Understanding the *"sorting hat"*: the role of family and caste network in school choice decision*

Sukanta Bhattacharya†
Aparajita Dasgupta‡
Kumarjit Mandal §
and
Anirban Mukherjee ¶

Abstract

Sorting of students over school type by their social standing creates the problem of social segregation and educational inequality. This paper, both theoretically and empirically, identifies factors which are responsible for sorting of students by socio-economic groups across different types of schools. We elaborate on the sorting mechanism by analyzing schools choice decisions where parents decide on the following: whether to send their children to private/public school and conditional on that whether to send them to good or bad school. We find that household characteristics such as the number of siblings, caste identity affect these two choices differently. For example we find that general caste students are more likely to attend private school but within the private category they are more likely to end up in bad schools. We develop our theory using a two period household level optimization framework and validate the prediction using a school survey data (Young Lives Survey).

*This is a preliminary draft

†University of Calcutta. Email:bsukanta@gmail.com
‡Population Council, New Delhi. Email:aparajita.dasgupta@email.ucr.edu
§University of Calcutta. Email:kumarjitm@hotmail.com
¶University of Calcutta. Email:anirban1976@gmail.com

1 Introduction

The issue of school choice and its subsequent impact on the future generation's earning opportunity assumes a central role in the economics of education literature. Conventional wisdom suggests that poor parents cannot afford good quality education for their children leading to restricted economic growth of income for the next generation. This mechanism which is often termed as poverty trap does not allow poor household to break free from the low level equilibrium (Galor and Zeira, 1993). In the current paper however we try to go beyond this position and see whether non-pecuniary factors such as caste identity, household size, sibling composition, gender composition and gender of the children play important roles in the process of school choice. This work allows us to understand whether between two households of same wealth position but different caste status or family structure one has a higher chance of breaking free from the poverty trap than the other.

Unlike Harry Potter's Hogwarts School of Witchcraft and Wizardry where students were sorted into different school houses using a *sorting hat*, school choice decisions by regular parents follow a rational, optimizing pattern by taking return to education in consideration. In this paper we analyze the school choice decision by parents using a human capital investment framework. We develop a theoretical model and validate the results using data from the Indian state of Andhra Pradesh. In a country like India, the parents have the choice of sending their kids to government schools or private schools. Since the former provides free (or very low cost) access to schools any rational decision making based only on the cost aspect would suggest that all parents would go for government schools – private schools do not stand a chance to survive in a competitive environment. Nevertheless, the private schools, in spite of their high cost, exist in the market because they provide (or believed to provide) better quality of education than

Young Lives is an international study of childhood poverty, following the lives of 12,000 children in 4 countries (Ethiopia, India, Peru and Vietnam) over 15 years. www.younglives.org.uk Young Lives is funded from 2001 to 2017 by UK aid from the Department for International Development (DFID), co-funded by the Netherlands Ministry of Foreign Affairs from 2010 to 2014, and by Irish Aid from 2014 to 2015.The views expressed are those of the author(s). They are not necessarily those of, or endorsed by, Young Lives, the University of Oxford, DFID or other funders

their government counterpart. In the Indian context there are many studies which suggest that private schools on an average are better than the government schools. In the literature (and also in popular belief) private school system is synonymous with good quality education and investing in private education is worthwhile if the parents can afford it.

In this backdrop the contribution of our paper are two folds. We reckon that the public/private choice and the school quality choice are two related but different choices. In our theoretical model we show that they follow distinctly different rationale. This position is supported by our empirical results where the factors which affect the public/private choice in one way, affect the school quality choice in another way. Our second contribution is to theoretically place social factors such as ethnic identity and household structure in school choice decisions and empirically validate the theoretical predictions.

The existing literature on school choice touches upon different fields of economics. There has been an emphasis on schooling in the growth literature as it determines the future educational attainment and the gains in productivity for the society (Lucas Jr, 1993). However, schooling in its bare minimum form is not necessarily translated into higher growth. The quality of schooling matters (Pritchett, 2001). Studies have found that the quality of schooling often is responsible for the learning gaps in the children leading to differences in the future earnings of the pupil (Glewwe et al., 2014). Given that future return to education is too noisy to predict, people often choose schools based on the quality parameters which can be thought to predict future earnings. However, the perception about school quality in absence of full information often depends on signals whose interpretations become subjective. The most obvious signal of school quality is whether the school is run by a public or a private body. There has been a long standing debate about the relative difference in schooling quality between private and public where the school quality is being measured by pupil-teacher ratio, class size, teachers' salaries and experience (James and Woodhead, 2014). The authors find that students enrolled in private schools learn better than the students in public schools and this learning gap is often termed as private school premium (Checchi and Jappelli, 2003; Singh, 2013). The choice of school crucially depends on the perceived school-quality which is affected by both subjective (e.g.

belief about work culture within the school) and objective indicators (e.g. the school resources/infrastructure in each geographical area). It is possible to link the school choice to aggregate school resources as well as individual decision-making by appropriate combinations of both these factors (Checchi and Jappelli, 2003). It has also been observed that school choices are often guided by the soft qualities such as uniforms, disciplines rather than the hard qualities such as teaches' qualifications and in terms of the soft qualities private schools usually fare better than their public counterparts (Azim Premzi Foundation Report, 2013).

Besides wealth, sibling composition and household size becomes an important factor in school choice decision. Butcher and Case (1994) find that in the United States between 1920-1965 women's educational choices have been affected by sex composition of her siblings while that of men have not. They found that women raised only with brothers received more education than women raised with sisters. Black et al. (2005) found that in Norway for an extended period birth order has significant negative effect on children's education.

Given the importance of the quality of school in future outcome, it is also pertinent to ask who ends up in bad schools and who in good schools. If the kids from the upper social classes end up in good schools while kids from lower socio-economic background end up in bad schools, the trend in social mobility will take a regressive turn. This phenomenon, known as sorting, is well researched in the context of developed countries (Card and Rothstein, 2007; Urquiola, 2005; Burgess et al., 2004) . However, unlike India where parents can choose any school for their kids, the parents in developed countries are generally constricted to send their kids to the public school in their residence school districts. So their choices of school is largely constrained by their choices of residence. Hence, in developed countries poor districts mostly cater to poor students while districts with high real estate price mostly take rich students. This makes the Indian case, where neighborhood choice and school choice are two distinct decisions, even more interesting to study.

