GENDER BIAS IN EDUCATION IN WEST BENGAL

Chayanika Mitra

**A THESIS SUBMITTED TO THE
INDIAN STATISTICAL INSTITUTE
IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR
THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY**

**INDIAN STATISTICAL INSTITUTE
KOLKATA**

June, 2019

Dedicated to my parents, Mr. Goutam Mitra and Mrs. Sumita Mitra,

and my brother, Mr. Agantuk Mitra.

Acknowledgements

It is my great pleasure to submit this thesis at the end of five years of rigorous research at the Indian Statistical Institute. It is also an occasion to express my gratitude to the persons I had the privilege to be associated with during the course of my work.

First and foremost is my Ph.D. supervisor Professor Amita Majumder, who not only introduced me to the world of research, but also guided me throughout with a keen interest and active support at every stage of my work. She had introduced me to the avenue of Applied Econometrics and taught me to handle a new problem on my own. I owe an immense debt of gratitude to her. To Professor Nityananda Sarkar I want to express my deep respect and gratitude not only for his erudition but also for his philanthropic ideals, which has been a source of inspiration for me. I am indebted to Professor Dipankor Coondoo for the help he has provided in so many ways. I would also like to thank Professor Arpita Dhar for motivating me to appear for the ISI entrance exam

Most of the learning processes at the time of my Ph.D. have happened while discussing with my colleagues of our institute. I would also like to give them special thanks for these stimulating conversations. I would like to thank Mr. Arindam Paul, Dr. Tanmoy Das, Ms. Madhuparna Srivastava and Mr. Abhinandan Sinha for these discussions. I would specially thank Dr. Jayant Jha for clarifying statistical doubts and helping me to design the outline of the thesis. I would also like to give special thanks to Dr. Sandip Sarkar, who had introduced me to STATA, which is the major programming language in my thesis. Moreover, I would like to thank the faculty members of Economic Research Unit, ISI, Kolkata. I would like to thank many of my friends at my hostel who actually helped me a lot in the arduous journey and kept my morale high during difficult times. I would like to specially thank Dr. Jayant Jha, Ms. Tamalika Koley, and Ms. Sreoshi Banerjee. I take this opportunity to extend my sincere thanks to the anonymous referees for their insightful comments on the papers based on some chapters of the thesis.

I would like to express my deepest gratitude to my parents Mr. Goutam Mitra and Mrs. Sumita Mitra and my brother Mr. Agantuk Mitra for their love and mental support for research. Without their support, the long journey of enlightenment could not have been completed.

I am grateful to the Indian Statistical Institute, Kolkata, India, for supporting me financially through a fellowship and for providing me infrastructural facilities.

Table of Contents

List of Tables

List of Appendices

Chapter 1

Introduction and Chapter-wise Summary

1.1 Introduction

There exists a vast literature with evidences of gender discrimination against females in allocation of goods and services at regional/national level in India. Evidence of gender bias on Indian data include, among many others, studies by Subramanian and Deaton (1991), Azam and Kingdon (2013), Kingdon (2005), Drèze and Kingdon (2001), Zimmerman (2012) and Lancaster, Maitra, and Ray (2008). India has a history of preference for sons over daughters for cultural and economic reasons. This is manifested through high birth sex ratios, which is possibly a result of female foeticide. The child sex ratio is within the normal natural range in all eastern and southern states of India, but is significantly higher in certain western and particularly northwestern states (Census 2011).

Discrimination against women is also evident from gender pay gap, which refers to the difference in earnings between women and men in the paid employment and labor market. For the year 2013, the gender pay gap in India was estimated to be 24.81% (Varkkey and Korde, 2013). Some of the main factors that contribute to the existing gender pay gap in India are the following: (i) Occupational segregation: this is mainly concentrated in rural areas in the agricultural sector, where the rate of female participation in the paid labor market is generally low (Elborgh-Woytek et al., 2013); (ii) Cultural barriers and social norms (Nussbaum et al., 1995) and (iii) Education and training, investment in which has also been strongly in favour of men, as they are brought up with the expectation of being bread earners, while women are instead viewed as "future homemakers", for whom education may not be as essential (Rustagi, 2005; Ghosh, 2010). Moreover, households consider investment on a girl child's education an additional monetary burden over and above the dowry and expenditure on marriage.

This thesis focuses on gender bias in the field of education, where a substantial number of studies on India point to the existence of such bias. Broadly, the research papers on gender bias in education can be categorized into two groups of issues: (i) school participation and (ii) expenditure on education.

In the context of school participation (enrolment rate and dropout behaviour), Kambhampati and Pal (2001), and Pal (2004), using data from rural West Bengal, found significant evidence that paternal and maternal education explain gender differences in both school enrolment and attainment. While father's education level has a positive impact on boys' schooling only and does not have any perceptible impact on girls, mother's literacy significantly enhances the probability of school enrolment among girls, but it is non-significant for boys. Evidences of female backwardness in education are found in studies by Balatchandirane (2003), which reports that 'South Asia records a net enrolment rate for girls which is 20 per cent less than that for boys', and by Bandyopadhyay and Subrahmanian (2008) noting that India constitutes almost 30% of the illiterate world population and 70% of them are women. Drèze and Kingdon (2001) found strong evidence of sharp gender bias against girls in school participation in rural north Indian states of Uttar Pradesh, Bihar, Madhya Pradesh and Rajasthan. They found parental education, both maternal and paternal, as important determinants of probability of school participation. In their paper, school infrastructure and household characteristics were the major determinants of a child to be enrolled in a school and to continue study further. However, the effects of these factors were more prominent for girls than for boys. Prakash et al. (2017), based on data from villages in Karnataka, found that vulnerability of adolescent girl children is a reason for discontinuing school. There are official evidences as well. The Govt. of India (2013) reported very low enrolment rate in the primary and secondary education level for female students compared to male students. According to the "Educational statistics at a glance" Report (2016) by the Ministry of Human Resources, India, the dropout rate among boys and girls were 4.53% and 4.14%, respectively. The topmost reasons for dropping out are lack of interest in studies, economic condition of the parents, lack of parental motivation, cost of schooling, migration of family and demand for child labour for domestic and other purposes. After dropping out from upper primary classes, students have been found to be generally helping the parents in earning money, either by getting involved in agriculture and allied activities or directly by working to earn money. Kambampati (2008) concluded that mothers and fathers in India make different decisions for girls vis-à-vis boys and that the variables reflecting mothers' autonomy vary in their impact.

In the context of intra household expenditure on education, Kingdon (2005) pointed out that the standard Engel curve approach fails to detect gender bias in the allocation of resources for education in a household. Education expenditure is different from other item specific expenditures in a household as a substantial percentage of households report zero spending on educational items. Gender bias in education expenditure is possibly due to two major reasons. First, the households do not spend on education for a girl child, which leads to zero spending. Second, conditional on positive spending, the spending on a girl child is less than that on a boy child in a household. Hence, the analysis of the educational items is not straightforward like other household items. Azam and Kingdon (2013), using both household and individual level data for various states of India, used a Double Hurdle Model (Cragg, 1971) to separate out the decision of children schooling into (i) whether to spend (enrol) and (ii) how much to spend. These studies found little evidence of gender-bias in the enrolment decision in age group 5–9, while pro-male bias in education expenditure conditional on enrolment was observed in several states. Zimmerman (2012) found that discrimination against girls increases with age and is robust to the statistical method and the expenditure measure at the all-India level, although state-level results are more sensitive to the statistical measures used. Lancaster, Maitra, and Ray (2008), using a two-stage collective approach to household behaviour, found evidence of significant gender bias in educational attainment with the bias running in favour of boys in Bihar and Maharashtra for the age group of 10-16 years. However, they found that gender bias actually runs the other way in rural Kerala and there is no gender difference in education expenditure in urban samples. The direct and indirect costs involved in the completion of the education of a child are major reasons for the presence of gender disparity in the resource-constrained households in India (Mohanty, 2006). Mohanty (2006) also pointed out that the probability of a male child being enrolled to school is higher than that of a female child and parents make pro male biased decision in the expenditure on education.

In this thesis, we confine our study to the state of West Bengal. One reason is that extensive research work to examine gender bias in the context of education for the state of West Bengal is not much in number. We have referred to some studies above. There are few other evidences of the presence of gender bias as follows. First, Tisdell (2002) pointed out that there exists a tendency of favouring sons over daughters, which results in the deprivation of female children in terms of education, health and other welfare parameters in the rural villages in West Bengal. Second, almost 75% of out-of-school children in India were from West Bengal, the sex ratio was very poor and only 50 girls for each of 100 boys belonging to the age group of 5-14 years were in school in West Bengal (Kabeer, 2003). Third, in the context of marriage of girls at early age, West Bengal (53%) came fifth after Bihar, Jharkhand, Rajasthan and Andhra Pradesh (Ghosh and Kar, 2010). Finally, in 2013-14 the dropout rates at the upper primary, secondary and grades XI – XII levels in West Bengal was much higher than the figures at the national level.[1]

Further, we have analyzed the problem of gender bias in education for both sectors in West Bengal. In the context of opportunities in and attitude towards education, the difference between the rural and the urban sector is prominent in India. West Bengal is no exception. On the demand side, the students in the rural sector get lesser opportunities and exposure. The difference between the monthly per capita expenditure in the rural and the urban sector is high, indicating the economic constraints faced by the parents in the rural sectors while educating their children. The children in the rural sector are less motivated to complete their studies compared to that in the urban sector as the parents in the rural sector are less educated than parents in the urban sector. On the supply side, the availability of nearby schools and the availability of good teachers in these schools are the major constraints in the rural sector. A school in every village in the rural sector is not always available. Hence, children have to travel to another village to study. As a result, parents find it more difficult to send their daughters to schools. The children consider travelling to schools is a waste of money and time as the schools are far away from their residences (Sahapur and Omprakash, 2017). In the context of availability of good teachers, most of the teachers look for jobs in the urban sector due to more opportunities and higher wage. The facilities provided by the school in the rural sector are technologically orthodox and financially constrained. The school performance score is overall better in the urban sector than that in the rural sector (Chavan and Chavan, 2018). As a result, the good students in the rural sectors fail to

[1] Source; Department of School Education and Literacy, Ministry of Human Resource Development, Government of India, 2015.

compete with their counterparts in the urban sector. Apart from that, the opportunities to get absorbed in the labour market are also lesser in numbers for rural students. Some students, who have the privilege, migrate to nearer big cities for jobs. However, this is not applicable for all students in the rural sector.

This thesis is an endeavour to explore the issue of gender bias in the field of education in the following directions: (i) gender bias among students, both at school and college level, in terms of participation in the education system, as well as expenditure allocation; (ii) the role of mother's bargaining power in allocating educational expenditure to boys and girls; and (iii) dropout behaviour of children.

