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

Soham Sahoo*

January, 2015

Abstract

This paper explores the gender inequality within households in the decision of private versus government school choice in India. Using a three-period longitudinal data on rural households from Uttar Pradesh, a northern state of India, this paper estimates a household fixed effects model that also takes into account non-random enrollment decision. The results show that there is an intra-household gender gap of 5.4 percentage points in private school enrollment among the children aged 6 to 19 years. The female disadvantage in private school choice is significant among both younger and older children, and is rising over time. Moreover, a larger cost-difference between private and government schools is associated with a significantly higher gender gap. The village-specific difference in direct cost, particularly school fees, is the most prominent factor in explaining the gender gap. Robustness analysis considering the potential endogeneity of the cost variables indicates that the coefficients are unlikely to be driven by omitted variables.

Keywords: Private schooling, gender, schooling cost, longitudinal data, sample selection

JEL codes: I20, I24, I25

*Economics and Planning Unit, Indian Statistical Institute, Delhi. Email: soham9r@isid.ac.in.
I am grateful to Abhiroop Mukhopadhyay for his guidance in this paper. I would also like to thank participants at the European Society for Population Economics Conference (Braga), International Economic Association World Congress (Jordan), European Population Conference (Budapest), and Jawaharlal Nehru University - Young Scholars' Seminar (Delhi) for their comments. All errors are my own.

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 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; Muralidharan and Kremer, 2008; Tooley et al, 2007; 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 school-

[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.

[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.

ing 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, 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

[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

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 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

market. However, it is likely that the matching takes place based on years of schooling rather than the quality of schooling.

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

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 level 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 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 2 describes the dataset and provides a descriptive analysis of trends in private school participation. Section 3 lays out the empirical model. Sections 4 presents the results on gender gap in private school choice. Section 5 discusses the relationship between gender gap and the difference in

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

cost of private and government schooling. Section 6 concludes.

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]

The summary statistics of variables used in this study are presented in *Table* A1. 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* 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

[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.

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* 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.

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

[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).

examine the relationship between schooling cost and gender gap.

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 ($P = 0$).

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

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

The choice of private versus government school is modeled by *Equation* (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, 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

[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.

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).

3.2 The Selection Problem

We can estimate *Equation* (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* (1), or by 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 deci-

[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.

sions 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* A2). 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.

$$S^*_{cht} = \mathbf{Z}_{cht}\gamma + \delta_h + u_{cht} \; ; \; S_{cht} = 1[S^*_{cht} > 0]. \tag{2}$$

The enrollment decision modeled in *Equation* (2) has a form similar to the school choice model. The set of explanatory variables, **Z**, contains all the variables that are present in **X**; but for ease of identification, it also contains an additional variable that is validly excluded from *Equation* (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* (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* (1)) as a linear model based on the selected sample, the Heckit model requires that we estimate the selection equation (*Equation* (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.

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* (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

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

way to 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* (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,

$$\delta_h = \eta_0 + \overline{\mathbf{Z}}_h \eta + e_h, \tag{3}$$

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:

$$S^*_{cht} = \eta_0 + \overline{\mathbf{Z}}_h \eta + \mathbf{Z}_{cht}\gamma + v_{cht}; S_{cht} = 1[S^*_{cht} > 0], \tag{4}$$

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

[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.

given $\mathbf{Z}_{ch}$ and v_{cht} is linear. Thus, we have the following relationship:

$$E(\alpha_h \mid \mathbf{Z}_{ch}, v_{cht}) = \psi_0 + \overline{\mathbf{X}}_h\psi + \pi_t v_{cht}, \tag{5}$$

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}:

$$E(\varepsilon_{cht} \mid \mathbf{Z}_{ch}, v_{cht}) = E(\varepsilon_{cht} \mid v_{cht}) = \rho_t v_{cht}. \tag{6}$$

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* (5) and (6), and combine them to get the following relation:

$$\begin{aligned} E(\alpha_h + \varepsilon_{cht} \mid \mathbf{Z}_{ch}, S_{cht} = 1) &= \psi_0 + \overline{\mathbf{X}}_h\psi + (\pi_t + \rho_t)E(v_{cht} \mid \mathbf{Z}_{ch}, S_{cht} = 1) \\ &= \psi_0 + \overline{\mathbf{X}}_h\psi + \zeta_t\lambda_{cht}, \end{aligned} \tag{7}$$

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:

$$E(P_{cht} \mid \mathbf{X}_{ch}, S_{cht} = 1) = \psi_0 + \overline{\mathbf{X}}_h\psi + \mathbf{X}_{cht}\beta + \zeta_t\lambda_{cht}. \tag{8}$$

A consistent estimate of β can be obtained through *Equation* (8) following a few steps. First, the reduced form sample selection equation (*Equation* (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* (8)) as additional regressors. Then *Equation* (8) is estimated by pooled OLS method. 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* (1).