Sorting can also take place along the gender line. Long and Conger (2013) find strong gender sorting across U.S. schools within sectors and types with higher gender sorting in counties that have higher

shares of enrollment in private and non-regular public schools. This sorting occurs even after accounting for parental preferences for school attributes for their sons and daughters. This issue is also related to the issue of differential investment in girl's education. Azam and Kingdon (2013), using data from India Human Development Survey(2005) find a difference in education expenditure by ages 10-14, and a difference in enrollment by ages 15 - 19. One of the reasons behind this is that boys are more likely to be sent to private schools, which is generally considered to be of higher quality than free public schools.Evidence from developing countries finds differential investments in childhood health, gender differences in household inputs such as educational investments, parental time, and food.

Jensen (2012) finds that increasing awareness of employment opportunities for women increased enrollment and body mass index of younger school-aged girls (ages 5-15) pointing that parents invest in girls more when the potential of their future returns are improved. In a similar line of research Munshi and Rosenzweig (2006) find that increasing returns to English combined with caste-based networks are driving increased school enrollment and investment for girls in cities in India. The idea is that while boys can get blue collar jobs through caste network, girls may not as those jobs are not deemed suitable for them. On the other hand, girls can work in white collar jobs which require the knowledge of English language. Hence, when there is increased returns to English language skills in India there is increase in girls educational investments in English medium schools.

The decision about the choice of school is typically taken by the parents in a household with the objective of enhancing future earnings prospect of the children. A two-period utility function of the household with appropriate resource constraint would serve as the benchmark model to capture this dynamism. The constraint set should incorporate the control variables signifying two choices of the household in search of school quality. They have to decide whether to choose a public or a private school for their kids' education. This is a binary choice problem. Now within the spheres of both private and public, there might be various schools of different qualities depending on the soft and hard qualities. The constraint set needs to incorporate the control variable with appropriate index for this quality difference between the schools. A simple analytical solution might be obtained

for school quality choice for the public and private schools separately. The main findings from the analytical model show amongst others that the choice of good quality school is dependent on the wealth of the household. The implicitly suggests the presence of sorting.

The present study has made an attempt to verify the results from the analytical model in the light of the Young Lives (YLS) data set . The indexation for good quality school has been carried out by averaging the school infrastructure , teacher quality and school facilities . To find the determinants of choice for the good quality school a set of probit regressions has been carried out. The results from the econometric exercise broadly corroborate the results from the analytical exercise. The suggestion that the household wealth plays an important role in determining the choice of quality school has been validated by the data. The presumption of private school premium might also find support in the data. The remaining part of the study is organized as follows: the next section describes the analytical model, section 3 discusses the YLS data set, section 4 focuses on the econometric results and section 5 concludes the study.

2 Model

We set up a two-period model of household decision making about investment in education vis-a-vis physical capital. The preference of a household is characterized by the utility function

$$U(c_1, c_2) = c_1^{\sigma} + \delta c_2^{\sigma} \tag{1}$$

where c_t denotes household consumption in period t, $t = 1, 2$ and $\sigma \in (0, 1)$ denotes the elasticity of intertemporal substitution. At the beginning of period 1, a decision-making household is endowed with wealth w. The wealth may be utilized either by spending on current consumption, or by investing on a risk-free asset that fetches a gross return of R per unit next period. Additionally, the household may also decide to invest in child's education.

There are two types of schools in the society where children may be sent. The government schools are free and therefore does not affect the amount of wealth available for current consumption or investment. There are also private schools which charge fee for admitting children. The type of school a household chooses for its child is denoted by i

where $i \in \{0,1\}$. In our model, $i = 0$ stands for government school and $i = 1$ stands for private school. The cost of sending a child to a type i school is $f.i$ with $f > 0$. So if a family chooses to send its child to a type i school and decides to invest s in the risk-free asset, then the wealth available for current consumption is $w - f.i - s$.

However, there is another aspect of a household's school choice. Both private and government schools exhibit variation in quality. We denote the quality of a school by an index $q \in \left[\underline{q}, \overline{q}\right]$. If a child is sent to a school with higher quality, it improves her learning and as a result increases her earning potential in future. However, a good quality school requires more effort from the child and other family members and thus reduces its ability to transform current wealth into current consumption. If a family chooses type i school of quality q for its child, then its current consumption can be written as

$$c_1 = \beta(q)\left[w - f.i - s\right] \tag{2}$$

where $\beta(.)$ denotes the factor that converts current wealth into current consumption. We assume that $\beta' < 0$. Specifically for the sake of simplicity we choose

$$\beta(q) = \frac{A}{q}. \tag{3}$$

On the other hand, the same child's expected future return from education is

$$\gamma_i(q,\theta).Y$$

We assume that a child's future earning is positively linked with the quality of school q and a socio-cultural parameter $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ which captures the capabilty of transforming the learning into income earning potential, i.e. $\frac{\delta\gamma_i}{\delta q} > 0$ and $\frac{\delta\gamma_i}{\delta\theta} > 0$. Moreover, we also assume that for every (q,θ), $\gamma_1(q,\theta) > \gamma_0(q,\theta)$. This assumption captures the almost near-universal belief that private schooling gives higher return in the long-run. Once again, we choose a simple multiplicative form for the coefficient $\gamma_i(q,\theta)$ in

$$\gamma_i(q,\theta) = \alpha_i q\theta \tag{4}$$

with $\alpha_1 > \alpha_0 > 0$. Finally, we can interpret Y as the minimum future wage for a child. A child with characteristic $\underline{\theta}$ going to a government

Notice that this assumption may not be true ex-post. However, in this model we try to examine the household's school choice decision for the child and in that decision making process the household's belief about return from different types of school is what matters.

school of quality $\underline{q}$ earns $\gamma_0\left(\underline{q},\underline{\theta}\right)Y$ in her working life. Without loss of generality, we choose parameter values such that $\gamma_0\left(\underline{q},\underline{\theta}\right)=1$.