1.2 Data Used

The empirical analysis is based on two types of data provided by the National Sample Survey Office (NSSO), India. The first source is household-level unit record data on consumption expenditure and the second is individual level data within households mainly on education expenditure and educational attainment. For Chapter 2, the latest available (at the time of the study) 68th round (2011-12) household-level consumption expenditure data have been used. Chapter 3 uses the 64th round (July 2007-June 2008) education expenditure and educational attainment data. This was the latest data available at the time of the work. The analysis of Chapter 4 is based on 68th round (2011-2012) consumption expenditure data. Finally, Chapter 5 uses the 71st round (January 2014-June 2014) data on education expenditure, which became available during the final stages of this work. Appendix I, at the end of this chapter, gives a detailed description of the data used.

1.3 Organisation of Thesis

The organization of the thesis is as follows. **Chapter 2** deals with identification of the educational items, which are exclusively consumed or used by the 'students'. **Chapter 3** examines the issue of gender bias in below class-10 and above class-10 levels of the education system in West Bengal. **Chapter 4** is concerned with gender bias in the context of intra household allocation. **Chapter 5** analyzes the dropout behaviour of children from two different perspectives, namely, the parents' perspective and child's perspective. **Chapter 6** concludes the thesis with a brief discussion on the future works. Chapters 2-5 have Appendices at the end of the respective chapters.

The next section provides a brief overview and summary of Chapters 2-5 of the thesis.

1.4 Chapter-wise Summary

Chapter 2: (Identification of Child Educational items from Household Level Data: The Outlay Equivalent Ratio Approach)[2]

In this chapter we identify the educational items which are used exclusively by the 'students' in a household from *household-level* data and examine the presence of gender bias in the expenditure on each of the educational items separately. Typically, household surveys on consumer expenditure report expenditures on individual items only at the household-level and not at the individual level. Hence, separating out educational items, exclusively for children and young adults, from household level data is not straightforward. To identify such items from the 68th round (2011-12) household-level consumption expenditure data, we first select the items under

[2] An earlier version of this chapter has been published as 'Majumder, A and Mitra, C. (2016): Gender bias in household education expenditure: the case of West Bengal, *Indian Growth and Development Review*, 9(2), 129-150'.

‘educational expenditure’ that closely relate to the items that are marked exclusively as ‘student items’ in the individual level 64th round (July 2007-June 2008) education expenditure data. We then verify that these are indeed ‘student exclusive items’ using the Outlay Equivalent Ratio approach of Deaton (1989).[3]

The idea behind the outlay equivalence ratio (Deaton, 1989) approach is – children do not have resources to meet their requirements and they have to depend on adults of the household. Hence, given household income, adults have to cut back from their private expenditure (expenditure on the goods exclusively consumed by the adults of the household) in order to fulfill the requirements of children. Consider first the question of whether a particular group of goods can be treated as ‘adult goods’. An additional child brings additional needs (education expenditure in our case), so that expenditure on all adult goods can be expected to fall. This effect acts exactly like a reduction in income. As a result, addition of a child to a household gives a negative income effect for those goods, which are exclusively consumed by the adults of a household. The ‘Outlay Equivalent Ratio’ of an item is simply the ratio of outlay change due to an additional person and marginal propensity to spend on the item. Thus, If ‘j’ is an adult good and ‘s’ is a child demographic category (student), then π_{js}, the outlay equivalence ratio, is likely to have a negative sign as increase in the number of children will decrease expenditure on adult goods. If a household favours boys over girls, then households will reduce the expenditure on adult goods more for boys than for girls. Using Working-Leser forms of Engel curves for various items of consumption, Subramanian and Deaton (1991) and Deaton (1997) found evidences of pro-male

[3] The reason behind trying to identify items which are consumed exclusively by the ‘students’ by matching these two data sets is that for our work on intra household allocation in Chapter 4 we need information on expenditure on such items as well as on other items. While the consumer expenditure survey data provides information on all items, but does not give explicit information on items consumed exclusively by the ‘students’, the education expenditure data gives individual level information on consumption of educational items only, and not on other items of consumption.

bias in rural Maharashtra in the 10–14 year age group through identification of adult goods such as adult clothing, alcohol, tobacco, entertainment etc.

We use this approach to identify educational items that are exclusively for students. We argue that, if 'i' is an educational item and 's' is a student group then $\pi_{is,}$ the outlay equivalent ratio of item i with respect to demographic group s, is likely to be positive, but at the same time the additional expenditure will be financed by cutting down expenditure on adult goods, so that π_{js} will have a negative sign, where 'j' is an adult good. So, for our case, to detect gender bias, if $\pi_{is_m} > \pi_{is_f}$ and $\pi_{js_m} < \pi_{js_f}$ where 'm' denotes male student and 'f' denotes female student, we say there is pro male bias.

We may also mention that because a large percentage of households reports zero expenditures on educational items, the possibility of selection bias is high. We, therefore, use Heckman's (1979) two-step procedure for estimation. Also, in place of the Working-Leser form we use the log-quadratic budget share form, underlying the Quadratic Almost Ideal Demand System (QUAIDS) [Banks, Blundell and Lewbel (1997)].

From the analysis we find that all educational items relate exclusively to students of a household. Educational items such as 'book', 'stationery and photocopy charges', 'tuition fees' and 'private coaching fees', are found to be students' items. 'Transport cost', 'Pan', 'Tobacco' and 'Alcohol' are found to be 'adult goods' in respect of household members belonging to the age group of 5-29 years. We find evidence of pro male bias in expenditure on educational items and the extent of gender bias is more in the age groups of 10-16 years and 17-29 years in both sectors in West Bengal. This corroborates the findings of Kingdon (2005, 2013).

Chapter 3 (Gender Bias in Below Class-10 and Above Class-10 Levels: Application of Heckman's Two Step Model and Multinomial Logit Model using Individual Level Education Expenditure Data)[4]

In India, the education system has two parts: below class-10 level and above class-10 level. These two levels are different, because at below class-10 level, the decision of a household involves two stages: first, whether to send the child to school (or similar educational institution) or not and second, given that the child will be sent to an educational institute, the amount to be spent on the child for education [Kingdon (2005); Azam and Kindgon (2013); Deaton (1997)]. For this section the analysis can be further split up into classes 1-8 and classes 9-10, in view of the Right to Education act (2005). The reason is the following: according to the Right to Education (RTE) act (2005), every child from 6-14 years should be provided free elementary schooling by Central and State government bodies. The government education bodies will not charge any fees like tuition fees, capitation fees etc. from the students belonging to this age group. Households' decision in the above class 10 levels, on the other hand, involves the issue of choice of specialization (e.g., Science, Arts, and Commerce). Given the fact that higher studies call for higher costs and inter-specialization differences in cost are rather high, gender bias may have some role in the choice of specializations. Hence, it is important to study the nature of gender bias at this level separately.

In the below class-10 level, the two stage decision process of the household has been estimated by the Heckman’s two-step model. As the decision of a household will still involve two stages, the same model has been used separately for the two different sections i.e., below class 8 and class 9-10 in below class-10 level. Gender bias is tested by hypotheses testing at both the participation stage and the outcome stage.

[4] This chapter has been published as ‘Majumder, A., and Mitra, C. (2017): Gender Bias in Education in West Bengal, *Journal of Quantitative Economics*, 15(1), 173-196’.

For the above class-10 level, we use a model analogous to the wage differential model of Brown et al. (1980). This approach enables us to assess the relative importance of within- and between-subject expenditure allocation effects. The total differential of the education expenditure between male and female is decomposed into four categories – Within Explained (*WE*), Within Unexplained (*WU*), Between Explained (*BE*) and Between Unexplained (*BU*). In the full model, *WE* and *BE* capture the expenditure differentials due to differences in measured characteristics between males and females, while those that are unexplained (*WU* and *BU*) reflect differential returns to the measured characteristics and may be the result of discrimination. The estimation of this model requires within-subject expenditure regressions and a technique for predicting the educational distribution of females assuming that they are subject to the same structure of determinants as males. The multinomial logit model captures the way the variables affect the probability of an individual studying a particular subject, treating the subject choice as endogenously determined.

Our model allows us to examine two issues that are often of concern in studies of discrimination. First, we are able to divide the total differential into explained and unexplained portions. Here we explicitly include the contribution of stream wise segregation and thus avoid the bias that is implicit when stream is ignored or incorrectly included. The second calculation examines the separate effects of within- and between-stream expenditure differences. The analysis is based on 64th round (July 2007-June 2008) education expenditure and educational attainment data.

This study establishes that there exists pro male bias in the decision to send the child to school and spending on education in the below class-10 level in both the rural and the urban sectors for children belonging to the age groups 10-14 years and 15-16 years, which corroborates the findings of Kingdon (2005). For the above class-10 level, the detailed analysis indicates evidence of discrimination against girl students at the level of selection of subject streams such as Science and Commerce and also in the allocation of expenditure within a subject stream in the urban areas in the state of West Bengal during 2007-08.[5]

Chapter 4 (Gender Bias in Education in West Bengal: An Application of Intra-Household Collective Model)

The recent literature on household demand models focuses on intra-household models. In intra-household models, each individual household member is characterized by his/her own rational preferences. This is in contrast with the conventional unitary models, where a household as a whole is assumed to be a single decision maker. Pareto efficient models assume that husband and wife jointly choose an efficient allocation of resources through a ‘sharing rule’, although they have separate utility functions. ‘Sharing rule’ of allocation of resources is often interpreted as an indicator of the bargaining power and governs the within-household distribution (apportionment) of household resources. An attractive feature of the resource-sharing rule is that it is expressed in monetary terms. The sharing rule is also useful for recovering information about the economic well-being of household members [Cherchye et al. (2015)]. Identification and estimation of sharing rule empirically is not, however, straightforward. As noted earlier, typically, consumption expenditure surveys report data at the household level and hence, individual level information is not available. Various methods have so far been proposed to resolve this problem.

[5] The data did not permit examination of discrimination in the rural sector in our framework at the above class-10 level.

One of the methods of estimating sharing rule is to assume that the unobservable sharing rule interacts with individual incomes *a la* Barten (Barten 1964, Perali 2008) and that at least one assignable good, or, equivalently, two exclusive goods[6] are observable (Aria, Atella and Peralli (2004)). We follow this line to estimate sharing rule in the state of West Bengal. The underlying assumption is that parents first agree on the sharing rule, determining the direction and amount of transfer between members and then they maximize their individual egoistic utility functions subject to their private budget constraint. After having estimated the sharing rule and the budget shares, the contribution of parents (individual) on expenditure on education of children is determined. The methodology is the following. First, the sharing rule parameters (that determine the apportionment of resources between members) within a household are estimated in an intra-household collective framework. Here, we have taken the households with a structure of husband, wife and at least one male and one female child. Next, using this sharing rule information, households are categorised into two types, viz., the ones in which the husband has a higher share of resources (higher bargaining power) and the ones in which the wife has a higher share of resources. Finally, the budget share function of education expenditure is estimated for each of these two types of households and gender bias is assessed through testing of equality of allocation of expenditure for male and female children. The analysis is based on a Quadratic Almost Ideal Demand System (QUAIDS) [Banks, Blundell and Lewbel (1997)] and the 68th round (2011-12) household level consumption expenditure data.

[6]A good is exclusive when one identifiable member of the household only consumes a strictly private good. A good is assignable when each member of the household consumes a strictly private good in observed proportions.

