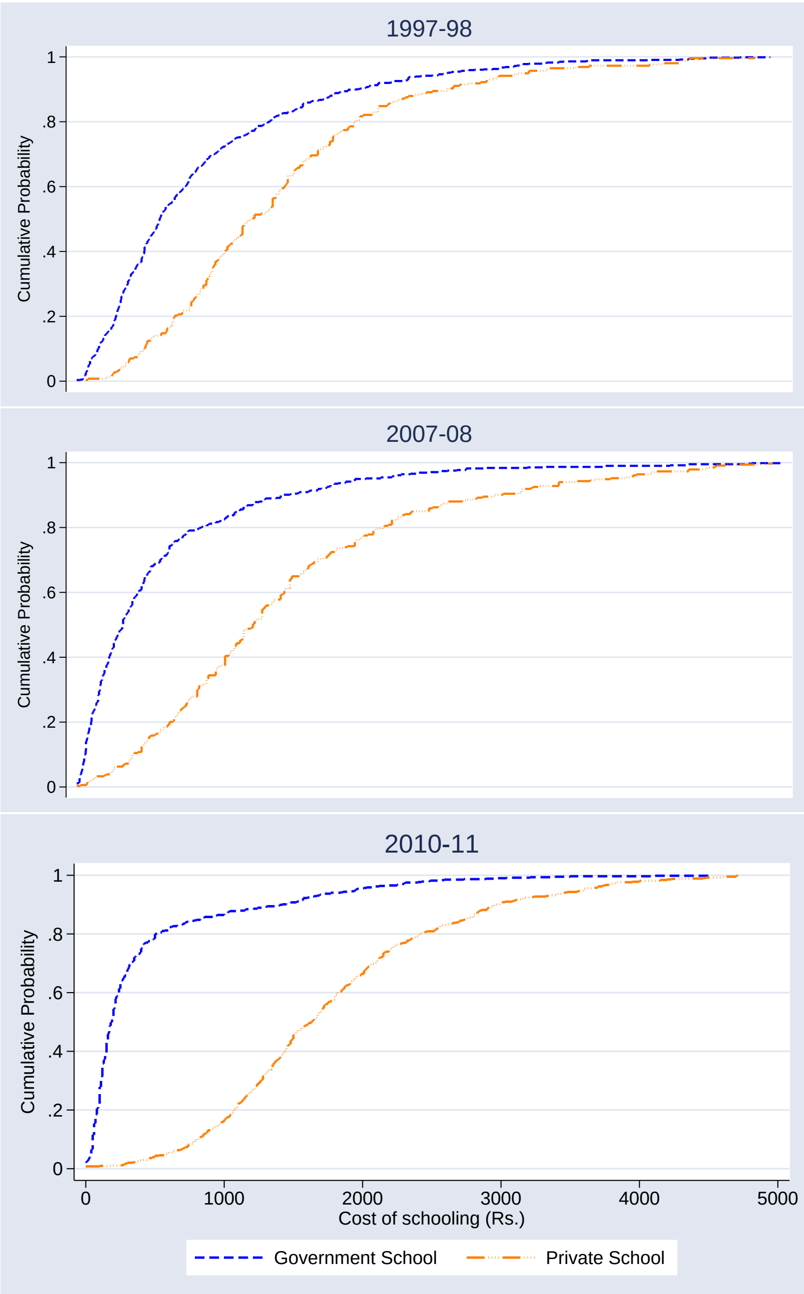
Source: SLC data.

Figure 4.2: Private school enrollment rates for enrolled children by gender (age 6-19 years)



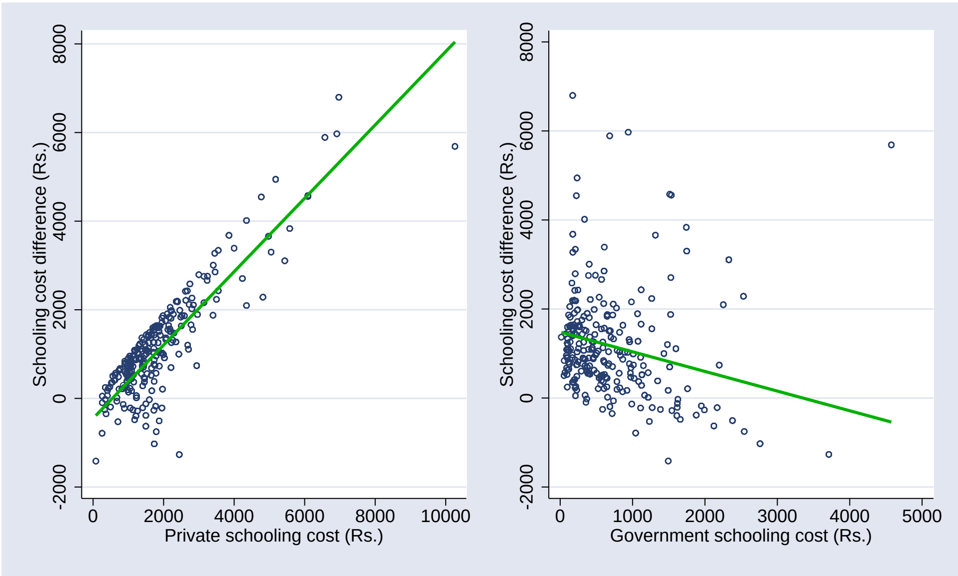
Source: SLC data.