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.

[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.

Table 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 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. 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.

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* 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* 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.

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.[19]

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.[20] This indicates that the government policies of providing free and compulsory education may have been effective in reducing the cost of government schooling.[21] 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* 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.

[19]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.

[20]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.

[21]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.[22] 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* A3 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

[22] 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.

community level characteristics.[23]

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.[24] *Figure* 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. 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

[23]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 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.

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

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 phenomenon, 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.[25] I construct a school quality index by principal component analysis of several characteristics related to the school infrastructure and teachers.[26] The cumulative distribution functions of the quality index have been plotted in *Figure* 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

[25]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.

[26]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.

of private schools.[27] 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.

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* (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 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

[27] 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).

in the average quality of private and government schools in the model, and interact it with the female dummy.[28] Since this variable is available only for the later two years, I exclude the data of 1997-98 from this part of the analysis.[29] Moreover, once the sample selection model following Wooldridge's method is estimated, I find that the inverse mills ratios are neither individually nor jointly significant in this case.[30] Therefore, I estimate a standard fixed effects model considering the enrolled children.

Table 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

[28]As described in footnote 26 in page 22, this variable is an index of school quality. The information is obtained from the DISE data for the years 2007-08 and 2010-11.

[29]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.

[30]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 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.

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.[31]

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 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

[31] 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.

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.

5.3.1 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* 6. The findings are almost identical with the last set of results shown in *Table* 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.

5.3.2 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}.[32] 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* 7 assuming $\delta = 1$.

Each row of *Table* 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.[33] The final column identifies an interval which is bounded

[32]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.

[33]Note that in Section 5.3.1, village level time-variant unobservables are controlled by including village-by-year fixed effects. Column (2) of *Table* 7 refers to the specification which further controls for

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 school choice.

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 differ-

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.

ence 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.

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 have become prevalent in the choice of more expensive and perceptibly better quality schools. The apparent progress in the overall enrollment rates of girls is shrouded by a more nuanced gender story that needs careful consideration by the policy makers to achieve the goal of equality in educational opportunity.

References

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

[2] 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.

[3] 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.

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

[5] 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.

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

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

[8] Clark, Shelley (2000), "Son Preference and Sex Composition of Children: Evidence from India", *Demography*, 37(1): 95-108.

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

[10] 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.

[11] 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.

[12] Duflo, Esther (2012), "Women Empowerment and Economic Development", *Journal of Economic Literature*, 50(4): 1051-1079.

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

[14] 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.

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

[16] 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.

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

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

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

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

[21] 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.

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

[23] 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.

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

[25] 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.

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

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

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

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

[30] 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.

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

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

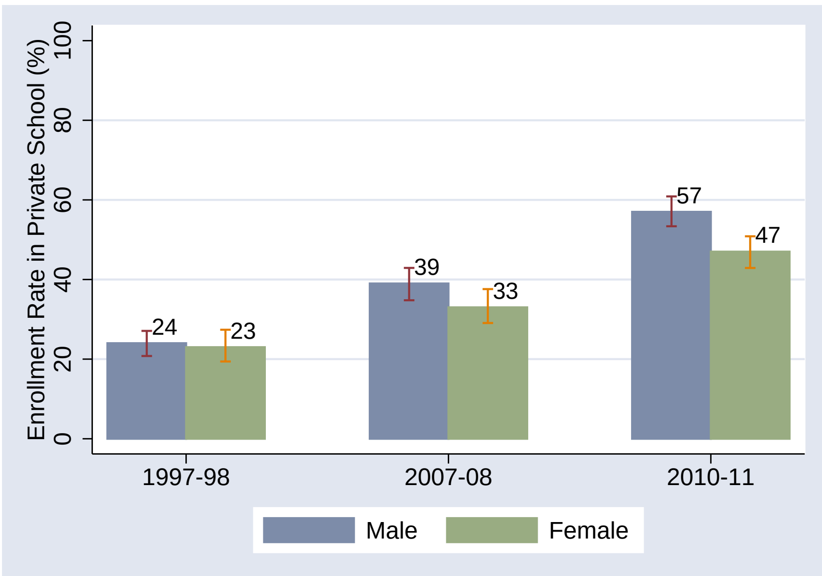
[37] Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts.