The findings that emerge are: pro male bias in education expenditure is observed in households where husbands enjoy higher bargaining power. However, when the wife is the decision maker, no gender bias in education expenditure is observed.

Chapter 5 (Dropout Behaviour of Children: The Case of West Bengal)

In most of the empirical works, the *incidence* of dropout is analysed taking the dependent variable to be a dichotomous variable. It is, however, more important from policy perspective to track the paths for the chosen reasons for the child to drop out [Christen Bradley and Renzulli (2011)]. In this chapter, we address the dropout problem from two different perspectives. First, keeping in mind that the socio-economic and socio-demographic backgrounds of a household are influential factors for the dropout decision, we use a Probit regression to identify those household factors which prevent the child from going to school, using household level information available in our dataset. Second, in line with the study by Bradley and Renzulli (2011), we classify the reasons, as stated by the students, into certain groups of reasons for dropout. There are four reason categories for the boy child and five reason categories for the girl child. We try to relate the reasons for dropout (as specified by the child) with the background of the household the child belongs to and with the school infrastructure provided to him/her, through a Multinomial Logit model.

According to Bradley and Renzulli (2011), the “push” factor group refers to the supply side factors that discourage a student from going to school. These factors are generally related to facilities provided by school authority: the environment of classes, medium of instruction, unable to cope up with studies, unfriendly atmosphere, quality of teachers etc.

The "pull" factors refer to the reasons for dropout as a cost-benefit analysis [Stearns and Glennie (2006)]. In the context of India, almost every household has more than one child. Normally, in middle class and lower middle class families, the elder child in the household takes more responsibilities for the household. For example, a student may have to work with family members in farming, take care of younger siblings or may have to go out to work for earning money for the family. Consequently, he is unable to devote the required time and effort needed to succeed in studies. This leads to failure and discontinuity in studies.

"Opted out" factors refer to individual specific factors. Apart from push and pull factors a student may lose interest in studies or may think that a certain level of achievement in education is sufficient for him/ her.

"Other reasons" refer to the reasons that do not fall into the above categories. For example, it might be a tradition in the family of a student not to study further.

For a boy child, the reasons for dropping out, as stated by the child, are categorized into the above-mentioned four groups - push factor, pull factor, opted out factor and other factors. For a girl child, another category is added to these four categories, which is "female-specific reasons". Prakash et al. (2017) noted that in addition to the four reasons, there are some hurdles that only a girl child has to face in India. It might so happen that due to reasons like early marriage, lack of female teachers, improper toilet facilities and distance between school and residence, the girl child has to leave school. These reasons become particularly relevant when the girl child attends adolescence. The data set used in this chapter is the 71st round (January 2014-June 2014) data on education expenditure.

The findings from this study are: in terms of incidence of dropout, children from affluent families and educated parents are less likely to drop out in both the rural and the urban sector. However, belonging to a large family and having many younger siblings negatively affect school participation in the rural sector, while for urban children these variables do not make any difference.

From the children's perspective, for boys who drop out, reporting about "push factor", "pull factor" or "opted out" are less probable compared to "others" if they belong to affluent families and have educated parents. They are also less likely to report these factors if the school is a non-public school providing midday meals.

The same implications hold for girls with similar family background, in terms of the above choices. The additional choice, viz., "female specific reasons" is also less likely, compared to "other" factors, as these families are generally liberal (in terms of marrying off a girl at an early age) and these girls generally go to schools with adequate infrastructure. This is corroborated by the fact that "going to a public school" has a positive impact on the probability of choosing this option, and children from poorer section of the society go to public schools, in general. It is interesting to note that mid-day meal plays no significant role in reporting "female specific reasons". Thus, public school infrastructure emerges as a major factor contributing to female dropout because of 'female specific reasons'. Given that infrastructure includes 'adequate toilet facilities', which are lacking in most public schools[7], this study contributes to the justification for the demand of proper toilet facilities for female children at school.

[7]According to a report (2014) prepared by the Department of School Education and Literacy under the Union Human Resource Development, West Bengal features in the second spot in government schools lacking girls' toilets.

APPENDIX I

Description of the Data

In this appendix, a brief description of the structure of two different types of data sets, viz., the household level consumption expenditure data and the individual level data on the participation and the expenditure in education, is provided.

1. 68th Round Consumption Expenditure Data

The 68th round (July 2011-June 2012) consumption expenditure data provide information on household level expenditure on different items like food, tobacco, alcohol, fuel, medical, education, recreation etc. with some specific household features (such as household size, household type, land holding, etc.) and individual demographic features (such as age, sex, marital status, education levels etc.) for all states for both rural and urban sectors.

1.1 Sampling Design of 68th round

A stratified multi-stage design is adopted for the 68th round survey. The First Stage Units (FSU) are the 2001 census villages in the rural sector and Urban Frame Survey (UFS) blocks are in the urban sector. The Ultimate Stage Units (USU) are households in the both sectors. Within each district of a State/ Union Territory (UT), two basic strata are formed: i) rural stratum comprising of all rural areas of the district and (ii) urban stratum comprising of all the urban areas of the district. The sub-strata are formed by dividing the sample size of each of the rural and the urban strata by 4 and arranging the group of villages and UFS within the sub strata in such a way that each sub strata would likely to have equal population. The total number of sample FSUs are allocated to the States and UTs in proportion to the population as per census 2001 subject to a minimum sample allocation to each State/ UT. State/ UT level sample size are allocated between two sectors in proportion to population as per census 2001 with double weightage to urban sector. Within each sector of a State/ UT, the respective sample sizes are specified for the different strata/ sub-strata in proportion to the population as per census 2001. For the rural sector, from each stratum/ sub-stratum, required number of sample villages are selected by probability proportional to size with replacement (PPSWR), size being the population of the village as per Census 2001. For the urban sector, UFS 2007-12 phase is used for all towns and cities and FSUs are selected from each stratum/sub-stratum by using Simple Random Sampling Without

Replacement (SRSWOR). From each sub-stratum the sample households for each of the schedules are selected by SRSWOR.

2. Participation and Expenditure in Education Data of 64th round and 71st round

The 64th round (July 2007-June 2008) and 71st round (January 2014-June 2014) data provide detailed information on the educational attainment and expenditure on education for all states/UT in India. Moreover, from these data sets a complete profile of household education expenditure pattern at the individual level can also be obtained. The major contents of these surveys are information on the extent of use of the educational infrastructure by the current status of educational attainment, the expenditure on educational items and the reasons for dropouts or discontinuity in education. The sampling design of these two rounds is similar.

Sampling Design of 64th round and 71st round

A stratified multi-stage design was adopted for these rounds of survey. The first stage units (FSU) are the census villages in the rural areas and Urban Frame Survey (UFS) blocks in the urban areas. The ultimate stage units (USU) are households in both rural and urban areas. First-stage units (FSUs) are selected by Probability Proportional to Size with Replacement (PPSWR). Stratification of the households was done based on whether or not the household has any student (aged 5-29 years) receiving technical/professional or general education. Eight households were selected using Simple Random Sampling without Replacement (SRSWOR) from each sample village/block. Information about particulars of household characteristics like household size, household type, religion, social group, distance from nearest school having primary/upper primary/secondary level classes, whether the household has a computer/ any access to internet and household's usual consumer expenditure in a month etc. were collected. The education particulars of the household members, aged 5-29 years, who were currently attending educational institutions at primary level and above, were recorded. Information on course, level, class/grade/year, type of institution, medium of instruction, etc., were also collected on basic course. Particulars of private expenditure on education of the household members were collected. Particulars of household members, aged 5 - 29 years, who were currently not attending any educational institution, were also collected along with information on whether or not they ever enrolled, age at entry in school, age of discontinuation / drop-out etc.

Chapter 2

Identification of Child and Adult Educational items from Household Level Data: The Outlay Equivalent Ratio Approach

2.1 Introduction

The objective of this chapter is to *identify* the educational items *exclusively* for 'students' from *household level* consumer expenditure data. It also tests the difference in the allocation of resources on education, with respect to these items, between a male student and a female student for both sectors in West Bengal, through 'Outlay Equivalent Ratios'.

The 'Outlay Equivalent Ratio' of an item is defined as the ratio of outlay change due to an additional person and marginal propensity to spend on the item. (Deaton, 1989). The idea behind the outlay equivalent ratio (Deaton, 1989) approach is – children do not have resources to meet their requirements and they have to depend on adults of the household. Hence, given household income, adults have to cut back from their private expenditure (expenditure on the goods exclusively consumed by the adults of the household) in order to fulfill the requirements of children. Consider first the question of whether a particular group of goods can be treated as 'adult goods'. An additional child brings additional needs (education expenditure in our case), so that expenditure on all adult goods can be expected to fall. This effect acts exactly like a reduction in income. As a result, addition of a child to a household gives a negative income effect for those goods, which are exclusively consumed by the adults of a household.

If a household favours boys over girls, then households will reduce the expenditure on adult goods more for boys than for girls. Using Working-Leser forms of Engel curves for various items of consumption, Subramanian and Deaton (1991) and Deaton (1989) find evidence of pro-male bias in rural Maharashtra in the 10–14 year age group through identification of adult goods such as adult clothing, alcohol, tobacco, entertainment etc. We use this approach to identify educational items that are exclusively for students (see section 2.2 below).

In the context of gender bias, Azam and Kingdon (2013), Kingdon (2005), Drèze and Kingdon (2001), using both household and individual level data for various states of India, focus particularly on education and adopt an Engel curve based methodology. These studies find little evidence of gender-bias in the enrolment decision in age group 5–9, while pro-male bias in education expenditure conditional on enrolment is observed in several states. Lancaster, Maitra, and Ray (2008), using a two-stage collective approach to household behaviour, find evidence of significant gender bias in educational attainment with the bias running in favour of boys in Bihar and Maharashtra. However, the gender bias actually runs the other way in rural Kerala and there is no gender difference in education expenditure in urban samples.

While the above studies concentrate only on children education and use the Working-Leser form of Engel curves,[8] we extend the scope by including young adults pursuing higher studies[9] and taking a budget share form, which is quadratic in logarithm of income. We also identify 'adult goods' along with 'educational items', with respect to the student demographic group, using the 'outlay equivalent ratio' approach.

Now, household surveys report expenditures only at the household level and not the expenditures that are incurred by individuals within a household separately. Hence, separating out educational

[8]The international evidence of gender bias in education using Working-Leser Engel curves include studies by Hori, Mitsuyama and Shimizutani (2015), Omolo (2012), Parpiev and Yusupov (2009), Burgess and Zhuang (2002), Deaton (1989), Himaz (2010), Aslam and Kingdon (2008), Choden and Sarkar (2013) for Japan, Kenya, Uzbekistan, rural China, Thailand andCôte de I′vore, Sri Lanka, Pakistan and Bhutan, respectively, among many others.