Figure 4.3: Cumulative distribution of direct expenditure in government and private schools



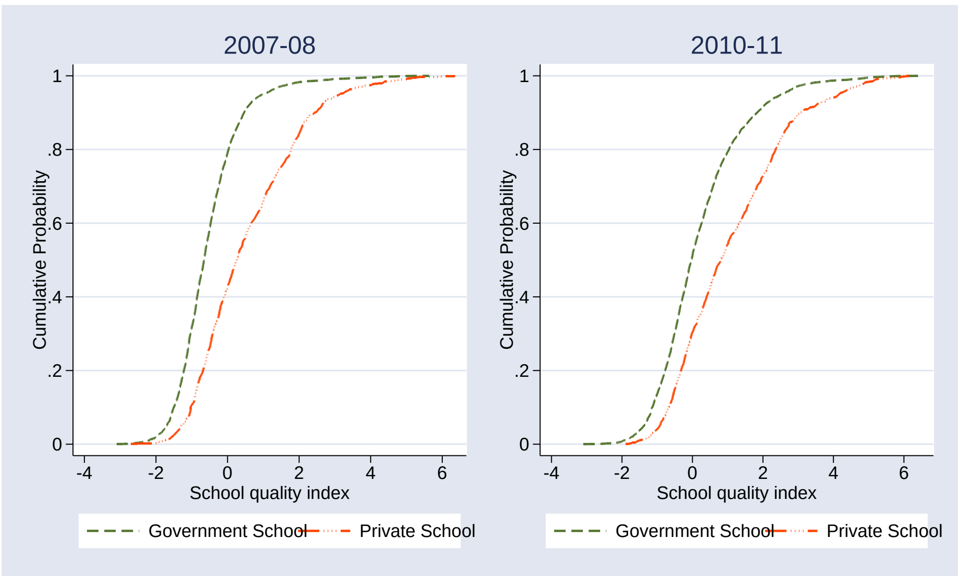
Source: SLC data.

Figure 4.4: Scatterplot with linear fit to show relation between schooling cost difference and private/government schooling cost



Source: SLC data.

Figure 4.5: Cumulative distribution of school quality in government and private schools



Source: DISE data (considering the SLC blocks).

Table 4.1: Effect of gender on private school choice within household

	Dep Var: Private school choice	
Variables	(1)	(2)
Female	-0.054***	-0.046***
	(0.016)	(0.015)
Age (years)	0.058***	0.012
	(0.022)	(0.016)
Age squared	-0.002**	-0.000
	(0.001)	(0.001)
Birth order	-0.032***	-0.033***
	(0.012)	(0.012)
Mother literate (dummy)	0.010	0.017
	(0.045)	(0.046)
Father literate (dummy)	-0.022	-0.040
	(0.045)	(0.044)
Household head literate (dummy)	-0.014	-0.019
	(0.063)	(0.063)
Household head female (dummy)	0.039	0.037
	(0.078)	(0.078)
Number of female children	-0.017	-0.016
	(0.019)	(0.019)
Number of male children	0.003	0.003
	(0.016)	(0.016)
Household size	0.005	0.007
	(0.007)	(0.007)
Wealth index	-0.005	-0.009
	(0.013)	(0.013)
Prop private schools	0.145**	0.124*
	(0.067)	(0.066)
Quality of government primary schools	0.002	0.001
	(0.012)	(0.012)
Village population (thousands)	0.003	-0.005
	(0.025)	(0.025)
Road access (pucca)	0.012	0.027
	(0.045)	(0.045)
Facility index	0.061**	0.056**
	(0.025)	(0.025)
Night lights	-0.012	-0.009
	(0.015)	(0.015)
Distance to district headquarter * Year 1997-98	-0.009**	-0.009**
	(0.004)	(0.004)
Distance to district headquarter * Year 2007-08	-0.004	-0.004
	(0.004)	(0.004)
Distance to district headquarter * Year 2010-11	-0.005	-0.005
	(0.004)	(0.004)
Sex ratio * Year 1997-98	-0.000	-0.000
	(0.001)	(0.001)
Sex ratio * Year 2007-08	-0.001	-0.001
	(0.001)	(0.001)
Sex ratio * Year 2010-11	-0.001	-0.001
	(0.001)	(0.001)
Constant	1.156**	1.284**
	(0.574)	(0.574)
District by time fixed effects	Yes	Yes
Household fixed effects	Yes	Yes
Observations	3,454	3,454
R-squared	0.261	0.258
Wald test p-value (Selection)	0.0152	
Wald test p-value (Fixed Effects)	0.0134	0.0185

Notes: Bootstrapped standard errors (clustered at the household level) based on 500 replications are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. Column (1) corrects for selection bias. Column (2) is without selection correction.