We can now formally state the household's choice problem. The household chooses both the type of school ($i\in\{0,1\}$) as well as the quality of school (q) it will send its child to at date 1. This reflects its investment in human capital. The household also decides on its investment in physical capital (s). These investments in turn determine the household's consumption at date 2 which can be expressed as

$$c_2 = Rs + \gamma_i\left(q,\theta\right).Y \tag{5}$$

The intertemporal budget constraint can thus be written as

$$\frac{c_1}{\beta\left(q\right)} + \frac{c_2}{R} = w - f.i + \frac{\gamma_i\left(q,\theta\right).Y}{R} \tag{6}$$

The household's choice problem is maximization of (1) subject to the constraint (6). Among household's choice variables, school type and school quality affect the household's budget constraint, but do not directly affect the utility. However, while school type affects the present value of life-time consumables of any household, school quality affects the same present value as well as the relative price between present and future consumption. In other words, school type only induces a wealth effect on consumption choice while school quality generates a wealth effect as well as a price effect.

In the household's choice problem, c_1, c_2 and q are continuous variables while i is a discrete variable. We follow a two step procedure to solve the household's problem. First for a given $i\in\{0,1\}$, we determine the household's optimal choices of c_1, c_2 and q. In the second step, we determine a household's optimal choice of i.

Given any $i\in\{0,1\}$, the optimality conditions for the household's choice problem are

$$\left(\frac{c_1}{c_2}\right)^{\sigma-1} = \frac{\delta R}{\beta\left(q\right)} \tag{7}$$

and

$$-\frac{c_1}{\left[\beta\left(q\right)\right]^2}\beta'\left(q\right) = \frac{\delta}{\delta q}\left[\gamma_i\left(q,\theta\right)\right].\frac{Y}{R} \tag{8}$$

along with (6). Since the utility function is strictly quasi-concave, the choice problem has a unique solution. Using the functions from (3) and (4), we can solve for the optimal choices of the households as

$$c_1^*\left(i\right) = \frac{A\alpha_i\theta Y}{R}$$

$$c_2^*(i) = R(w - f.i)$$

and

$$q^*(i) = \frac{R^{1-2\sigma}(w - f.i)^{1-\sigma} A^\sigma}{\delta\theta^{1-\sigma} Y^{1-\sigma}}$$

In our first proposition, we exhibit how for the children going to a particular type school, the choice of school quality is affected by changes in different parameters of the model.

Proposition 2.1 *For children going to a particular type of school (either government or private), the quality of school is positively associated with wealth level (w) and the household's current productivity (A) and negatively associated with the child's capability of transforming learning into income (θ) as well as the minimum future wage (Y).*

The results are along expected line. First consider an increase in w. If everything else remains same, this leads to an increase in c_1. However any increase in c_1, reduces the marginal rate of substitution between present and future consumption and the household's willingness to pay for future consumption in terms of present consumption rises. This leads to an increase in q as school quality increases future consumption at the same time making current consumption costlier. The other results in proposition 1 can be intuitively explained in similar manner.

We are now in a position to examine the choice of school type by a household. For a household that chooses school type $i \in \{0, 1\}$, the indirect utility funcrion can be written as

$$\begin{aligned} V_i(w, f, A, \alpha_i, \theta, \delta) &= [c_1^*(i)]^\sigma + \delta [c_2^*(i)]^\sigma \\ &= \left[\frac{A\alpha_i\theta Y}{R}\right]^\sigma + \delta [R(w - f.i)]^\sigma \end{aligned}$$

Hence, the incremental utility from choosing a private school for its child is given by

$$\begin{aligned} \Delta V &= V_1 - V_0 \\ &= \left[\frac{A\alpha_1\theta Y}{R}\right]^\sigma + \delta [R(w - f)]^\sigma - \left[\frac{A\alpha_0\theta Y}{R}\right]^\sigma - \delta [Rw]^\sigma \end{aligned}$$

Since $\sigma < 1$, ΔV is strictly increasing in w. Thus the benefit for going to a private school is higher for wealthier households. Notice that if α_1 is so high relative to α_0 such that

$$\left[\frac{A\theta Y}{R}\right]^\sigma (\alpha_1^\sigma - \alpha_0^\sigma) \geq \delta [Rf]^\sigma$$

then every household sends its child to private school and the government schools have no takers. However, if

$$\left[\frac{A\theta Y}{R}\right]^{\sigma}(\alpha_1^{\sigma}-\alpha_0^{\sigma})<\delta\left[Rf\right]^{\sigma} \tag{9}$$

holds, then for any $w \leq f$, $\Delta V < 0$. But as $w \to \infty$, $\Delta V > 0$ follows from $\alpha_1 > \alpha_0$. Since ΔV is continuous and strictly increasing in w, there exists a critical $w_c \in (f, \infty)$ such that $\Delta V \geq 0$ if and only if $w \geq w_c$. This leads to our second proposition.

Proposition 2.2 *Suppose 9 holds. Then there exists a critical wealth level $w_c \in (f, \infty)$ such that a household sends its child to private school if and only if its wealth level $w \geq w_c$. Moreover, w_c falls as θ or A increases, but rises as f increases.*

The households with higher wealth are more likely to send their children to private schools. This is what one expects given diminishing marginal utility of present consumption and higher expected future return from private schools. However, an increase in θ may induce a household to send its child to a private school who would not have done so otherwise. θ is a socio-cultural parameter in our model which captures the household's capability of transforming education into income earning potential. We can think of θ as a parameter representing the social standing of the household which eventually matters when the child enters the job market. For example, we expect that general castes would have a higher θ than the backward castes. People living in urban areas are expected to have higher θ than people residing in rural areas. This is because educated urban youth face a better chance to match the skills learnt in school with the job they ultimately land than their rural counterparts. We thus expect that children from general categories or children in urban households are more likely to attend private schools even when we control for family wealth or income.