[9]We call this extended group 'students'.

items, exclusively for children and young adults, from household level data is not straightforward. To identify such items from the 68th round (2011-12) household-level consumption expenditure data, we first select the items under 'educational expenditure' that closely relate to the items that are marked exclusively as 'student items' in the individual level 64th round (July 2007-June 2008) education expenditure data. We then verify that these are indeed 'student exclusive items' using the Outlay Equivalent Ratio approach of Deaton (1989).[10] For items to serve as possible 'adult goods', we have considered items that have been identified as 'adult goods' in other studies on Indian data (Subramanian and Deaton (1991)). Typically, as a large percentage of households report zero expenditure on educational items as well as on adult goods, the possibility of selection bias is high. To address the issue of selection bias we use Heckman's (1979) two-step procedure for estimation.

The outline of the chapter is the as follows: in Section 2.2 the methodology is discussed, Section 2.3 presents the analysis of results along with a brief description of the data set, and Section 2.4 ends the chapter with some concluding remarks. Appendix II, at the end of this chapter, presents the derivation of some expressions used in the text.

2.2 Methodology

The methodology is based on four main features:

(1) Engel curve approach is used to capture parental preference for student's welfare and to find the existence of male student favouritism in the field of education.

[10] The reason behind trying to identify items which are consumed exclusively by the 'students' by matching these two data sets is that for our work on intra household allocation in Chapter 4 we need information on expenditure on such items as well as on other items. While the consumer expenditure survey data provides information on all items, but does not give explicit information on items consumed exclusively by the 'students', the education expenditure data gives individual level information on consumption of educational items only, and not on other items of consumption.

(2) In case of exclusive student goods, the addition of a student will increase the resource allocated for students, at the same time reducing the resource allocated for adult goods under the assumption that students do not earn. The underlying analysis is done by the outlay equivalent ratio approach.

(3) If a household favours males over females, then that household is likely to sacrifice more for a male student's education than that for a female student and the expenditure on education for a male student would be more than that on female students.

(4) To address selection bias, Heckman's two step procedure has been used.

2.2.1 Outlay equivalent ratio:

We start with a general form used in Deaton (1989), where household expenditure on good 'i' is a function of total household expenditure and other demographic characteristics, given by

$$p_i q_i = f(x, \underline{n}, z, u), \tag{2.1}$$

where, $p_i q_i$ is the household expenditure on good 'i' (here educational items), x is the total household expenditure, $\underline{n}$ is a vector that represents demographic composition of the household, z is the vector of other household characteristics, u is the term that represents unobservable taste variations. $\underline{n}$ consists of a list of household members belonging to different age-sex categories such that n_s is the number of members belonging to s^{th} demographic category of a household. Here, we have taken six age categories: 5-9 years, 10-16 years and 17-29 years for each sex.[11] Hence, the effect of increase in the number of members in the s^{th} category on household expenditure on good 'i' keeping other things constant is given by $\frac{\partial(p_i q_i)}{\partial(n_s)}$. Income effects are

[11]Following the Majumder and Mitra (2017) study, where age groups 5-9 years, 10-16 years and 17-29 years were considered to examine gender bias at below class 10 and above class 10 levels, here we treat these same age groups as 'students'. The levels and the corresponding age groups are given in Appendix II, Table A2.3.

given by marginal propensity to spend i.e., $\frac{\partial(p_i q_i)}{\partial(x)}$. The ratio $\frac{\frac{\partial(p_i q_i)}{\partial(n_s)}}{\frac{\partial(p_i q_i)}{\partial(x)}}$ relates the effect of additional member of type 's' with the income effect and this ratio conveys how much of the total budget would have to be increased to generate the same additional expenditure on good 'i' as would the addition to the household of one more person of type 's' (Deaton (1989)). The expression of π ratio is the following

$$\pi_{is} = \frac{\frac{\partial(p_i q_i)}{\partial(n_s)}}{\frac{\partial(p_i q_i)}{\partial(x)}} \div \frac{n}{x}. \tag{2.2}$$

So, if 'i' is an educational item and 's' is a student group then π_{is} is likely to be positive, but at the same time the additional expenditure will be financed by cutting down expenditure on adult goods, so that π_{js} will have a negative sign, where 'j' is an adult good. Therefore, to detect gender bias, if $\pi_{is_m} > \pi_{is_f}$ and $\pi_{js_m} < \pi_{js_f}$ where 'm' denotes male student and 'f' denotes female student, we say there is pro male bias.

2.2.2 Heckman's two-step procedure:

Heckman's model is based on a simple two-stage decision framework: the probability of zero or positive spending is first addressed using a Probit model (*participation equation*) and then given that the probability is positive, the factors that influence the amount of spending are studied using an Ordinary Least Squares (OLS) method (*outcome equation*). Thus, the model has two equations: i.e., participation equation and outcome equation.

There are two latent variables Y_1*and Y_2*.

The first latent variable i.e., Y_1^* is a variable which depends on factors that determine whether a household would spend on education or not, but Y_1^*cannot be observed. The decision is represented by a binary variable Y_1, where $Y_1=1$ means that Y_1^* has a strictly positive value and the household is willing to send their child to an educational institution.

$$Y_1 = 1 \quad if \quad Y_1^*>0$$

$$= 0 \; if \quad Y_1^* \leq 0.$$

The *participation equation* is given by

$$Y_1^* = X_1^{*/}\beta_1 + \varepsilon_1 \tag{2.3}$$

Where, ε_1 is a standard normal variate and X_1^* is a vector of independent variables that determine whether a household would send the child to educational institute or not.

The second latent variable Y_2^* (share of monthly spending of the household for education) can only be observed when the household is spending on education and cannot be observed for those who do not spend $(Y_1 = 0)$. The outcome equation represents the observed value of Y_2^* dependent on a vector of independent factors X_2^*.

Thus, the *outcome equation* is given by

$$Y_2^* = X_2^{*/}\beta_2 + \varepsilon_2 \tag{2.4}$$

$$Y_2 = Y_2^* \; if \quad Y_1^*>0$$

$$= 0 \quad if \quad Y_1^* \leq 0,$$

Where, ε_2 is a normal variate with mean 0 and variance σ^2.

The OLS regression of Y_2 on X_2^* will be inconsistent when the error terms in equation (2.3) and equation (2.4) are correlated.

The conditional censored mean in this framework will be as follows

$$E[Y_2 \,|X_1^*, X_2^*, Y_1^*] = E\big[X_2^{*/}\beta_2 + \varepsilon_2 \,\big|X_1^{*/}\beta_1 + \varepsilon_1 \geq 0]$$

$$= X_2^{*/}\beta_2 + E[\varepsilon_2 \,|\varepsilon_1 \geq -X_1^{*/}\beta_1 \;]. \tag{2.5}$$

To obtain $E[\varepsilon_2 \,|\varepsilon_1 \geq -X_1^{*/}\beta_1 \,]$ Heckman (1979) assumed joint normality of the error terms and arrived at the following expression[12]:

$$E[Y_2\,|X_1^*, X_2^*, Y_1^*] = X_2^{*/}\beta_2 \;\; + \sigma_{12} \;\; \lambda(X_1^{*/}\beta_1),$$

where, $\lambda\left(X_1^{*/}\beta_1\right) = \phi(X_1^{*/}\beta_1)/\Phi(X_1^{*/}\beta_1)$, which is the Inverse Mill's Ratio[13] and σ_{12} is the covariance between the two error terms.

Hence, the estimating equation is given by

$$Y_2 = X_2^{*/}\beta_2 + \sigma_{12}\,\lambda(X_1^{*/}\beta_1) + u, \qquad (2.6)$$

Where, u is the error term.

The estimate of $\widehat{\beta_1}$ is obtained by a Probit regression and estimate of $\widehat{\beta_2}$ is obtained by applying OLS to equation (2.6).

In our context, first a Probit regression has been run to address the question whether a household will spend on children education ($w_i > 0$) or not ($w_i = 0$), w_i being the share of expenditure on i^{th} educational item for a given household. The same analysis is applicable for adult items also. The explanatory variables are education level of household head (*hdedu*)[14], household head's age (*hdage*), a dummy variable for categorising female headed households (*fehd= 1* if the household is female headed, *0* otherwise), a dummy variable to indicate whether any member of the household is a regular salary earner (*hsalary = 1* if yes, *0* otherwise), household size (*n*), logarithm of monthly per capita expenditure (*LnMPCE*), square of logarithm of monthly per capita expenditure (*Lsq*), proportions of household members belonging to the age groups 5-9 years (*Primary_m* and *Primary_f*), 10-16 years (*Secondary_m* and *Secondary_f*) and 17-29 years

[12] See section 16.5.3 in Cameron and Trivedi (2005) for a more detailed description.

[13] The inverse Mill's ratio takes account of the possible selection bias, arising out of a censored dependent variable, which causes a concentration of observations at zero values. Ignoring this leads to biased OLS estimates.

[14] Here education level of the household head is the years of schooling of the head.

(*High_m and High_f*). In order to calculate the π ratios, the following log-quadratic form of budget share equation for the Engel curve has been used:

$$w_i = \frac{p_i q_i}{x} = \alpha_i + \beta_{1i} \ln\left(\frac{x}{n}\right) + \beta_{2i}\left[\ln\left(\frac{x}{n}\right)\right]^2 + \eta_i \ln n + \sum_s \gamma_{is}\frac{n_s}{n} + \delta_i z + \mu_i \widehat{\lambda_\iota} + u_i \quad (2.7),$$

where, $\frac{x}{n}$ is per capita household expenditure, $\frac{n_s}{n}$ captures the demographic composition of a household. It is the fraction of household members belonging to s^{th} demographic group, $\ln n$ captures the effect of household size on the share of the expenditure on i^{th} item in education w_i, z is a vector of other demographic explanatory variables which includes household head's education, household head's occupation, $\widehat{\lambda_\iota}$ is the Inverse Mill's Ratio obtained from Probit regression.

Using equation (2.2) for the formulation of π ratio, we get the following expression,

$$\pi_{is} = \frac{\frac{1}{x}\{-\beta_{1i} - 2.\beta_{2i}\ln\left(\frac{x}{n}\right) + \gamma_{is} - \sum_{s\neq l}\gamma_{il}.n_l + \eta_i - n\mu_i\widehat{\beta_{1s}}\widehat{\lambda}_\iota[X_1^{/}\widehat{\beta_1} + \widehat{\lambda}_\iota]\}}{w_i + \beta_{1i} + 2.\beta_{2i}\ln\left(\frac{x}{n}\right) - x.\mu_i\widehat{\beta_{1x}}\widehat{\lambda}_\iota[X_1^{/}\widehat{\beta_1} + \widehat{\lambda}_\iota]}, \quad (2.8)$$[15]

where, the π ratios are obtained by substituting the estimated values of η_i, γ_{is} *and* γ_{il} for all $l \neq s$ from the second stage regression and w_i and $\frac{n_l}{n}$ are replaced by their sample means. $\hat{\beta}_{1s}$ is the estimated coefficient of s^{th} demographic group from the Probit model, $\hat{\beta}_{1x}$ is the estimated coefficient of per capita expenditure (*x/n*) obtained from the Probit model.

From equation (2.8) it is clear that the π ratios are functions of the estimated parameters and other household information; the corresponding standard errors are obtained using the Delta method. The *p*-values calculated from the standard errors obtained by the Delta method are used for hypothesis testing.[16]

---

[15]The derivation of the expression is given in Appendix II.