Table 4.2: Year wise estimated gender gap in private school choice

	(1)
Female * Year 1997-98	0.007
	(0.028)
Female * Year 2007-08	-0.050*
	(0.028)
Female * Year 2010-11	-0.089***
	(0.024)
Other control variables	Yes
District by time fixed effects	Yes
Household fixed effects	Yes
Observations	3,454
R-squared	0.263
Wald test p-value (Selection)	0.0597
Wald test p-value (Fixed Effects)	0.0285

Notes: Bootstrapped standard errors (clustered at the household level) based on 500 replications are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.3: Age-group wise estimated gender gap in private school choice

	(1)
Female * Primary	-0.068***
	(0.021)
Female * Post-primary	-0.042*
	(0.022)
Other control variables	Yes
District by time fixed effects	Yes
Household fixed effects	Yes
Observations	3,454
R-squared	0.263
Wald test p-value (Selection)	0.0154
Wald test p-value (Fixed Effects)	0.0275

Notes: Bootstrapped standard errors (clustered at the household level) based on 500 replications are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4.4: Analysis of schooling cost difference between private and government schools

	Dep Var: Cost difference (pvt-govt)		
Variables	(1)	(2)	(3)
Road access (pucca)	230.0***	76.17	201.8***
	(42.10)	(64.05)	(42.51)
Facility index	118.5***	100.7***	117.1***
	(14.15)	(20.06)	(14.18)
Night lights	49.68***	56.32***	50.67***
	(8.553)	(8.776)	(8.559)
Distance to district headquarter	1.300	-1.452	1.421
	(1.237)	(1.558)	(1.250)
Village population (thousands)	-64.05***	-29.94**	-70.90***
	(13.31)	(14.46)	(13.39)
Sex ratio	-0.994***	-0.692***	-0.991***
	(0.131)	(0.145)	(0.133)
Prop private schools	-266.7***	-205.5***	-250.3***
	(52.08)	(60.07)	(51.72)
Quality of government primary schools	-40.07***	-4.945	-40.56***
	(10.54)	(12.88)	(10.52)
School quality diff		320.1***	
		(60.05)	
Female			11.39
			(32.16)
Age (years)			5.122
			(7.202)
Birth order			26.23
			(16.46)
Mother literate (dummy)			88.44**
			(41.66)
Father literate (dummy)			-42.23
			(41.87)
Household head literate (dummy)			36.12
			(42.53)
Household head female (dummy)			4.077
			(61.39)
Number of female children			-19.69
			(17.22)
Number of male children			-12.12
			(17.08)
Household size			-3.939
			(6.209)
Wealth index			33.35***
			(10.12)
Post-primary age-group	-254.6***	-271.9***	-258.3***
	(29.95)	(37.22)	(53.16)
Constant	2,540***	1,768***	2,489***
	(156.6)	(187.9)	(188.8)
District fixed effects	Yes	Yes	Yes
District by time fixed effects	Yes	Yes	Yes
Observations	5,175	3,282	5,175
R-squared	0.288	0.231	0.293

Notes: Robust standard errors are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. Regressions in columns (1) and (3) consider all three rounds of SLC data. In column (2), an additional variable capturing the block level difference in quality of private and government schools is included. This variable is sourced from the DISE data, and is not observed for 1997-98. Hence this regression considers only the sample of 2007-08 and 2010-11.

Table 4.5: Interaction of gender with cost difference between private and government schools

	Dependent Variable: Private school choice						
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.065***	-0.021	-0.020	-0.031	-0.049	-0.060	-0.040
	(0.023)	(0.053)	(0.054)	(0.049)	(0.057)	(0.048)	(0.057)
Total cost diff		-0.001					
		(0.002)					
Female * Total cost diff		-0.003**					
		(0.002)					
Direct cost diff			-0.002				
			(0.002)				
Female * Direct cost diff			-0.004**				
			(0.002)				
Fees diff				-0.003			-0.004
				(0.002)			(0.002)
Female * Fees diff				-0.006**			-0.006**
				(0.002)			(0.003)
Books & uniform diff					-0.001		0.001
					(0.004)		(0.005)
Female * Books & uniform diff					-0.003		0.002
					(0.005)		(0.005)
Transport & others diff						0.002	0.004
						(0.004)	(0.004)
Female * Transport & others diff						-0.003	-0.001
						(0.005)	(0.005)
School quality diff	0.186*	0.199*	0.198*	0.205*	0.186	0.186*	0.204*
	(0.112)	(0.113)	(0.113)	(0.112)	(0.114)	(0.113)	(0.112)
Female * School quality diff		0.011	0.013	0.019	-0.000	-0.000	0.019
		(0.043)	(0.042)	(0.042)	(0.045)	(0.044)	(0.042)
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District by time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,312	2,312	2,312	2,312	2,312	2,312	2,312
R-squared	0.617	0.621	0.622	0.623	0.617	0.617	0.624

Notes: Robust clustered standard errors (clustered at the household level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. All cost differences are expressed in hundreds of rupees.