3 Data

The data used in this study comes from the Young Lives study which was collected between 2002 and 2011 in the state of Andhra Pradesh. The sites were selected from three different agro-climatic areas and

CoastalAndhra, Rayalseema and Telangana (Young Lives 2007)

had a pro-poor bias with districts and sites being ranked according to a number of development indicators(Kumra 2008).The administrative sub-districts (mandals) are the primary sampling units in our sample. We use data of the younger cohort of children born between January 2001 and June 2002. We make use of the rich demographic array of indicators from the household survey for example parental/caregiver education, wealth index of the household, caste, religion, household head's gender, number of siblings, sibling composition, child anthropometry, a host of school level outcomes (cognitive outcomes and test scores in mathematics, Telugu and English,type site(rural/urban),region/community type, whether member of any social group, number of household members giving financial support to the child, the number of school going kids present in the household, birth order of the child, whether household suffered from any major bad event in the last four years etc. Additionally we use the separate schooling data collected through visits to the schools of a randomly selected sub-sample of the Younger Cohort in 2011.Attrition rates in the data is very low – 1930 children(96 per cent) in the Younger Cohort sample could be followed in 2009.Overall attrition by the third round was 2.2% (with attrition rate of 2.3 per cent for the younger cohort) over the eight-year period.In 2011, the Young Lives study randomly sampled 247 schools which were being attended bychildren in the Younger Cohort.The sampling frame consisted of all the Younger Cohort(YC) children who were still enrolled in school in Round 3 (2009-10) and were going to school within Andhra Pradesh. The sample included 952 children across 247 schools.The school-level survey was conducted between December 2010 and March 2011, i.e. in the school year immediately following the third wave of household-level data collection (Singh, 2013). The survey captured detailed school-level differences in infrastructure and funding, teacher qualifications and characteristics, classroom characteristics, teaching procedures and children's subjective experiences of schooling. It administered questionnaires to all school principals, teachers and detailed information on the mathematics teachers of the sample children from the younger cohort.

Andhra Pradesh is divided into 23 administrative districts that are further subdivided into mandals. Generally, there are between 20 and 40 villages in a mandal. In total, there are 1,125 mandals and 27,000 villages in Andhra Pradesh(Kumra 2008)

4 Results

The primary objective of this paper is to analyze how family characteristics and caste network influence the school choice decision made by the parents for their children. In our theory we analyze the choice as a decision to invest in human capital with the objective of maximizing family income. Sending a child to school requires spending of family resource and like any other investment decision the opportunity cost of spending the resource for education plays a critical role. In this section, we empirically examine how different family characteristics that capture the investment mechanism becomes important. We know from our theory and existing literature that factors such as family wealth, number of children, sibling composition, parental education, membership in caste network need to be examined for their roles in school choice decision.

In the last section we described the data. Let us now describe the formulation of index. We classify good school based broadly on two indices: teacher quality and school infrastructure. For construction of teacher quality index we use information on the highest level of education completed of the teacher and highest level of teacher training. The infrastructure index uses the following information at the facility level: number of fully covered separate rooms for teaching, having a secure compound wall/fence, availability of electric connection, alternative sources of power, availability of drinking water in school premises, availability of drinking water on the day of the survey, whether classes sharing the same classroom, having computer, internet facility, first aid, play material, sports kit, music kit, art material, school bell, television and functioning toilets. In a later stage we also breakdown the infrastructure index in *infrastructure* and *facility* indices. In the *infrastructure* group we include information mainly on school building and toilets while under *facility* we include computer, internet facility, first aid, play material, sports kit, music kit, art material, school bell, television etc.

For each indicators we normalize their value [0 1] and use the composite index for defining a *good school*(if the value is greater than the mean score). For making the indicators comparable we take the actual value of the above mentioned categories and normalize them by the

following formula for a variable x

$$I_x = \frac{x - \underline{x}}{\overline{x} - \underline{x}} \tag{10}$$

where $\underline{x}$ denotes the minimum value of x and $\overline{x}$ denotes the maximum.

In this section we report the results from running the following Probit model

$$Pr(z = 1) = \Phi(X\beta) \tag{11}$$

where z takes the value 1 when a student attends a good school (government school for the first regression) and 0 otherwise. X represents variables of interest such as family size, family asset, parental education, network etc and other relevant controls. We start by estimating the equation for both the rural and urban areas. Eventually we test the equation separately for rural and urban areas and include different controls for testing robustness.

In our theory, we have looked at two relevant school choices: public vs private and good school vs bad school. Conventional belief equates private school with an indicator of good school – a belief which is confirmed by our data on the basis of average values. Nevertheless, there are several dimensions - other than quality- believed to be embedded in a private school. Hence, the choice of school organization (private or public)needs to be analyzed separately. This is why our theory centers around two distinct decisions: private vs public and good vs bad. In the empirical section we treat these two separately as well. Also, we empirically analyze the school choice for rural and urban setting separately.

We start by trying to find what type of households send their kids to private schools. We find that general caste households are less likely to send their kids to government schools. Similarly, families with more educated primary caregiver for their children (this will be parents or close relatives) are more likely to send their children to private schools. Similarly, male children are more likely to end up in private schools while kids with siblings are less likely to go to private schools.

In our study a school can be characterized along three dimensions: quality (score) location (rural/urban) and organization (government

or private). In table 1 we present the summary statistics. This is a student level data where we collate their family characteristics, location, caste and school score. Note that, data is not available for all variables for all students. In table 1, we present the summary statistics for the whole sample.

We start by looking at the factors influencing the first decision variable: private or public school. We find that wealthy families are more likely to send their kids to private schools. General caste households are also likely to send their kids to private schools. We also find that male children, children without siblings and children with educated primary caregivers (usually parents) are more likely to attend private schools. However, we maintain the point that even if school organization (government or private) signals quality, there is high degree of variation of quality within each type. This justifies our separate analysis of decisions regarding public/private and good school/bad school choice. In table 2 we represent the distribution between good and bad school across government and private categories. We find that among the bad schools 53% are private while 47% are government when we consider rural and urban area together. However, the balance is more tilted towards the private schools in the urban area than in the rural. Among the urban good schools 84% are private while 16% are public. On the other hand, among the rural good schools, 56% is private and 44% are government. We also emphasize the multi-dimensionality of school quality.

In tables 7 and 8 we present the mean qualities of government and private schools which are measured along three dimensions – teacher's quality, school facility and schools infrastructure. In all these categories an average private school fares better than its government counterpart. The gap between private and government is very high when it comes to facility or infrastructure. For teacher's quality, the average private school is marginally better than the government one.

In the next step we see how different family characteristics affects the probability of going to good school for different settings such as rural and urban. We further analyze the decision in the following subcategories: rural government, rural private, urban government and urban private.