[38] 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.

Figure 1: Enrollment rates by gender (age 6-19 years)

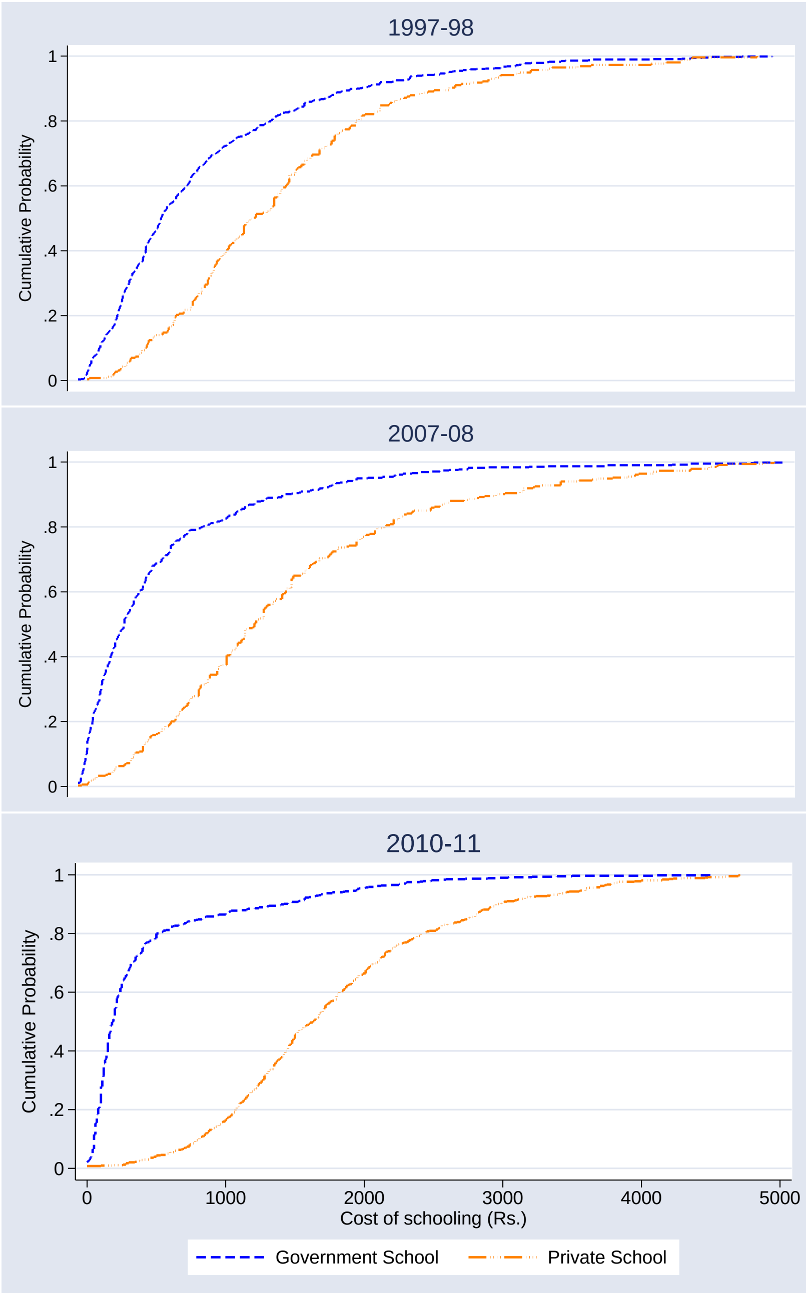
Source: SLC data.