Table 4.6: Robustness: Controlling for village level unobservables

	Dependent Variable: Private school choice						
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.067***	-0.026	-0.024	-0.034	-0.058	-0.064	-0.047
	(0.024)	(0.052)	(0.053)	(0.048)	(0.057)	(0.047)	(0.056)
Total cost diff		-0.002					
		(0.002)					
Female * Total cost diff		-0.003**					
		(0.002)					
Direct cost diff			-0.003				
			(0.002)				
Female * Direct cost diff			-0.004**				
			(0.002)				
Fees diff				-0.004			-0.004
				(0.003)			(0.003)
Female * Fees diff				-0.006***			-0.006***
				(0.002)			(0.002)
Books & uniform diff					-0.003		-0.002
					(0.004)		(0.005)
Female * Books & uniform diff					-0.002		0.003
					(0.005)		(0.005)
Transport & others diff						-0.000	0.003
						(0.004)	(0.005)
Female * Transport & others diff						-0.004	-0.002
						(0.005)	(0.004)
Female * School quality diff		0.014	0.015	0.023	0.003	0.003	0.026
		(0.042)	(0.042)	(0.041)	(0.044)	(0.043)	(0.042)
Other control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Village by time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,312	2,312	2,312	2,312	2,312	2,312	2,312
R-squared	0.635	0.639	0.639	0.641	0.635	0.635	0.641

Notes: Robust clustered standard errors (clustered at the household level) are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%. All cost differences are expressed in hundreds of rupees.

Table 4.7: Robustness: Assessing the extent of omitted variable bias

Variable of Interest	Uncontrolled Coefficient ($Standard\ Error$) [R^2]	Controlled Coefficient ($Standard\ Error$) [R^2]	Identified Set $R_{max} = 0.7$, $\delta = 1$
	(1)	(2)	(3)
Female * Total cost diff	-0.0053 (0.001) [0.016]	-0.0030 (0.002) [0.663]	[-0.0030, -0.0029]
Female * Direct cost diff	-0.0057 (0.001) [0.015]	-0.0037 (0.002) [0.663]	[-0.0037, -0.0036]
Female * Fees diff	-0.0077 (0.001) [0.013]	-0.0060 (0.002) [0.664]	[-0.0060, -0.0059]
Female * Books & uniform diff	-0.0121 (0.003) [0.010]	-0.0008 (0.005) [0.662]	[-0.0008, -0.0002]
Female * Transport & others diff	-0.0141 (0.005) [0.006]	-0.0026 (0.005) [0.662]	[-0.0026, -0.0020]

Notes: Each row corresponds to a separate regression of private school choice which estimates the interaction between female dummy and the cost difference variable. The "uncontrolled coefficient" is from the regression which does not have any other control variable. The "controlled coefficient" is obtained from the regression which controls for the full set of explanatory variables (as in *Table* 4.1) and also includes school quality difference interacted with gender, household fixed effects and village-by-agegroup-by-time fixed effects. These regressions consider the 2007-08 and 2010-11 sample from the SLC data. For each of these regressions, only the coefficient of the interaction between female dummy and the respective cost variable is reported here. The "identified set" contains the true effect once the potential omitted variable bias is taken into account (Oster, 2013). These results are obtained through the Stata command `psacalc`.

Table 4.$A1$: Summary statistics

Variables	1997-98			2007-08			2010-11		
	Obs	Mean	SD	Obs	Mean	SD	Obs	Mean	SD
Child level variables									
Enrollment	1893	0.60	0.49	1525	0.68	0.47	1757	0.73	0.44
Private school enrollment	1142	0.24	0.43	1030	0.36	0.48	1282	0.52	0.50
Female	1893	0.46	0.50	1525	0.48	0.50	1757	0.48	0.50
Age (years)	1893	11.64	3.86	1525	12.74	3.95	1757	12.48	4.04
Birth order	1893	2.32	1.52	1525	2.25	1.30	1757	2.41	1.57
Mother literate (dummy)	1893	0.15	0.36	1525	0.19	0.40	1757	0.24	0.43
Father literate (dummy)	1893	0.50	0.50	1525	0.57	0.50	1757	0.61	0.49
Household level variables									
Household head literate (dummy)	1893	0.45	0.50	1525	0.47	0.50	1757	0.52	0.50
Household head female (dummy)	1893	0.04	0.20	1525	0.09	0.29	1757	0.07	0.25
Number of female children	1893	1.72	1.34	1525	1.74	1.15	1757	1.94	1.52
Number of male children	1893	1.99	1.31	1525	1.84	1.12	1757	1.97	1.28
Household size	1893	8.76	4.66	1525	7.96	3.06	1757	9.18	4.47
Wealth index	1893	-0.60	1.43	1525	-0.13	1.90	1757	0.79	2.47
Religion - Hindu	1893	0.94	0.23	1525	0.94	0.23	1757	0.96	0.19
Religion - Muslim	1893	0.06	0.23	1525	0.06	0.23	1757	0.04	0.19
Caste - General	1893	0.19	0.39	1525	0.17	0.38	1757	0.16	0.37
Caste - SC/ST	1893	0.26	0.44	1525	0.28	0.45	1757	0.29	0.46
Caste - Backward	1893	0.55	0.50	1525	0.55	0.50	1757	0.54	0.50
Village level variables									
Proportion of private schools	1893	0.21	0.35	1525	0.14	0.27	1757	0.22	0.27
Quality of government primary schools	1893	-1.55	2.42	1525	0.84	1.53	1757	1.06	1.21
School quality difference	0			1525	0.95	0.39	1757	0.93	0.67
Village population (thousands)	1893	1.58	0.77	1525	2.82	1.42	1757	2.98	1.48
Road access (*pucca*)	1893	0.31	0.46	1525	0.52	0.50	1757	0.92	0.27
Facility index	1893	-0.10	1.18	1525	-0.09	1.26	1757	0.12	1.09
Night lights	1893	2.12	2.06	1525	3.62	3.32	1757	5.19	4.02
Distance to district headquarter	1893	33	17	1525	32	17	1757	32	17
Sex ratio	1893	907	101	1525	898	107	1757	903	107
Total cost difference	1893	702	1401	1525	1276	1251	1757	1861	1208
Direct cost difference	1893	646	1197	1525	1210	1178	1757	1599	945
Fees difference	1893	436	795	1525	727	933	1757	958	726
Books & uniform cost difference	1893	210	564	1525	484	431	1757	641	470
Transport & other cost in difference	1893	56	292	1525	66	211	1757	263	551