[16]A brief description of Delta method is given in Appendix II.

For testing gender discrimination, we need to test whether (i) the π ratio for a male student is greater than that of a female student for a given educational item that is exclusively for students, and (ii) the π ratio for a male student has a larger negative value than that of a female student for adult goods. The tests are:

(i) $H_0: \pi_{is_m} = \pi_{is_f}$ against $H_1: \pi_{is_m} > \pi_{is_f}$, i being an educational item, and

(ii) $H_0: \pi_{js_m} = \pi_{js_f}$ against $H_1: \pi_{js_m} < \pi_{js_f}$, j being an adult good,

where, s_m and s_f refer to male and female categories, respectively. Rejection of H_0 (in both cases) would imply pro-male gender bias in expenditure on educational items.

2.3 Results and Discussion

In this study, NSSO (National Sample Survey Office) 68th round consumption expenditure data collected during July 2011 to June 2012 have been used. The consumption expenditure data collected by NSSO consist of household level expenditure on food, clothing, recreation, education and other consumption items along with household demographic characteristics. A brief description of the explanatory variables is reported in Table 2.1a.[17] In West Bengal, 6310 households are surveyed of which 3566 are in the rural sector and 2744 are in the urban sector. Among them only 32.14% and 25.78% spend on education in the rural and the urban sector, respectively.[18] In Table 2.1b the percentages of zero values are reported for each educational

[17] It may be mentioned that this data set does not provide information on 'distance from school to residence'. So, although this would have been a very relevant explanatory variable, we are unable to include it in our analysis.

[18]The choice of items has been guided by the Majumder and Mitra (2017) study based on NSSO 64th round (2007-2008) individual level survey data exclusively on education expenditure. In NSSO consumption expenditure data the education group does not include transport cost, but we have included it in this study. Further, we have combined

item. It can be seen that the frequency of zero spending for educational items is more in the rural sector than that in the urban sector.

For each item mentioned in Appendix II, Table A2.2 along with three adult items mentioned before, the budget share equation is estimated using the Heckman two step estimation method, adjusting for sampling weights. The estimation is carried out in two stages – Probit and Ordinary Least Squares.

The results of estimation are reported in Table 2.2, Table 2.3 for educational items and Table 2.4, Table 2.5 for adult items for the rural and urban sectors, respectively.

For the Probit regression, the coefficients of *LnMPCE* for all items are positive and significant for both sectors and the coefficients of *Lsq* are negative and significant for both sectors which implies that the items – 'Books', 'Stationery and photocopy charges', 'Private coaching', 'Tuition fees', 'Transport cost', 'Pan', 'Tobacco' and 'Alcohol' are all normal goods.

The coefficients of *lnhhsize* ($\ln n$) are significant for both sectors, positive for all items except for the transport cost and negative, significant for all adult items. This can be explained using Kingdon (2005), where the result can be considered to be the consequence of economies of scale. That is, a higher household size leads to more sharing of goods and services, which allows the households to be able to spend more on educational items by cutting back expenditure on other goods like furniture, recreation etc. and adult items. This explanation becomes more plausible by looking at positive and significant coefficients of the proportion of household members belonging to 'primary', 'secondary' and 'high' education groups for both sexes for almost all other items. However, the coefficients of these fractions are negative, significant (except *Primary_m* and *Primary_f*) for adult items, leading to the explanation that the expenditure on

---

the expenditure on first-hand books and second-hand books as expenditure on books. The list of education items under the two datasets are given in Appendix II, Tables A2.1 and A2.2.

these items are exclusively for adults of the household, but that of '*High_m'* are positive, significant which implies that the male members of a household belonging to the age group of 17-29 years are likely to consume Tobacco.

For 'books' and 'transport cost', the coefficients are negative and significant implying that the explanation of economies of scale mentioned above is not applicable for these items. Hence, it is clear that the sharing of educational items is generally confined to the members of a household aged 5-29 years for both sexes for both sectors, whereas, that is not the case for the adult items in India. This justifies treating the children and young adults aged between 5-29 years as 'students' and the group above 30 years old as 'adults' with respect to educational items.

The coefficients of household head's education (*hdedu*) and those of the dummy variable *hsalary* (= 1 if any member of the household is a regular salary earner, 0 otherwise) are positive and significant for both sectors. The coefficients of the dummy variable *fehd* (whether the household is female headed or not) are non-significant in the rural sector for all educational items, but in the urban sector they are positive and significant for stationery and tuition fees and non-significant for books and private coaching. However, the coefficients of this dummy variable are negative, significant for adult items in both sectors. Thus, in the urban sector expenditure on education is taken as investment for welfare of the children, more so if the decision maker of the household is female and also households with female heads have a negative impact on spending on adult items. The coefficients of household head's age (*hdage*) are negative and significant for all items in the rural and urban sectors, except for 'books', where the coefficient is positive and significant. On the other hand, for adult items, the coefficients of *hdedu* and *hdage* are negative, significant in both sectors leading to the fact that households with older and wiser heads are less likely to spend on adult items.

For the second stage OLS regression, the pattern of signs of coefficients is generally the same as in the case of the Probit model for the common explanatory variables considered.

Here, household head's occupation has been incorporated using the variable 'household type', available in the NSSO data. This variable has been transformed into following dummies:[19]:

hdself = 1 if household head is *self-employed in agriculture/ non-agriculture, 0* otherwise*;*

hdwage = 1 if household head is *a regular wage/ salary earner, 0* otherwise, and

hdcasual =1 if household head is *a casual labourer in agriculture/ non-agriculture, 0* otherwise.

In the rural sector *hdself* turns out to be non-significant for all items, while *hdwage* and *hdcasual* turn out to be generally positive and significant. For the urban sector, on the other hand, all the dummy variables have positive and significant coefficients generally for all items, with few exceptions. However, *hdwage* turns out to be positive, significant for both sectors.

The Inverse Mill's Ratio and Wald statistics are significant in both sectors.

The estimated coefficients from the Heckman two step method are used to calculate the π ratios using equation (2.8) and are reported in Tables 2.6 and 2.7 for the rural and the urban sectors, respectively.

The pattern of the π ratios is discussed separately for each educational item considered.

For expenditure on 'books', the π ratios are positive and significant for age groups 5-9 years, 10-16 years and 17-29 years for both sexes and for both sectors. Hence, an increase in the number of members belonging to the age groups 5-29 years, who are expected to be involved with educational activities, leads to positive income effects for the expenses on books. Thus, books can be labeled as a 'student goods'.

[19]Note that in the OLS regression *fehd* and *fsalary* have been excluded and there is no constant term in the regression, as the three dummies for occupation are exhaustive.

For the expenditure on ‘Stationery and photocopy charges’, the π ratios are positive and significant for the age groups 10-16 years and 17-29 years for both sexes and for both sectors. For females in the age group 5-9 years the π ratio is positive and significant. For the rest of the age groups these are non-significant. Thus, ‘stationery and photocopy charges’ can be treated as a ‘student good’ with respect to the age groups 10-29 years.

Similar arguments can be made for expenditure on ‘tuition fees’ and ‘private coaching’ as the pattern of signs of π ratios are like those obtained from expenditure on ‘stationery and photocopy charges’ in both sectors for both sexes, although the values of the π ratios are different. So, ‘tuition fees’ and ‘private coaching’ can be treated as ‘student goods’.

In contrast, a completely opposite pattern of signs of π ratios are obtained for π ratios from expenditures on ‘transport costs’, ‘pan’, ‘tobacco’ and ‘alcohol’. The π ratios are negative and significant for the age groups 5-29 years. Therefore, ‘transport cost’, ‘pan’, ‘tobacco’ and ‘alcohol’ can be labelled as ‘adult good’. A possible explanation for ‘transport cost’ to be an adult good is that we do not have separate information on transport costs specifically for students, for commuting to educational institutes. Thus, an increase in the number of members belonging to the age group 5-29 years leads to a cut back in the expenditure on the adult items to finance their educational expenses.

To detect gender discrimination against female students, we have tested whether the π ratio for male members is greater than that for female members for each age group, for each educational item, and for both sectors. We have also tested the difference between the π ratios for males and females for each adult item using the values from Tables 2.6 and 2.7. The results are reported in Tables 2.8 and 2.9 for the rural and the urban sectors, respectively.

For expenditure on books, the π ratios for male are significantly greater than that for female for the age groups between 5-29 years in the rural sector and the urban sector. The results indicate discrimination against females. In the urban sector the magnitude is higher compared to the rural sector. The probable reason might be the difference in the supply side conditions of books in these two sectors, as in the urban sector the availability of books (both first hand and second hand) is better than that in the rural sector.

For expenditure on stationery and photocopy charges, the differences in the π ratios between males and females are positive and significant for the age group 10-29 years for both sectors (for 5-9 years, are positive, but non-significant for both sectors), establishing the fact that households discriminate between spending on male and female students for these educational items, with a pro male bias.

On 'tuition fees', the differences in the π ratios between male and female members are positive and significant for the age groups 5-29 in both sectors. For 'private tutor and coaching', the differences in π ratios are positive and significant for the age group 10-29 years in the rural sector and in the urban sector. Thus, both items exhibit pro male bias.

Gender discrimination (pro-male bias) is further supported in the age groups of 10-16 and 17-29 years by the difference in π ratios between males and females for adult items. While the differences are negative and significant for all items for these age groups, indicating greater sacrifice on adult goods for male students, for the age group 5- 9 years the differences are non-significant.

2.4 Conclusion

This chapter attempts to capture gender bias in expenditure on education based on the Engel curve approach and outlay equivalent ratios using household level consumer expenditure data. In the existing literatures, this approach has been used mainly to identify adult goods such as alcohol, tobacco, adult clothing etc. In this chapter we have tried to use the methodology in the context of education expenditure, which is likely to be incurred on students of a household (here 'students' include children as well as young adults). Moreover, since a considerable number of households report zero expenditure on educational items, we use Heckman's two-step procedure to estimate the Engel curves. In the event of the non-availability of individual level data, this method could be used to address the issue of gender bias.

We find that all educational items considered here, such as books, stationery and photocopy charges, tuition fees and private coaching fees, are found to be exclusively 'student' items. Transport cost, Pan, Tobacco and Alcohol are found to be adult goods. We find evidence of households cutting back expenditure on adult items to finance the educational expenses when there is an increase in the number of members belonging to the age group of 5-29 years. We also find evidence of pro male bias in expenditure on educational items and the extent of gender bias is more in the age groups of 10-16 years and 17-29 years in both sectors in West Bengal. This corroborates the findings of Kingdon (2005, 2013). A possible reason for absence of gender bias in the age group of 5-9 years is the low cost of education at this level. With higher level of education, the cost is expected to increase and hence gender bias tends to become prominent.