For the rural-urban combined sample (Table 9 and 10) we run the Probit with siblings entering the regressions differently. In table 9 we include sibling dummy as control while in table 10 we include the number of siblings. For robustness, we also do a regression with the sibling composition as the control (not reported here). But that does not change the result. We find that wealthy families are more likely to send their kids to good school. Similarly, education of the primary caregiver is also positively contributing to the probability of going to good school. Sibling dummy on the other hand is negatively related. Remember, these factors had a similar effect on the probability of going to private schools.

In table 10 we report the result with the number of siblings as control and find that the number of siblings has a positive impact on the probability. The only other variable that is significant in this regression is rural location which has a negative impact. Note that the effect of the number of siblings is different from the regression with the probability of government school as the dependent variable. In that regression kids with siblings are more likely to end up in a government school (table 2). But an average government school is worse than the an average private school. How do we reconcile these two results? Also note that if we use the number of siblings as the control we get a positive result. This means that having a sibling is not enough. The number of sibling is critical. Using our theoretical model we can reason this in the following way. Sending a kid to good school is costly – both in terms of money and the time it requires to study. If a child puts more effort for study (as is required in a good school) there will be less effort for home production. However, with more siblings around to share the burden the parents can afford to send children to good schools.

Then we look at the choice for sending to a good school among the students attending government schools (table 10). We find that wealth is positively affecting the probability of going to good schools. We also find that urban students are more likely to attend good schools. Then we look in the sub groups within the urban and rural categories. For rural government schools (Table 11) only wealth has a positive significant effect.

In table 12 and 13, we look at the students of rural private and urban private schools. We find that among the students attending private schools general caste students are less likely to attend good schools. Also, consistent with the earlier results, in rural areas kids of more educated parents are likely to attend good schools. We do not report the results with sibling composition as they do not change any result. For urban settings, (table 13) we find that general caste is still negatively significant. But education of caregivers do not matter anymore. But the number of school going children in the family negatively affects the probability of going to a good school.

Next, instead of defining good schools by composite index we define good schools separately based on good facilities (computer etc), good infrastructure (rooms, toilet etc) and teacher's qualification. First we report the result when good school is defined on the basis of good facility. For rural private schools, kids from male headed household go to good schools. No other factors matter.For rural government schools, wealth positively affects the probability of going to good schools while caregiver's education and the number of school going kids negatively affects it. For urban private schools none of the factors matter.

Next, we define schools based on infrastructure. First we look at the private urban schools. We find that household size and caregiver's education have negative impact on the probability of going to good schools. General caste families on the other hand are less likely to send their kids to good schools. For the rural good schools – "good" being defined on the basis of infrastructure – group membership and caregiver's education have positive impact on the probability. Students who get support from household are less likely to attend good schools. For rural, government schools none of the factors matter.

In the last part, we define good schools based on teacher's qualification. We find that urban private general caste people are less likely to send their kids to good school. For rural private, only group membership has a negative impact while for rural government nothing matters.

5 Conclusion

This paper analyzes the school choice decision using a theoretical model and tests the testable implications using data. We analyze the decision making using an inter-temporal optimization model that sees children's education as investment in human capital. The trade-off from the decision making comes from two channels: the resource saved from not sending a child to a costly school can be invested in physical capital and the child's saved time can be used for household work. Unlike the standard literature, we use two stage decision making to distinguish between the decisions of sending a kid to a private schools and sending her to a good/bad school once the public/private decision has been taken. We find that the way the social and cultural variables affect the public/private school decision are not the same they affect good/bad school choice. Our empirical results support these theoretical results. This allows us to conclude that for given wealth families from different socio-cultural background send their children to different quality schools which have profound effect on their future earning in the micro level and social mobility in the macro level. By analyzing the school choice mechanism we create a space in the policy regime to discuss appropriate policies.

References

Azam, M. and Kingdon, G. G. (2013). Are girls the fairer sex in india? revisiting intra-household allocation of education expenditure. *World Development*, 42:143–164.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, pages 669–700.

Burgess, S., McConnell, B., Propper, C., and Wilson, D. (2004). Sorting and choice in english secondary schools. *Centre for Market and Public Organisation Working Paper*, 4:111.

Butcher, K. F. and Case, A. (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, 109(3):pp. 531–563.

Card, D. and Rothstein, J. (2007). Racial segregation and the black–white test score gap. *Journal of Public Economics*, 91(11):2158–2184.

Checchi, D. and Jappelli, T. (2003). School choice and quality. Technical report, IZA Discussion paper series.

Galor, O. and Zeira, J. (1993). Income distribution and macroeconomics. *The review of economic studies*, 60(1):35–52.

Glewwe, P., Krutikova, S., and Rolleston, C. (2014). Do schools reinforce or reduce learning gaps between advantaged and disadvantaged students? evidence from vietnam and peru.

James, Z. and Woodhead, M. (2014). Choosing and changing schools in indias private and government sectors: Young lives evidence from andhra pradesh. *Oxford Review of Education*, 40(1):73–90.

Jensen, R. (2012). Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India. *The Quarterly Journal of Economics*, 127(2):753–792.

Long, M. C. and Conger, D. (2013). Gender sorting across k–12 schools in the united states. *American Journal of Education*, 119(3):349–372.

Lucas Jr, R. E. (1993). Making a miracle. *Econometrica: Journal of the Econometric Society*, pages 251–272.

Munshi, K. and Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *The American Economic Review*, pages 1225–1252.

Pritchett, L. (2001). Where has all the education gone? *The World Bank Economic Review*, 15(3):367–391.

Singh, A. (2013). Size and sources of the private school premium in test scores in india. *Young Lives Working Paper*, 98.

Urquiola, M. (2005). Does school choice lead to sorting? evidence from tiebout variation. *American Economic Review*, pages 1310–1326.