Source: SLC data for all variables except *School quality difference* (DISE data), *Night lights* (NOAA website), and *Sex ratio* (Census 2001 data).

Table 4.*A*2: Comparison of means of child and household level characteristics between enrolled and not-enrolled children

Variables	Enrolled (1)	Not-enrolled (2)	Difference (1)-(2)	p-value (H0: Difference=0)
Female	0.439	0.536	-0.096	0.000
Age	11.499	13.761	-2.261	0.000
Birth order	2.500	1.988	0.512	0.000
Mother literate	0.237	0.116	0.120	0.000
Father literate	0.651	0.374	0.276	0.000
Household head literate	0.539	0.360	0.180	0.000
Household head female	0.065	0.069	-0.004	0.559
Number of female children	1.807	1.777	0.030	0.457
Number of male children	1.982	1.856	0.126	0.001
Household size	8.850	8.299	0.551	0.000
Wealth index	0.229	-0.430	0.659	0.000
Religion - Muslim	0.051	0.049	0.001	0.843
Caste - SC/ST	0.245	0.341	-0.096	0.000
Caste - Backward	0.559	0.525	0.033	0.023

Notes: Out of 5175 children in all three years, 3454 (66.74 percent) children are enrolled and 1721 (33.26 percent) children are not enrolled in any school.

Table 4.A3: Child level regression of schooling expenditure (on direct cost) for children in government and private schools

Variables	Schooling Expenses	
	Government	Private
	(1)	(2)
Road access (pucca)	64.44	-77.75
	(47.07)	(148.3)
Facility index	15.04	59.83
	(16.72)	(92.11)
Night lights	-9.032	14.58
	(6.974)	(19.97)
Distance to district headquarter	-1.422	-5.354
	(1.238)	(3.294)
Village population (thousands)	1.261	-16.10
	(15.23)	(49.84)
Sex ratio	-0.294*	-0.886**
	(0.152)	(0.384)
Prop private schools	-155.2***	-303.6*
	(55.22)	(175.0)
Quality of government primary schools	-44.74***	-72.07**
	(10.46)	(30.50)
Female	-123.3***	-273.3***
	(32.77)	(104.5)
Age (years)	163.4***	122.0***
	(8.850)	(17.34)
Birth order	91.17***	-18.01
	(19.67)	(48.12)
Mother literate (dummy)	205.5***	537.4***
	(42.39)	(113.7)
Father literate (dummy)	203.1***	-8.161
	(51.86)	(124.9)
Household head literate (dummy)	-38.09	24.81
	(45.34)	(110.9)
Household head female (dummy)	58.51	200.5
	(57.12)	(150.5)
Number of female children	-84.29***	-77.66
	(21.92)	(54.37)
Number of male children	-108.9***	-11.32
	(22.53)	(52.08)
Household size	1.306	-10.60
	(5.741)	(19.76)
Wealth index	64.80***	225.8***
	(16.25)	(35.81)
Constant	-1,204***	4,124***
	(232.7)	(1,021)
District fixed effects	Yes	Yes
District by time fixed effects	Yes	Yes
Observations	2,099	1,290
R-squared	0.458	0.254

Notes: Robust standard errors are in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Bibliography

[1] Afridi, Farzana, Vegard Iversen and M. R. Sharan (2013), "Women political leaders, corruption and learning: Evidence from a large public program in India", working paper, IZA and International Growth Centre (IGC).

[2] Alderman, Harold and Elizabeth M. King (1998), "Gender Differences in Parental Investment in Education", *Structural Change and Economic Dynamics*, 9(4): 453–468.

[3] Alderman, Harold, Peter F. Orazem and Elizabeth M. Paterno (2001), "School Quality, School Cost and the Public/Private School Choices of Low-Income Households in Pakistan", *Journal of Human Resources*, 36(2): 304–326.

[4] Altonji, Joseph G., Todd E. Elder and Christopher R. Taber (2005), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools", *Journal of Political Economy*, 113(1): 151–184.