Table 2.1a: Descriptive Statistics of Variables

Variables	MEAN		STANDARD DEVIATION	
	Rural (3566)*	Urban (2744)*	Rural	Urban
LnMPCE	4.218	4.623	.484	.501
Lsq	15.343	16.102	.188	.291
Lnhhsize	.958	.783	.599	.607
Primary_m	.056	.039	.177	.166
Secondary_m	.132	.113	.286	.309
High_m	.195	.184	.351	.416
Primary_f	.048	.032	.159	.135
Secondary_f	.131	.100	.290	.295
High_f	.208	.175	.309	.351
Hdedu	5.289	7.384	3.382	3.589
Hdage	46.825	49.917	13.244	14.415

*****Figures in the parentheses are the number of sample households.

Table 2.1b: Percentages (%) of households reporting Zero expenditure in educational items

ITEMS	RURAL	URBAN
1. BOOKS	29.67	21.60
2. STATIONERY/PHOTOCOPY CHARGES	37.67	24.85
3. TUITION FEES	50.64	37.67
4. PRIVATE COACHING	50.98	32.37

Table 2.2: Heckman Two-step Estimation Results: (Rural)

Dependent Variable: Budget Share of Educational Items

EXPLANATORY VARIABLES	BOOKS		STATIONERY		TUITION		Pvt. COACHING		TRANSPORT	
	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS
LnMPCE	0.176 (0.001)***	0.597 (0.000)***	0.577 (0.005)***	0.013 (0.059)*	0.012 (0.000)***	0.819 (0.001)***	0.794 (0.044)**	0.025 (0.001)***	0.154 (0.001)***	0.045 (0.000)***
Lsq	-0.011 (0.001)***	-0.005 (0.000)***	-0.057 (0.000)***	-0.001 (0.011)**	-0.001 (0.014)**	-0.081 (0.000)***	-0.088 (0.015)**	-0.0004 (0.004)***	-0.037 (0.022)**	-0.004 (0.006)***
lnhhsize	0.157 (0.002)***	0.007 (0.005)***	0.638 (0.007)***	0.080 (0.000)***	0.001 (0.005)***	0.639 (0.015)**	0.541 (0.025)**	0.051 (0.035)**	-0.084 (0.019)**	-0.057 (0.002)***
Primary_m	0.393 (0.003)***	0.056 (0.001)***	0.315 (0.000)***	-0.0003 (0.204)	0.003 (0.006)***	0.057 (0.001)***	0.306 (0.002)***	0.045 (0.001)***	-0.315 (0.000)***	-0.006 (0.059)*
Secondary_m	0.668 (0.000)***	0. 025 (0.005)***	0.750 (0.003)***	0.001 (0.001)***	0.022 (0.012)**	0.551 (0.003)***	0.411 (0.015)**	0.062 (0.004)***	-0.052 (0.002)***	-0.027 (0.001)***
High_m	0.203 (0.048)**	0. 055 (0.001)***	0.140 (0.031)**	0.001 (0.007)***	0.027 (0.001)***	0.121 (0.041)*	0.079 (0.001)***	0.008 (0.031)**	0.436 (0.277)	-0.001 (0.002)***
Primary_f	0.061 (0.008)***	0.0002 (0.011)**	0.202 (0.000)***	0.0003 (0.228)	0.011 (0.003)***	0.288 (0.002)***	0.384 (0.001)***	0.012 (0.019)***	-0.256 (0.043)**	0.012 (0.064)*
Secondary_f	0.071 (0.003)***	0.002 (0.005)***	0.830 (0.034)**	0.015 (0.000)***	0.021 (0.025)**	0 .618 (0.022)***	0.434 (0.008)***	0.009 (0.001)***	-0.096 (0.029)**	0.008 (0.089)*
High_f	0.590 (0.015)**	0. 005 (0.003)***	0.395 (0.027)**	0.004 (0.028)**	0.002 (0.003)***	0.391 (0.049)**	0.303 (0.039)**	0.025 (0.013)**	-0.090 (0.019)**	0.018 (0.332)
Hdedu	0.008 (0.019)**	0.011 (0.009)***	0.056 (0.001)***	0.056 (0.000)***	0.040 (0.022)**	0.054 (0.034)**	0.024 (0.001)***	0.072 (0.001)***	0.002 (0.111)	0.0004 (0.205)
Hdage	-0.011 (0.011)**	-0.227 (0.022)**	-0.008 (0.001)***	-0.008 (0.002)***	-0.011 (0.004)***	-0.054 (0.013)**	-0.007 (0.009)***	-0.015 (0.011)**	-0.001 (0.059)*	0.0005 (0.339)
hsalary	0.163 (0.001)***		0.002 (0.000)***		0.152 (0.033)**		0.120 (0.000)***		0.116 (0.000)***	

Table 2.2 (Continued): Heckman Two-step Estimation Results: (Rural)

Dependent Variable: Budget Share of Educational Items

EXPLANATORY VARIABLES	BOOKS		STATIONERY		TUITION		Pvt. COACHING		TRANSPORT	
	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS
Fehd	-0.054 (0.091)*		0.032 (0.219)		0.171 (0.607)		-0.266 (0.098)*		-0.190 (0.211)	
Λ		0. 003 (0.000)***		0.056 (0.000)***		0.027 (0.000)***		0.003 (0.000)***		0.334 (0.000)***
constant	-0.908 (0.021)**		-1.276 (0.032)**		0.980 (0.046)**		3.098 (0.026)**		0.093 (0.000)***	
Hdself		0.003 (0.059)*		-0.087 (0.079)*		0.049 (0.098)*		0.003 (0.101)		0.019 (0.509)
hdwage		0.019 (0.033)**		0.001 (0.001)***		0.005 (0.016)**		-0.0001 (0.405)		0.001 (0.021)**
hdcasual		0.067 (0.001)***		-0.009 (0.122)		0.007 (0.000)***		0.056 (0.011)**		0.064 (0.003)***
WALD Statistics ***(χ^2_{17})***	730.42 (0.001)***		782.60 (0.000)***		652.44 (0.000)***		995.45 (0.000)***		887.94 (0.000)***	

*Figures in the parentheses are the p-values

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.3: Heckman Two-step Estimation Results: (Urban)

Dependent Variable: Budget Share of Educational Items

EXPLANATORY VARIABLES	BOOKS		STATIONERY		TUITION		Pvt. COACHING		TRANSPORT	
	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS
LnMPCE	0.373 (0.000)***	0.562 (0.001)***	0.018 (0.000)***	0.001 (0.000)***	0.280 (0.001)***	0.052 (0.000)***	0.172 (0.000)***	0.031 (0.002)***	0.270 (0.000)***	.0 685 (0.001)***
Lsq	-0.040 (0.000)***	-0.005 (0.000)***	-0.022 (0.018)**	-0.001 (0.021)**	-0.023 (0.030)**	-0.007 (0.001)***	-0.011 (0.016)**	-0.003 (0.047)**	-0.041 (0.029)**	-0.006 (0.008)***
lnhhsize	0. 432 (0.001)***	0.511 (0.002)***	0.111 (0.000)***	0.069 (0.007)***	0.226 (0.003)***	0.042 (0.005)***	0.161 (0.029)**	0.009 (0.021)**	-0.084 (0.001)***	-0.030 (0.027)**
Primary_m	0. 405 (0.011)**	0.012 (0.020)**	0.368 (0.033)**	-0.0005 (0.042)**	0.193 (0.002)***	0.005 (0.000)***	0.024 (0.048)**	0.002 (0.398)	-0.186 (0.004)***	-0.012 (0.007)***
Secondary_m	0.314 (0.001)***	0.002 (0.029)**	0.961 (0.016)**	0.046 (0.005)***	0.215 (0.009)***	0.031 (0.002)***	0.286 (0.022)**	0.069 (0.004)***	-0.144 (0.019)**	-0.004 (0.006)***
High_m	0. 090 (0.011)**	0.169 (0.008)***	0.170 (0.004)***	0.011 (0.003)***	0.142 (0.032)**	0.023 (0.004)***	0.055 (0.031)**	0.054 (0.002)***	-0.144 (0.002)***	-0.015 (0.002)***
Primary_f	0. 671 (0.005)***	0.012 (0.003)***	0.151 (0.001)***	-0.0006 (0.209)	0.360 (0.019)**	0.039 (0.017)**	0. 459 (0.015)**	0.004 (0.003)***	-0.489 (0.013)**	-0.126 (0.044)**
Secondary_f	0. 644 (0.001)***	0.026 (0.004)***	0.190 (0.019)**	0.078 (0.020)**	0.146 (0.004)***	0.030 (0.001)***	0.160 (0.005)***	0.009 (0.006)***	-0.106 (0.001)***	-0.017 (0.004)***
High_f	0. 330 (0.008)***	0.089 (0.019)**	0.350 (0.002)***	0.065 (0.017)**	0.546 (0.022)**	0.019 (0.002)***	0.476 (0.000)***	0.001 (0.000)***	-0.304 (0.003)***	-0.034 (0.001)***
Hdedu	0.029 (0.004)***	0.077 (0.008)***	0.074 (0.013)**	0.088 (0.002)***	0.066 (0.001)***	0.031 (0.038)**	0.034 (0.004)***	0.0002 (0.029)**	0.029 (0.005)***	0.0005 (0.541)
Hdage	0. 004 (0.013)**	-0.035 (0.001)***	-0.003 (0.000)***	-0.009 (0.002)***	-0.007 (0.041)**	-0.007 (0.001)***	-0.005 (0.007)***	-0.0003 (0.011)**	0.005 (0.001)***	0.0001 (0.711)
hsalary	0.181 (0.039)**		0.001 (0.044)**		0.010 (0.002)***		0.032 (0.026)***		0.002 (0.018)**	

Table 2.3 (Continued): Heckman Two-step Estimation Results: (Urban)

Dependent Variable: Budget Share of Educational Items

Fehd	-0.017 (0.098)*		0.074 (0.0028)**		0.097 (0.001)***		-0.101 (0.197)		-0.153 (0.305)	
Λ		0.110 (0.001)***		0.024 (0.000)***		0.129 (0.025)***		0.078 (0.002)***		0.314 (0.003)***
Constant	-6.57 (0.000)***		-1.02 (0.000)***		2.067 (0.000)***		0.379 (0.008)***		1.153 (0.000)***	
Hdself		0.0004 (0.048)**		0.003 (0.001)***		0.041 (0.001)***		0.0005 (0.214)		0.026 (0.001)***
hdwage		0.102 (0.003)***		0.809 (0.040)**		0.907 (0.027)**		0.0003 (0.488)		0.078 (0.024)**
hdcasual		0.0002 (0.111)		0.007 (0.302)		0.187 (0.011)**		0.007 (0.001)***		0.0007 (0.001)***
WALD Statistics (χ^2_{17})	818.73 (0.000)***		778.83 (0.001)***		619.32 (0.001)***		666.03 (0.006)***		701.96 (0.001)***	

*Figures in the parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.4: Heckman Two-step Estimation Results: (Rural)