Table 1: **Summary Statistics**

Variable	Obs	Mean	Std. Dev.	Min	Max
goodschool	953	0.6977964	0.4594541	0	1
hhsize	953	5.555089	2.429977	2	22
wi1	953	0.3892927	0.190981	0.0111111	0.8796296
hindu	953	0.8982162	0.3025226	0	1
gen_caste	953	0.1752361	0.3803685	0	1
shigh	953	2.725079	4.052954	0	15
region	953	22.02833	0.8226284	21	23
male	953	0.5362015	0.4989496	0	1
bwght	356	2740.067	544.2736	1000	4500
brothers	951	0.3322818	0.6448165	0	6
sisters	951	0.4374343	0.7691604	0	5
schkid	944	0.7245763	1.002854	0	6
siblings	953	1.351522	0.4776961	1	2
genstruc	953	1.786988	0.7543701	1	3
grpmem	951	0.3280757	0.4786388	0	2
badevent	953	0.5152151	0.5000309	0	1
rural	953	0.8100735	0.3924488	0	1

Table 2: Probability of going to government school

Probit regression		Number of obs	949
		LR chi2(13)	434.14
		Prob > chi2	0
Log likelihood = -414.75787		Pseudo R2	0.3436
govt_school	Coef.	Std. Err.	z	P>z
hhsize	0.0369099	0.0227366	1.62	0.105
wi1	**-3.246064*** **	0.3418082	-9.5	0
hindu	-0.2452353	0.1779389	-1.38	0.168
gen_caste	**-0.543707*** **	0.1402641	-3.88	0
shigh	**-0.1021552*** **	0.0150697	-6.78	0
region	**-0.4654067*** **	0.0712671	-6.53	0
male	**-0.5105428*** **	0.1122608	-4.55	0
have_siblings	**0.2989668** **	0.1173525	2.55	0.011
brothers	0.0921504	0.0849726	1.08	0.278
sisters	0.0256743	0.0788923	0.33	0.745
genstruc	-0.0117342	0.0757034	-0.16	0.877
grpmem	0.1134876	0.1059768	1.07	0.284
badevent	0.3030584	0.1121578	2.7	0.007
_cons	12.14242	1.594075	7.62	0

Table 3: Rural and urban combined

	Private	Govt	Total
Bad School	12 (4.17)	276 (95.83)	288 (100)
Good School	353 (53.08)	312 (46.92)	665 (100)
Total	365 (38.3)	588 (61.7)	953 (100)

Table 4: Urban Area

	Private	Govt:	Total
Bad School	105(88.24)	14(11.76)	119 (100)
Good School	52 (83.87)	10(16.13)	62(100)
Total	157 (86.74)	24(13.26)	181(100)

Table 5: Rural Area

	Private	Govt:	Total
Bad School	130(20.57)	502(79.43)	632 (100)
Good School	78(55.71)	62(44.29)	140(100)
Total	208 (26.94)	564(73.06)	772(100)

Table 6: Private School quality According to Different parameters

Variable	Obs	Mean	Std. Dev.	Min	Max
good_facility	365	0.6191781	0.4862555	0	1
school_infrastructure	365	0.6671233	0.2250182	0.25	1
good_infrastructure	365	0.8931507	0.3093456	0	1
Teacher quality	365	0.660274	0.4742663	0	1

Table 7: Government School quality According to Different parameters

Variable	Obs	Mean	Std. Dev.	Min	Max
good_facility	588	0.1462585	0.353666	0	1
school_infrastructure	588	0.2614796	0.1989539	0	1
good_infr	588	0.2397959	0.4273223	0	1
Teacher quality	588	0.4132653	0.4928389	0	1

Table 8: Whole Sample: Urban and Rural

Probit regression	Number of obs	=	940	
	LR chi2(15)	=	129.96	
	Prob > chi2	=	0	
Log likelihood = -508.38441	Pseudo R2	=	0.1133	

goodschool	Coef.	Std. Err.	z	P>z
hhsize	0.022576	0.0248531	0.91	0.364
wi1	**1.461536*** **	0.3462972	4.22	0
hindu	0.2492496	0.1637665	1.52	0.128
gen_caste	-0.0061579	0.1470229	-0.04	0.967
shigh	**0.0500356*** **	0.0157075	3.19	0.001
region	0.0996606	0.061838	1.61	0.107
male	0.1415585	0.0991353	1.43	0.153
brothers	-0.0912198	0.0841422	-1.08	0.278
sisters	0.1018612	0.0803413	1.27	0.205
schkid	-0.0437354	0.0710871	-0.62	0.538
siblings	**0.2354613** **	0.1116081	2.11	0.035
genstruc	-0.0395512	0.0688066	-0.57	0.565
grpmem	0.0136779	0.0973168	0.14	0.888
badevent	-0.0441211	0.1047342	-0.42	0.674
rural	-0.4587301	**0.1974186**	**-2.32**	**0.02**
_cons	-2.54015	1.414264	-1.8	0.072

Table 9: Urban and Rural: Sibling dummy as control

Probit regression			Number of obs	909
			LR chi2(17)	125.83
			Prob > chi2	0
Log likelihood = -493.45726			Pseudo R2	0.1131

goodschool	Coef.	Std. Err.	z	P>z
hhsize	-0.0055243	0.0288979	-0.19	0.848
wi1	**1.485179*** **	0.3530537	4.21	0
hindu	0.2333858	0.1667167	1.4	0.162
gen_caste	-0.0537435	0.1500954	-0.36	0.72
shigh	**0.0506396** **	0.0160115	3.16	0.002
region	0.1002761	0.0626661	1.6	0.11
male	0.1538462	0.1009937	1.52	0.128
brothers	-0.0717538	0.0879664	-0.82	0.415
sisters	0.1007952	0.0832007	1.21	0.226
schkid	-0.0236912	0.0736059	-0.32	0.748
have_siblings	**-0.2392979** **	0.1141519	-2.1	0.036
genstruc	-0.034313	0.0702494	-0.49	0.625
grpmem	-0.0112017	0.0999605	-0.11	0.911
badevent	-0.0647237	0.1076188	-0.6	0.548
hdsex	0.1991745	0.1673825	1.19	0.234
hhsupp	0.0374332	0.038523	0.97	0.331
rural	**-0.4456068** **	0.2012855	-2.21	0.027
_cons	-2.25433	1.424509	-1.58	0.114