[5] Anderson, Siwan and Mukesh Eswaran (2009), "What determines female autonomy? Evidence from Bangladesh", *Journal of Development Economics*, 90(2): 179–191.

[6] Andrabi, Tahir, Jishnu Das and Asim I. Khwaja (2013), "Students Today, Teachers Tomorrow: Identifying constraints on the provision of education", *Journal of Public Economics*, 100: 1–14.

[7] Angrist, Joshua and Jorn-Steffen Pischke (2009), "Mostly harmless econometrics: An empiricist's companion", Princeton University Press, Princeton.

[8] ASER (2005 to 2013), Annual Status of Education Report, Pratham Resource Centre, New Delhi.

[9] Azam, Mehtabul (2012), "The Impact of Indian Job Guarantee Scheme on Labor Market Outcomes: Evidence from a Natural Experiment", IZA Discussion Paper 6548.

[10] Azam, Mehtabul and Geeta Gandhi Kingdon (2013), "Are Girls the Fairer Sex in India? Revisiting Intra-Household Allocation of Education Expenditure", *World Development*, 42: 143–164.

[11] Banerjee, Abhijit and Esther Duflo (2005), "Growth Theory through the Lens of Development Economics", in P. Aghion and S. N. Durlauf (eds.) *Handbook of Economic Growth*, 1A, North-Holland, Amsterdam: 473–552.

[12] Banerjee, Abhijeet and Esther Duflo (2011), *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, PublicAffairs, United States.

[13] Basu, Deepankar and Robert de Jong (2010), "Son Targeting Fertility Behavior: Some Consequences and Determinants", *Demography*, 47(2): 521–536.

[14] Blumberg, R. L. (1988), "Income under female versus male control: hypotheses from a theory of gender stratification and data from the third world", *Journal of Family Issues*, 9(1): 51–84.

[15] Bommier, Antoine and Sylvie Lambart (2000), "Education Demand and Age at School Enrollment in Tanzania", *The Journal of Human Resources*, 35(1): 177–203.

[16] Brown, Philip H. and Albert Park (2002), "Education and poverty in rural China", *Economics of Education Review*, 21(6): 523–541.

[17] Burde, Dana and Leigh L. Linden (2012), "The Effect of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan", Working Paper 18039, National Bureau of Economic Research.

[18] Burke, Kathleen and Kathleen Beegle (2004), "Why Children Aren't Attending School: The Case of Northwestern Tanzania", *Journal of African Economics*, 13(2): 333–355.

[19] Chamberlain, Gary (1980), "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47(1): 225–238.

[20] Chiappori, Pierre-Andr (1988), “Nash-bargained household decisions: A Comment”, *International Economic Review*, 29(4): 791–796.

[21] Clark, Shelley (2000), “Son Preference and Sex Composition of Children: Evidence from India”, *Demography*, 37(1): 95–108.

[22] Colclough, Christopher, Geeta Kingdon and Harry Patrinos (2010). “The Changing Pattern of Wage Returns to Education and its Implications”, *Development Policy Review*, 28(6), pp 733-747.

[23] Das Gupta, Monica (1987), “Selective Discrimination against Female Children in Rural Punjab, India”, *Population and Development Review*, 13(1): 77–100.

[24] Dasgupta, Aparajita (2013), “Can major public works policy buffer negative shocks in early childhood? Evidence from Andhra Pradesh, India”, Young Lives Working Paper 112.

[25] Deolalikar, Anil B. (1993), ”Gender Differences in the Returns to Schooling and in School Enrollment Rates in Indonesia”, *Journal of Human Resources*, 28(4): 899–932.

[26] Desai, Sonalde, Amaresh Dubey, Reeve Vanneman, Rukmini Banerji (2008), “Private Schooling in India: A New Educational Landscape”, in *India Policy Forum*, edited by Suman Bery, Barry Bosworth and Arvind Panagariya, Sage, New Delhi.

[27] Duflo, Esther (2001), “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment”, *American Economic Review*, 91(4): 795–813.

[28] Duflo, Esther (2012), “Women Empowerment and Economic Development”, *Journal of Economic Literature*, 50(4): 1051–1079.

[29] Duraisamy, P. (2002), “Changes in Returns to Education in India, 1983-94: By Gender, Age-cohort and Location”, *Economics of Education Review*, 21(6): 609–622.

[30] Dustmann, Christian and Mara Engracia Rochina-Barrachina (2007), "Selection Correction in Panel Data Models: An application to the estimation of females' wage equation", *Econometrics Journal*, 10: 263–293.

[31] Filmer, Deon (2007), "If You Build It, Will They Come? School Availability and School Enrollment in 21 Poor Countries", *Journal of Development Studies*, 43(5): 901–928.

[32] Gertler, Paul and Paul Glewwe (1990), "The Willingness to Pay for Education in Developing Countries: Evidence from Rural Peru", *Journal of Public Economics*, 42(3): 251–275.

[33] Glick, Peter (2008), "What Policies will Reduce Gender Schooling Gaps in Developing Countries: Evidence and Interpretation", *World Development*, 36(9): 1623–1646.

[34] 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(1): 118–145.