Dependent Variable: Budget Share of Adult Items

EXPLANATORY VARIABLES	Pan		Tobacco		Alcohol	
	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS
LnMPCE	0.187 (0.000)***	0.507 (0.002)***	0.011 (0.000)***	0.108 (0.011)**	0.455 (0.001)**	0.035 (0.000)***
Lsq	-0.095 (0.001)***	-0.028 (0.019)**	-0.001 (0.004)***	-0.121 (0.000)***	-0.020 (0.001)***	-0.081 (0.006)***
Lnhhsize	-0.046 (0.002)***	-0.052 (0.001)***	-0.062 (0.006)***	-0.095 (0.004)***	-0.040 (0.005)***	-0.016 (0.044)**
Primary_m	-0.032 (0.007)***	0.087 (0.351)	-0.023 (0.002)***	-0.129 (0.122)	-0.089 (0.001)***	0.108 (0.125)
Secondary_m	-0.135 (0.009)***	-0.005 (0.004)***	-0.033 (0.001)***	-0.035 (0.018)**	0.066 (0.015)**	-0.209 (0.008)***
High_m	0.424 (0.113)	0.015 (0.188)	0.025 (0.003)***	0.024 (0.007)***	0.133 (0.034)**	0.589 (0.002)***
Primary_f	0.183 (0.221)	0.070 (0.291)	-0.011 (0.128)	-0.238 (0.199)	0.188 (0.608)	0.079 (0.172)
Secondary_f	-0.031 (0.002)***	-0.049 (0.004)***	-0.039 (0.018)***	-0.039 (0.003)***	-0.057 (0.002)***	-0.530 (0.006)***
High_f	0.024 (0.506)	-0. 011 (0.025)**	-0.007 (0.119)	-0.069 (0.000)***	-0. 084 (0.753)	-0.073 (0.009)***
Hdedu	-0.118 (0.021)**	0.126 (0.276)	-0.331 (0.041)**	-0.212 (0.001)***	-0.073 (0.004)***	-0.140 (0.003)***
Hdage	0.012 (0.002)***	0.052 (0.338)	-0.008 (0.039)**	-0.025 (0.005)***	-0.158 (0.298)	-0.035 (0.037)**
Hsalary	0.024 (0.003)***		0.082 (0.003)***		0.152 (0.003)***	
Fehd	-0.146 (0.008)***		-1.148 (0.001)***		-0. 828 (0.002)***	
λ		0. 114 (0.001)***		0.019 (0.021)**		0.014 (0.005)***

Table 2.4 (continued): Heckman Two-step Estimation Results: (Rural)

Dependent Variable: Budget Share of Adult Items

EXPLANATORY VARIABLES	Pan		Tobacco		Alcohol	
	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS
Constant	-1.918 (0.000)***		-1.551 (0.013)**		-0.907 (0.005)***	
Hdself		0.109 (0.011)**		-0.003 (0.138)		0.003 (0.328)
Hdwage		0.085 (0.031)**		0.052 (0.001)***		0.065 (0.005)***
Hdcasual		0.015 (0.209)		-0.019 (0.002)***		0.027 (0.455)
WALD Statistics (χ^2_{17})	681.02 (0.000)***		563.92 (0.000)***		446.44 (0.003)***	

*Figures in the parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.5: Heckman Two-step Estimation Results: (Urban)

Dependent Variable: Budget Share of Adult Items

EXPLANATORY VARIABLES	Pan		Tobacco		Alcohol	
	PROBIT	OLS	PROBIT	OLS	PROBIT	OLS
LnMPCE	0.255 (0.000)***	0.183 (0.000)***	0.403 (0.000)***	0.042 (0.001)***	0.063 (0.000)***	0.209 (0.001)***
Lsq	-0.142 (0.007)***	-0. 105 (0.013)**	-0.079 (0.011)**	-0.051 (0.016)**	-0.212 (0.023)**	-0.011 (0.003)***
Lnhhsize	-0.017 (0.000)***	-0.067 (0.402)	-0.076 (0.003)***	-0.048 (0.001)***	-0.047 (0.005)***	-0.088 (0.002)***
Primary_m	-0 .015 (0.126)	0.087 (0.384)	0.307 (0.233)	-0.054 (0.098)*	0.041 (0.081)*	0.652 (0.087)*
Secondary_m	-0.039 (0.004)***	-0.005 (0.019)**	-0.602 (0.007)***	-0.397 (0.001)***	-0.022 (0.001)***	-0.009 (0.001)***
High_m	-0.103 (0.144)	0.015 (0.109)	0.206 (0.028)**	0.275 (0.033)**	0.016 (0.008)***	0.021 (0.015)***
Primary_f	0.119 (0.122)	0.070 (0.239)	-0.107 (0.268)	-0.069 (0.492)	0.430 (0.409)	0.067 (0.391)
Secondary_f	-0.138 (0.005)***	-0.049 (0.003)***	-0.802 (0.014)**	-0.423 (0.019)***	-0.145 (0.078)*	-0.020 (0.004)***
High_f	-0.019 (0.002)***	-0. 011 (0.024)**	-0.081 (0.004)***	-0. 495 (0.004)***	-0.034 (0.577)	-0.696 (0.001)***
Hdedu	-0.041 (0.005)***	0.126 (0.133)	-0.051 (0.018)***	-0.029 (0.008)***	-0.072 (0.013)***	-0.059 (0.003)***
Hdage	0.071 (0.004)***	0.052 (0.287)	-0.023 (0.311)	-0.011 (0.002)***	-0.007 (0.002)***	-0.107 (0.084)*
Hsalary	0.021 (0.003)***		0.082 (0.001)***		0.201 (0.001)***	
Fehd	-0.525 (0.017)**		-0.425 (0.046)**		-0.436 (0.002)***	
Λ		0. 114 (0.003)***		0.057 (0.002)***		0.085 (0.013)***
Constant	-1.755 (0.000)***		-1.664 (0.000)***		-0.903 (0.000)***	
Hdself		0.081 (0.013)**		-0.112 (0.704)		0.205 (0.0168)
Hdwage		0.125 (0.004)***		0.102 (0.004)***		0.577 (0.002)***
Hdcasual		0.015 (0.183)		-0.209 (0.126)		0.023 (0.311)
WALD Statistics (χ^2_{17})	494.87 (0.000)***		683.33 (0.000)***		511.66 (0.002)***	

*Figures in the parentheses are the p-values

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.6: The π Ratios: Rural Sector

ITEMS	PRIMARY (5-9 years)		SECONDARY (10-16 years)		HIGH (17-29 years)	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Books	0.29 (0.000)***	0.17 (0.002)***	0.16 (0.000)***	0.08 (0.000)***	0.21 (0.000)***	0.11 (0.009)***
Stationery, photocopying charges	0.17 (0.887)	0.13 (0.000)***	0.38 (0.001)***	0.19 (0.001)***	0.32 (0.000)***	0.21 (0.000)***
Tuition and other fees (school, college, etc.)	0.25 (0.000)***	0.11 (0.000)***	0.45 (0.000)***	0.32 (0.000)***	0.42 (0.000)***	0.24 (0.000)***
Private tutor/ Coaching centre	-0.23 (0.768)	-0.17 (0.111)	0.43 (0.002)***	0.27 (0.001)***	0.67 (0.000)***	0.52 (0.001)***
Transport cost	-0.56 (0.021)**	-0.18 (0.601)	-0.34 (0.000)***	-0.67 (0.000)***	-0.10 (0.002)***	-0.19 (0.001)***
Pan	-0.23 (0.506)	-0.45 (0.112)	-0.57 (0.002)***	-0.17 (0.011)**	-0.45 (0.002)***	-0.32 (0.031)***
Tobacco	-0.34 (0.709)	-0.49 (0.203)	-0.14 (0.004)***	-0.08 (0.012)**	-0.64 (0.023)**	-0.28 (0.027)**
Alcohol	-0.48 (0.119)	-0.89 (0.406)	-0.16 (0.014)**	-0.09 (0.067)*	-0.57 (0.002)***	-0.42 (0.001)***

*Figures in the parentheses are the *p*-values for testing $\pi = 0$.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.7: The π Ratios: Urban Sector

ITEMS	PRIMARY (5-9 years)		SECONDARY (10-16 years)		HIGH (17-29 years)	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Books	0.34 (0.000)***	0.17 (0.000)***	0.15 (0.000)***	0.14 (0.001)***	0.19 (0.000)***	0.02 (0.000)***
Stationery, photocopying charges	0.23 (0.101)	0.18 (0.789)	0.16 (0.000)***	0.07 (0.000)***	0.27 (0.006)***	0.19 (0.000)***
Tuition and other fees (school, college, etc.)	0.17 (0.000)***	0.03 (0.000)***	0.22 (0.000)***	0.19 (0.000)***	0.24 (0.000)***	0.15 (0.000)***
Private tutor/ coaching centre	0.09 (0.000)***	0.07 (0.321)	0.45 (0.000)***	0.32 (0.011)**	0.37 (0.000)***	0.23 (0.021)**
Transport cost	-0.13 (0.000)***	-0.19 (0.000)***	-0.36 (0.000)***	-0.25 (0.040)**	-0.28 (0.000)***	-0.36 (0.047)***
Pan	-0.36 (0.307)	-0.28 (0.432)	-0.38 (0.001)***	-0.32 (0.002)**	-0.39 (0.015)**	-0.18 (0.016)**
Tobacco	-0.59 (0.118)	-0.38 (0.309)	-0.19 (0.005)***	-0.09 (0.019)**	-0.33 (0.056)*	-0.29 (0.028)**
Alcohol	-0.79 (0.109)	-0.41 (0.705)	-0.11 (0.003)***	-0.05 (0.016)**	-0.31 (0.001)***	-0.15 (0.021)**

*Figures in the parentheses are the p-values of the test statistic π=0.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.8: Difference in π Ratios ($male - female$) for Gender Discrimination Test: Rural Sector

Items / Education level	Student items				Adult Items		
	Books	**Stationery, photocopying charges**	**Tuition and other fees (school, college, etc.)**	**Private tutor/ Coaching centre**	**Pan**	**Tobacco**	**Alcohol**
PRIMARY (5-9 years)	0.12 (0.000)***	0.04 (0.111)	0.14 (0.011)**	-0.06 (0.225)	0.22 (0.327)	0.15 (0.486)	0.41 (0.509)
SECONDARY (10-16 years)	0.08 0.000)***	0.19 (0.000)***	0.13 (0.000)***	0.16 (0.009)***	-0.40 (0.047)**	-0.06 (0.034)**	-0.07 (0.015)**
HIGH (17-29 years)	0.10 (0.000)***	0.11 (0.000)***	0.18 (0.001)***	0.15 (0.001)***	-0.13 (0.067)*	-0.23 (0.043)**	-0.15 (0.021)**

*Figures in the parentheses are the p-values for testing $H_0: \pi_{is_m} = \pi_{is_f}$ against $H_1: \pi_{is_m} > \pi_{is_f}$ for student goods and $H_0: \pi_{js_m} = \pi_{js_f}$ against $H_1: \pi_{js_m} < \pi_{js_f}$ for adult goods.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 2.9: Difference in π Ratios ($male - female$) for Gender Discrimination Test: Urban Sector