Table 10: Government Schools without sibling composition

Probit regression	Number of obs	555
	LR chi2(17)	27.7
	Prob ¿ chi2	0.0486
Log likelihood = -369.98293	Pseudo R2	0.0361

goodschool	Coef.	Std. Err.	z	P¿z
hhsize	-0.0114412	0.0328561	-0.35	0.728
wi1	**0.9274539****	0.4201656	2.21	0.027
hindu	-0.2542629	0.2151283	-1.18	0.237
gen_caste	-0.2428162	0.2227536	-1.09	0.276
shigh	-0.0095287	0.02206	-0.43	0.666
region	0.0431397	0.0769334	0.56	0.575
male	-0.0305449	0.1164739	-0.26	0.793
brothers	-0.0962165	0.1065625	-0.9	0.367
sisters	0.087458	0.0931792	0.94	0.348
schkid	0.0215505	0.0840817	0.26	0.798
siblings	0.1226997	0.1371653	0.89	0.371
genstruc	-0.0439507	0.0833412	-0.53	0.598
grpmem	-0.0026215	0.1143491	-0.02	0.982
badevent	-0.04154	0.1278822	-0.32	0.745
hdsex	0.1493124	0.1921324	0.78	0.437
hhsupp	0.0122404	0.0446405	0.27	0.784
rural	**-1.068338*****	0.4068836	-2.63	0.009
_cons	-0.0543024	1.781824	-0.03	0.976

Table 11: Rural government schools

Probit regression			Number of obs	532
			LR chi2(16)	11.09
			Prob > chi2	0.8039
Log likelihood = -363.07374			Pseudo R2	0.015

goodschool	Coef.	Std. Err.	z	P>z
hhsize	-0.0120596	0.0330548	-0.36	0.715
wi1	**0.9257615****	0.422637	2.19	0.028
hindu	-0.2290012	0.2191508	-1.04	0.296
gen_caste	-0.2597682	0.2261341	-1.15	0.251
shigh	-0.016659	0.0226362	-0.74	0.462
region	0.0638728	0.0779556	0.82	0.413
male	-0.0569862	0.1175809	-0.48	0.628
brothers	-0.1227795	0.1083866	-1.13	0.257
sisters	0.0711349	0.0941779	0.76	0.45
schkid	0.0319023	0.0850182	0.38	0.707
siblings	0.0988667	0.138249	0.72	0.475
genstruc	-0.0355041	0.0839289	-0.42	0.672
grpmem	-0.0126815	0.1150625	-0.11	0.912
badevent	-0.0560419	0.1285032	-0.44	0.663
hdsex	0.1193062	0.1943346	0.61	0.539
hhsupp	0.0069139	0.0449633	0.15	0.878
_cons	-1.493557	1.702066	-0.88	0.38

Table 12: Rural Private School

Probit regression		Number of obs	219	
		LR chi2(18)	59.35	
		Prob > chi2	0	
Log likelihood =	-16.84001	Pseudo R2	0.638	
goodschool	Coef.	Std. Err.	z	P>z
hhsize	0.9429509	0.5654117	1.67	0.095
wi1	-3.021688	2.244689	-1.35	0.178
hindu	0.0034069	0.7672595	0	0.996
gen_caste	**-3.059721****	1.346611	-2.27	0.023
shigh	**0.204219***	0.1056475	1.93	0.053
region	-1.937035	0.8635568	-2.24	0.025
male	0.5710004	1.890194	0.3	0.763
brothers	-3.806295	5.552212	-0.69	0.493
sisters	7.038441	628.1166	0.01	0.991
schkid	-1.278484	0.9799198	-1.3	0.192
siblings	-0.4858945	0.9538465	-0.51	0.61
genstruc	0.1847538	1.489559	0.12	0.901
grpmem	0	(omitted)		
badevent	0.0301058	0.6916483	0.04	0.965
hdsex	0	(omitted)		
hhsupp	0.7559224	0.6437499	1.17	0.24
sibcomp				
2	4.436239	6.247506	0.71	0.478
3	-8.067866	628.133	-0.01	0.99
4	-7.598118	628.1181	-0.01	0.99
5	0	(empty)		
6	1.059741	628.2106	0	0.999
_cons	43.46728	19.837	2.19	0.028

Table 13: Urban private schools

Probit regression			Number of obs	117
			LR chi2(14)	41.51
			Prob > chi2	0.0001
Log likelihood =	-17.93407		Pseudo R2	0.5365
goodschool	Coef.	Std. Err.	z	P>z
hhsize	1.01141	0.5331076	1.9	0.058
wi1	-0.4098919	3.412301	-0.12	0.904
hindu	-0.0133297	0.8046106	-0.02	0.987
gen_caste	**-2.516472****	1.126885	-2.23	0.026
shigh	0.1295414	0.0859656	1.51	0.132
region	**-1.381688****	0.6331049	-2.18	0.029
male	-0.2449549	0.8029945	-0.31	0.76
brothers	1.071217	0.9121165	1.17	0.24
sisters	0.899956	0.7720655	1.17	0.244
schkid	**-1.725396***	0.8855538	-1.95	0.051
siblings	-0.2912826	0.7699406	-0.38	0.705
genstruc	-0.4414701	0.5266668	-0.84	0.402
grpmem	0	(omitted)		
badevent	-0.2837245	0.8331998	-0.34	0.733
hdsex	0	(omitted)		
hhsupp	0.4663865	0.6787158	0.69	0.492
_cons	30.43594	14.87134	2.05	0.041

Table 14: Private and Rural: Good School defined by Facility

Probit regression		Number of obs	200	
		LR chi2(14)	18.53	
		Prob > chi2	0.1835	
Log likelihood = -128.72133		Pseudo R2	0.0672	
good_facility	Coef.	Std. Err.	z	P>z
hhsize	0.0703241	0.0627353	1.12	0.262
wi1	0.6880115	0.6868969	1	0.317
hindu	-0.2664357	0.4237753	-0.63	0.53
gen_caste	-0.1898148	0.2312702	-0.82	0.412
shigh	0.0431551	0.026777	1.61	0.107
region	0.3341481	0.1406664	2.38	0.018
male	-0.0388159	0.2263546	-0.17	0.864
schkid	-0.0097313	0.1400892	-0.07	0.945
have_siblings	0.0027437	0.1922371	0.01	0.989
genstruc	0.0828129	0.1373373	0.6	0.547
grpmem	-0.1610325	0.2048946	-0.79	0.432
badevent	-0.23591	0.2113217	-1.12	0.264
hdsex	**0.7756417****	0.3684818	2.1	0.035
hhsupp	0.0077892	0.0709863	0.11	0.913
_cons	-8.624588	3.200066	-2.7	0.007