[35] Goldin, Claudia and Lawrence F. Katz (2002), "The power of the pill: Oral contraceptives and women's career and marriage decisions", *Journal of Political Economy*, 110(4): 730–770.

[36] Grootaert, Christian, and Harry Patrinos (1999), *The Policy Analysis of Child Labor: A Comparative Study*, St. Martin's Press, New York.

[37] Handa, Sudhanshu (2002), "Raising Primary School Enrollment in Developing Countries: The Relative Importance of Supply and Demand", *Journal of Development Economics*, 69(1): 103–128.

[38] Harma, Joanna (2011), "Low cost private schooling in India: Is it pro poor and equitable?", *International Journal of Educational Development*, 31(4): 350–356.

[39] 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): 237–244.

[40] Heckman, James J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1): 153–161.

[41] Himanshu (2011), "Employment Trends in India: A Re-examination", *Economic and Political Weekly*, 46(37): 43–59.

[42] Hoddinott, John and Lawrence Haddad (1995), "Does female income share influence household expenditure? Evidence from Cote D'Ivoire", *Oxford Bulletin of Economics and Statistics*, 57(1): 77–96.

[43] Ilahi, Nadeem (1999), "Children's work and schooling under shocks: Does gender matter? Evidence from the Peru LSMS panel data", Background paper for Engendering Development, Technical Report, World Bank, Washington DC.

[44] Imbert, Clement and John Papp (2015), "Labor market effects of social programs: Evidence from India's employment guarantee", *American Economic Journal: Applied Economics*, 7(2): 233–63.

[45] Jalan, Jyotsna and Martin Ravallion (2003), "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching", *Journal of Business Economic Statistics*, 21(1): 19–30.

[46] Jayachandran, Seema (2014), "The Roots of Gender Inequality in Developing Countries", Working Paper 20380, National Bureau of Economic Research.

[47] Jensen, Robert (2002), "Fertility Preferences and Female Disadvantage: Equal Treatment, Unequal Outcomes?", mimeo, John F. Kennedy School of Government, Harvard University, Cambridge.

[48] Jensen, Robert (2010), "The (Perceived) Returns to Education and the Demand for Schooling", *Quarterly Journal of Economics*, 125(2): 515–548.

[49] Jha, Jyotsna and Ramya Subrahmanian (2005), "Secondary Education in the Indian State of Uttar Pradesh: Gender Dimensions of State Policy and Practice", United Nations Research Institute for Social Development (UNRISD).

[50] Kaur, Supreet (2014), "Nominal Wage Rigidity in Village Labor Markets", Working Paper 20770, National Bureau of Economic Research.

[51] Khera, Reetika and Nandini Nayak (2009), "Women workers and perceptions of the National Rural Employment Guarantee Act", *Economic and Political Weekly*, 44(43): 49–57.

[52] Kingdon, Geeta (1996), "The quality and efficiency of private and public education: A case-study of urban India", *Oxford Bulletin of Economics and Statistics*, 58(1): 57–82.

[53] Kingdon, Geeta Gandhi (1998), " Does the labour market explain lower female schooling in India?", *Journal of Development Studies*, 35(1): 39–65.

[54] Kingdon, Geeta, Harry Patrinos, Chris Sakellariou and Mans Soderbom (2008), "International Pattern of Returns to Education", mimeo, World Bank.

[55] Kingdon, Geeta Gandhi and Jeemol Unni (2001), "Education and Women's Labour Market Outcomes in India", *Education Economics*, 9(2): 173–195.

[56] Klonner, Stefan and Christian Oldiges (2014), "Safety net for India's poor or waste of public funds? Poverty and welfare in the wake of the world's largest job guarantee program", AWI Discussion Paper Series No. 564.

[57] Lancaster, Tony (2000), "The incidental parameter problem since 1948", *Journal of Econometrics*, 95(2): 391–413.

[58] Lavy, Victor (1996), "School Supply Constraints and Children's Educational Outcomes in Rural Ghana", *Journal of Development Economics*, 51(2): 291–314.

[59] Lincove, Jane Arnold (2009), "Determinants of Schooling for Boys and Girls in Nigeria under a Policy of Free Primary Education", *Economics of Education Review*, 28(4): 474–484.

[60] Luke, Nancy and Kaivan Munshi (2011), "Women as agents of change: Female income and mobility in India", *Journal of Development Economics*, 94(1): 1–17.

[61] Maitra, Pushkar, Sarmistha Pal and Anurag Sharma (2013), "Returns to Schooling, English Skills and Gender Gap in Private School Enrolment: Evidence from India", SSRN Working Paper, downloaded from http://ssrn.com/abstract=2341005.

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

[63] McElroy, Marjorie B. and Mary Jean Horney (1981), "Nash-bargained household decisions: Toward a generalization of the theory of demand", *International Economic Review*, 22(2): 333–349.

[64] Miguel, Edward and Michael Kremer (2004), "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities", *Econometrica*, 72(1): 159–217.

[65] Mundlak, Yair (1978), "On the Pooling of Time Series and Cross Section Data", *Econometrica*, 46(1): 69–86.