Figure 2: Private school enrollment rates for enrolled children by gender (age 6-19 years)



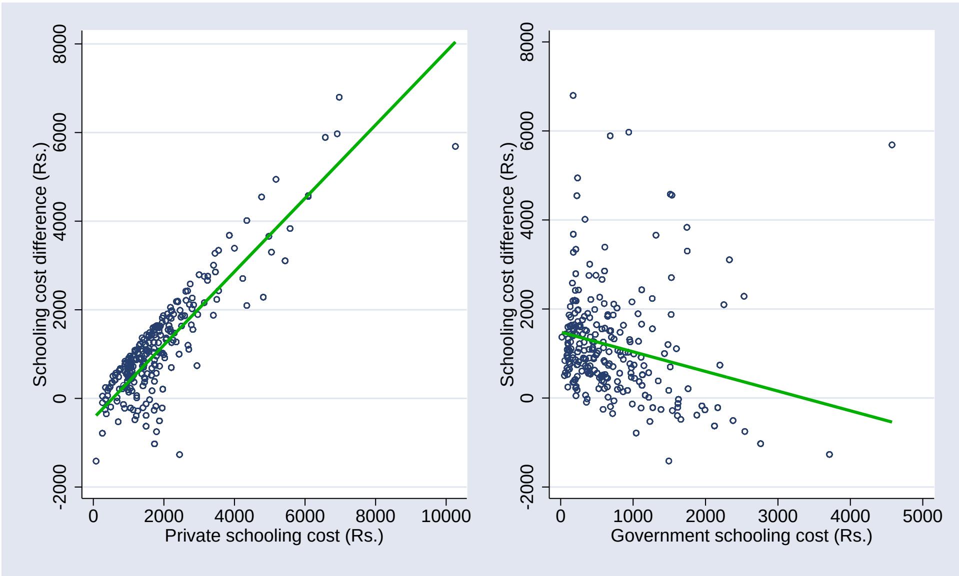
Source: SLC data.

Figure 3: Cumulative distribution of direct expenditure in government and private schools



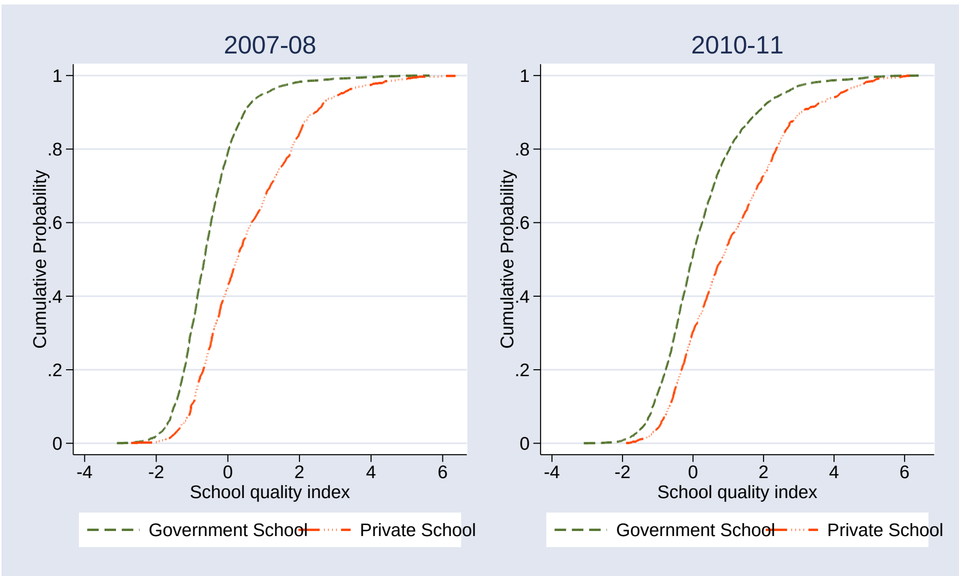
Source: SLC data.

Figure 4: Scatterplot with linear fit to show relation between schooling cost difference and private/government schooling cost



Source: SLC data.

Figure 5: Cumulative distribution of school quality in government and private schools



Source: DISE data (considering the SLC blocks).

Table 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 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 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: 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 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 6: Robustness: Controlling for village level unobservables

Variables	Dependent Variable: Private school choice (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 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* 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 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 A2: 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 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%.