Table 15: Government, Rural, Good School Defined by Facility

Probit regression		Number of obs	533	
		LR chi2(14)	19.23	
		Prob > chi2	0.1564	
Log likelihood = -212.27916		Pseudo R2	0.0433	
good_facility	Coef.	Std. Err.	z	P>z
hhsize	0.0485335	0.0386372	1.26	0.209
wi1	**1.098556****	0.5287253	2.08	0.038
hindu	-0.0739808	0.2798596	-0.26	0.792
gen_caste	-0.3415027	0.3221788	-1.06	0.289
shigh	**-0.0879906****	0.0346976	-2.54	0.011
region	0.0857118	0.0979393	0.88	0.381
male	-0.1334158	0.1467603	-0.91	0.363
schkid	**-0.1779886****	0.0873369	-2.04	0.042
have_siblings	0.0997776	0.157396	0.63	0.526
genstruc	-0.0050839	0.1013283	-0.05	0.96
grpmem	0.0705148	0.1425599	0.49	0.621
badevent	0.0084346	0.1622943	0.05	0.959
hdsex	-0.2892301	0.2681678	-1.08	0.281
hhsupp	-0.0495562	0.0525089	-0.94	0.345
_cons	-2.851907	2.133105	-1.34	0.181

Table 16: Private, Urban: Good School Defined by Infrastructure

Probit regression			Number of obs	155
			LR chi2(14)	27.52
			Prob > chi2	0.0165
Log likelihood = -47.725026			Pseudo R2	0.2238
good_infr	Coef.	Std. Err.	z	P>z
hhsize	**0.2555596****	0.1252679	2.04	0.041
wi1	-1.982258	1.765978	-1.12	0.262
hindu	-0.1578641	0.4617398	-0.34	0.732
gen_caste	**-1.250583*****	0.4594529	-2.72	0.006
shigh	**0.0901865****	0.0400711	2.25	0.024
region	-0.2114675	0.2233274	-0.95	0.344
male	-0.0538577	0.3606976	-0.15	0.881
schkid	-0.168679	0.2750722	-0.61	0.54
have_siblings	-0.4967461	0.3542358	-1.4	0.161
genstruc	-0.0733995	0.243779	-0.3	0.763
grpmem	0.1837014	0.4872679	0.38	0.706
badevent	0.4429551	0.5052547	0.88	0.381
hdsex	-0.4072005	0.5269938	-0.77	0.44
hhsupp	-0.16227	0.1830088	-0.89	0.375
_cons	7.344432	5.444644	1.35	0.177

Table 17: Private, Rural: Good School defined by Infrastructure

Probit regression			Number of obs	200
			LR chi2(14)	28.83
			Prob > chi2	0.011
Log likelihood = -43.747666			Pseudo R2	0.2478

good_infr	Coef.	Std. Err.	z	P>z
hhsize	0.1797752	0.1359007	1.32	0.186
wi1	-0.773571	1.209137	-0.64	0.522
hindu	0.2145489	0.668293	0.32	0.748
gen_caste	-0.5162877	0.35231	-1.47	0.143
shigh	**0.1162643****	0.0494031	2.35	0.019
region	0.3924769	0.2556415	1.54	0.125
male	-0.6526351	0.4308571	-1.51	0.13
schkid	0.2483252	0.2993691	0.83	0.407
have_siblings	-0.0748022	0.3348802	-0.22	0.823
genstruc	-0.1624507	0.2125204	-0.76	0.445
grpmem	0.7693403*	0.398954	1.93	0.054
badevent	0.0131919	0.3249954	0.04	0.968
hdsex	0.1582714	0.7588312	0.21	0.835
hhsupp	**-0.3304783****	0.1493307	-2.21	0.027
_cons	-7.082858	5.6836	-1.25	0.213

Table 18: Urban Private: Good School Defined by Teacher's quality

Probit regression		Number of obs	155
		LR chi2(14)	22.42
		Prob > chi2	0.0705
Log likelihood = -70.335804		Pseudo R2	0.1374
good_teach_quali	Coef.	Std. Err.	z
hhsize	0.0394274	0.0909779	0.43
wi1	-1.150615	1.399488	-0.82
hindu	0.1764445	0.3665487	0.48
gen_caste	**-0.7249831****	0.3206249	-2.26
shigh	0.0188224	0.0317807	0.59
region	-0.0814305	0.1699651	-0.48
male	0.339111	0.3116306	1.09
schkid	-0.2424023	0.2076875	-1.17
have_siblings	0.1548774	0.2767583	0.56
genstruc	0.0362152	0.2126151	0.17
grpmem	0.6413586	0.4464743	1.44
badevent	-0.3916267	0.3792628	-1.03
hdsex	0.4968192	0.48378	1.03
hhsupp	0.1866121	0.1513117	1.23
_cons	2.189666	4.060251	0.54

Table 19: Rural, Private: Good school by Teacher's quality

Probit regression		Number of obs	200
		LR chi2(14)	30.12
		Prob ¿ chi2	0.0073
Log likelihood = -121.60114		Pseudo R2	0.1102
good_teach_quali	Coef.	Std. Err.	z	P>z
hhsize	-0.0635892	0.0622165	-1.02	0.307
wi1	-1.009373	0.7157252	-1.41	0.158
hindu	-0.2886014	0.4761365	-0.61	0.544
gen_caste	-0.3810317	0.2313944	-1.65	0.1
shigh	0.0337074	0.0272628	1.24	0.216
region	-0.1545793	0.1437601	-1.08	0.282
male	-0.4095935	0.2320994	-1.76	0.078
schkid	0.0602125	0.1406313	0.43	0.669
have_siblings	-0.1030693	0.1975438	-0.52	0.602
genstruc	-0.2082422	0.1382932	-1.51	0.132
grpmem	-0.4663883**	0.2067981	-2.26	0.024
badevent	-0.4053011	0.2152159	-1.88	0.06
hdsex	0.6119562	0.349685	1.75	0.08
hhsupp	-0.0090514	0.071055	-0.13	0.899
_cons	5.064372	3.284526	1.54	0.123