[66] Muralidharan, Karthik and Michael Kremer (2008), " Public and Private Schools in Rural India", in *School Choice International*, edited by Paul Peterson and Rajashri Chakrabarti, MIT Press, Cambridge, Massachusetts.

[67] Muralidharan, Karthik and Nishith Prakash (2013), "Cycling to School: Increasing Secondary School Enrollment for Girls in India", Working Paper 19305, National Bureau of Economic Research.

[68] Neyman, J. and Elizabeth L. Scott (1948), "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16(1): 1–32.

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

[70] Nishimura, Mikiko and Takashi Yamano (2013), " Emerging Private Education in Africa: Determinants of School Choice in Rural Kenya", *World Development*, 43: 266–275.

[71] Orazem, Peter F. and Elizabeth M. King (2007), "Schooling in Developing Countries: The Roles of Supply, Demand and Government Policy", in T.P. Schultz and J. Strauss (eds.) *Handbook of Development Economics*, 4, North-Holland, Amsterdam: 3475–3559.

[72] Oster, Emily (2009), "Proximate Sources of Population Sex Imbalance in India", *Demography*, 46(2): 325–339.

[73] Oster, Emily (2013), "Unobservable Selection and Coefficient Stability: Theory and Validation", Working Paper 19054, National Bureau of Economic Research.

[74] Pal, Sarmistha (2010), "Public Infrastructure, Location of Private Schools and Primary School Attainment in an Emerging Economy", *Economics of Education Review*, 29(5): 783–794.

[75] Pankaj, Ashok and Rukmini Tankha (2010), "Empowerment effects of the NREGS on women workers: A study in four states", *Economic and Political Weekly*, 45(30): 45–55.

[76] Pitt, Mark M., Mark R. Rosenzweig and Donna M. Gibbons (1993), "The Determinants and Consequences of the Placement of Government Programs in Indinesia", *World Bank Economic Review,* 7(3): 319–348.

[77] Psacharopoulos, George (1994), "Returns to Investment in Education: a Global Update", *World Development*, 22(9): 1325–1343.

[78] Psacharopoulos, George and Harry A. Patrinos (2004), "Returns to Investment in Education: A Further Update", *Education Economics*, 12(2): 111–134.

[79] Qian, Nancy (2008), "Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance", *Quarterly Journal of Economics*, 123(3): 1251–1285.

[80] Quisumbing, Agnes and John Maluccio (2003), "Resources at marriage and intra-household allocation: evidence from Bangladesh, Ethiopia, Indonesia, and South Africa", *Oxford Bulletin of Economics and Statistics*, 65(3): 283–327.

[81] Rosenzweig, Mark R. and T. Paul Schultz (1982), "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India", *American Economic Review*, 72(4): 803–815.

[82] Sawada, Yasuyuki and Michael Lokshin (2009), "Obstacles to school progression in rural Pakistan: An analysis of gender and sibling rivalry using field survey data", *Journal of Development Economics*, 88(2): 335–347.

[83] Schultz, T. Paul (2004), "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education", *Journal of African Economies*, 13(2): ii95–ii148.

[84] Sheahan, Megan, Yanyan Liu, Christopher Barrett and Sudha Narayanan (2014), "The political economy of MGNREGS spending in Andhra Pradesh", IFPRI Discussion Paper 01371.

[85] Skoufias, Emmanuel (1993), "Labor market opportunities and intrafamily time allocation in rural households in South Asia", *Journal of Development Economics*, 40(2): 277–310.

[86] Stock, James H. and Yogo, Motohiro (2005), "Testing for Weak Instruments in Linear IV Regression", in D. W. K. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, Cambridge.

[87] Subramaniam, Ramesh (1996), "Gender-Bias in India: The Importance of Household Fixed-Effects", *Oxford Economic Papers*, 48(2): 280–299.

[88] Thomas, Duncan (1990), "Intra-Household Resource Allocation: An Inferential Approach", *Journal of Human Resources*, 25(4): 635–664.

[89] Tooley, James, Pauline Dixon and S. V. Gomathi (2007), "Private schools and the millennium development goal of universal primary education: a census and comparative survey in Hyderabad, India", *Oxford Review of Education*, 33(5): 539–560.

[90] van der Klaauw, Wilbert (1996), "Female labour supply and marital status decisions: A life-cycle model", *Review of Economic Studies*, 63(2):199–235.

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

[92] Woodhead, Martin, Melanie Frost and Zoe James (2013), "Does growth in private schooling contribute to Education for All? Evidence from a longitudinal, two cohort study in Andhra Pradesh, India", *International Journal of Educational Development*, 33: 65–73.

[93] Wooldridge, Jeffrey M. (1995), "Selection correction for panel data models under conditional mean independence assumptions", *Journal of Econometrics*, 68(1): 115–132.

[94] Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.

[95] World Bank (2012), World Development Report, World Bank, Washington DC.

[96] Yamaguchi, Kazuo (1989), "A Formal Theory for Male-Preferring Stopping Rules of Childbearing: Sex Differences in Birth Order and in the Number of Siblings", *Demography*, 26(3): 451–465.

[97] Zimmermann, Laura (2014), "Why guarantee employment? Evidence from a large Indian public-works program", mimeo, University of Georgia.