Items / Education level	Student Items				Adult Items		
	Books	**Stationery, photocopying charges**	**Tuition and other fees (school, college, etc.)**	**Private tutor/ Coaching centre**	**Pan**	**Tobacco**	**Alcohol**
PRIMARY (5-9 years)	0.17 (0.000)***	0.05 (0.206)	0.14 (0.001)***	0.02 (0.097)*	-0.08 (0.109)	-0.21 (0.156)	-0.38 (0.291)
SECONDARY (10-16 years)	0.01 (0.000)***	0.09 (0.011)**	0.03 (0.000)***	0.13 (0.000)***	-0.06 (0.011)**	-0.10 (0.031)**	-0.06 (0.002)***
HIGH (17-29 years)	0.17 (0.000)***	0.08 (0.000)***	0.08 (0.000)***	0.14 (0.005)***	-0.21 (0.003)***	-0.07 (0.027)**	-0.16 (0.041)**

*Figures in the parentheses are the p-values for testing $H_0: \pi_{is_m} = \pi_{is_f}$ against $H_1: \pi_{is_m} > \pi_{is_f}$ for student goods and $H_0: \pi_{js_m} = \pi_{js_f}$ against $H_1: \pi_{js_m} < \pi_{js_f}$ for adult goods.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

APPENDIX 1I

Derivation of Outlay Equivalent Ratio

The expression of π_{is} is as follows

$$\pi_{is} = \frac{\frac{\partial(p_i q_i)}{\partial n_s}}{\frac{\partial(p_i q_i)}{\partial x}} \div \frac{x}{n}$$

Now,

$$w_i = \frac{p_i q_i}{x} = \alpha_i + \beta_{1i}\ln\left(\frac{x}{n}\right) + \beta_{2i}\left[\ln\left(\frac{x}{n}\right)\right]^2 + \eta_i \ln n + \sum_s \gamma_{is}\frac{n_s}{n} + \delta_i z + \mu_i\widehat{\lambda_\iota} + u_i$$

where, $\widehat{\lambda_\iota} = \frac{\phi(X_1^{/}\widehat{\beta_1})}{\Phi(X_1^{/}\widehat{\beta_1})}$

For, the denominator

$$\frac{\partial(p_i q_i)}{\partial x} = w_i + x.\frac{\partial w_i}{\partial x}$$

$$\frac{\partial w_i}{\partial x} = \frac{\beta_{1i}}{x} + \frac{2.\beta_{2i}\ln\left(\frac{x}{n}\right)}{x} - \mu_i\widehat{\beta_{1x}}\widehat{\lambda_\iota}[X_1^{/}\widehat{\beta_1} + \widehat{\lambda_\iota}]$$

$$\frac{\partial(p_i q_i)}{\partial x} = w_i + \beta_{1i} + 2.\beta_{2i}\ln\left(\frac{x}{n}\right) - x.\mu_i\widehat{\beta_{1x}}\widehat{\lambda_\iota}[X_1^{/}\widehat{\beta_1} + \widehat{\lambda_\iota}]$$

Where, $\widehat{\beta_{1x}}$ is the estimated coefficient of log of monthly per capita expenditure obtained from Probit regression.

For the numerator,

$$\frac{\partial(p_i q_i)}{\partial n_s} \div \frac{x}{n} = \frac{n}{x}\{-\frac{\beta_{1i}}{n} - \frac{2.\beta_{2i}\ln\left(\frac{x}{n}\right)}{n} + \frac{\gamma_{is}}{n} - \sum_{s\neq l}\gamma_{il}\frac{n_l}{n} + \frac{\eta_i}{n} - \mu_i\widehat{\beta_{1s}}\widehat{\lambda_i}[X_1^{/}\widehat{\beta_1} + \widehat{\lambda_i}]\}$$

Hence, $\frac{\partial(p_i q_i)}{\partial n_s} \div \frac{x}{n} = \frac{1}{x}\{-\beta_{1i} - 2.\beta_{2i}\ln\left(\frac{x}{n}\right) + \gamma_{is} - \sum_{s\neq l}\gamma_{il}.n_l + \eta_i - n\mu_i\widehat{\beta_{1s}}\widehat{\lambda_i}[X_1^{/}\widehat{\beta_1} + \widehat{\lambda_i}]\}$

Where, $\widehat{\beta_{1s}}$ is the estimated coefficient of the fraction of household members belonging to s[th] demographic group obtained from the Probit regression.

Hence, $\pi_{is} = \frac{\frac{1}{x}\{-\beta_{1i} - 2.\beta_{2i}\ln\left(\frac{x}{n}\right) + \gamma_{is} - \sum_{s\neq l}\gamma_{il}.n_l + \eta_i - n\mu_i\widehat{\beta_{1s}}\widehat{\lambda_\iota}[X_1^{/}\widehat{\beta_1} + \widehat{\lambda_\iota}]\}}{w_i + \beta_{1i} + 2.\beta_{2i}\ln\left(\frac{x}{n}\right) - x.\mu_i\widehat{\beta_{1x}}\widehat{\lambda_\iota}[X_1^{/}\widehat{\beta_1} + \widehat{\lambda_\iota}]}$

The Delta Method

Let $T=(T_1,T_2,\dots T_k)$ be random variables with means $(\theta_1,\theta_2,....,\theta_k)$. Suppose there is a differentiable function $g(T)$ for which the estimation of variance is to be obtained. Define the partial derivatives as

$$g'_i(\theta)=\left.\frac{\partial g(T)}{\partial T_i}\right|_{T_1=\theta_1 \dots T_k=\theta_k}$$

Taking first order Taylor series approximation, we get

$$g(T) \approx g(\theta)+\sum_{i=1}^{k} g'_i(\theta)(T_i-\theta_i)$$

The expectation and the variance are the following

$$Eg(T) \approx g(\theta)+\sum_{i=1}^{k} g'_i(\theta)E(T_i-\theta_i)$$

$$= g(\theta)$$

The variance of $g(T)$ is as follows

Var [g(T)]

$$\approx E\left[(g(T)-g(\theta))^2\right]$$

$$\approx E\left[\left(\sum_{i=1}^{k} g'_i(\theta)(T_i-\theta_i)\right)^2\right]$$

$$=\sum_{i=1}^{k} g'_i(\theta)^2 VarT_i + 2\sum_{i>j} g'_i(\theta)g'_j(\theta)Cov(T_i,T_j)$$

Where, the last equality comes from expanding the square and using the definition of variance-covariance.

Using this concept, we get the expression of variance of π ratios,

$Variance\ (\pi_{is})$

$$=\frac{1}{x^2}\ [\left(\frac{\partial\pi_{is}}{\partial\eta_i}\right)^2.var(\eta_i)+\left(\frac{\partial\pi_{is}}{\partial\gamma_{is}}\right)^2.var(\gamma_{is})+\left(\frac{\partial\pi_{is}}{\partial\gamma_{ij}}\right)^2.var(\gamma_{ij})+\left(\frac{\partial\pi_{is}}{\partial\beta_{1i}}\right)^2.var(\beta_{1i})+$$

$$\left(\frac{\partial\pi_{is}}{\partial\beta_{2i}}\right)^2.var(\beta_{2i})+\left(\frac{\partial\pi_{is}}{\partial\mu_i}\right)^2.var(\mu_i)+\left(\frac{\partial\pi_{is}}{\partial\beta_{1s}}\right)^2.var(\beta_{1s})+\left(\frac{\partial\pi_{is}}{\partial\hat{\beta}_1}\right)^2.var(\hat{\beta}_1)+$$

$$\left(\frac{\partial\pi_{is}}{\partial\beta_{1x}}\right)^2.var(\beta_{1x})+2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\gamma_{is}}Cov(\eta_i,\gamma_{is})+2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\gamma_{ij}}Cov(\eta_i,\gamma_{ij})+$$

$$2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\beta_{1i}}Cov(\eta_i,\beta_{1i})+2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\beta_{2i}}Cov(\eta_i,\beta_{2i})+2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\mu_i}Cov(\eta_i,\mu_i)+$$

$$2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\beta_{1s}}Cov(\eta_i,\beta_{1s})+2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_1}Cov(\eta_i,\hat{\beta}_1)+2\frac{\partial\pi_{is}}{\partial\eta_i}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\eta_i,\beta_{1x})+$$

$2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\gamma_{ij}}Cov(\gamma_{is},\gamma_{ij}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\beta_{1i}}Cov(\gamma_{is},\beta_{1i}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\beta_{2i}}Cov(\gamma_{is},\beta_{2i}) +$

$2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\mu_{i}}Cov(\gamma_{is},\mu_{i}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\beta_{1s}}Cov(\gamma_{is},\beta_{1s}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Cov(\gamma_{is},\hat{\beta}_{1}) +$

$2\frac{\partial\pi_{is}}{\partial\gamma_{is}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Cov(\gamma_{is},\hat{\beta}_{1}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{ij}}.\frac{\partial\pi_{is}}{\partial\beta_{1i}}Cov(\gamma_{ij},\beta_{1i}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{ij}}.\frac{\partial\pi_{is}}{\partial\beta_{2i}}Cov(\gamma_{ij},\beta_{2i}) +$

$2\frac{\partial\pi_{is}}{\partial\gamma_{ij}}.\frac{\partial\pi_{is}}{\partial\mu_{i}}Cov(\gamma_{ij},\mu_{i}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{ij}}.\frac{\partial\pi_{is}}{\partial\beta_{1s}}Cov(\gamma_{ij},\beta_{1s}) + 2\frac{\partial\pi_{is}}{\partial\gamma_{ij}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Cov(\gamma_{ij},\hat{\beta}_{1}) +$

$2\frac{\partial\pi_{is}}{\partial\gamma_{ij}}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\gamma_{ij},\beta_{1x}) + 2.\frac{\partial\pi_{is}}{\partial\beta_{1i}}.\frac{\partial\pi_{is}}{\partial\beta_{2i}}Cov(\beta_{1i},\beta_{2i}) + 2\frac{\partial\pi_{is}}{\partial\beta_{1i}}.\frac{\partial\pi_{is}}{\partial\mu_{i}}Cov(\beta_{1i},\mu_{i}) +$

$2.\frac{\partial\pi_{is}}{\partial\beta_{1i}}.\frac{\partial\pi_{is}}{\partial\beta_{1s}}Cov(\beta_{1i},\beta_{1s}) + 2\frac{\partial\pi_{is}}{\partial\beta_{1i}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Co + 2\frac{\partial\pi_{is}}{\partial\beta_{1i}}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\beta_{1i},\beta_{1x}) +$

$2\frac{\partial\pi_{is}}{\partial\beta_{2i}}.\frac{\partial\pi_{is}}{\partial\mu_{i}}Cov(\beta_{2i},\mu_{i}) + 2\frac{\partial\pi_{is}}{\partial\beta_{2i}}.\frac{\partial\pi_{is}}{\partial\beta_{1s}}Cov(\beta_{2i},\beta_{1s}) + 2\frac{\partial\pi_{is}}{\partial\beta_{2i}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Cov(\widehat{\beta_{2i},\beta_{1}}) +$

$2\frac{\partial\pi_{is}}{\partial\beta_{2i}}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\beta_{2i},\beta_{1x}) + 2\frac{\partial\pi_{is}}{\partial\mu_{i}}.\frac{\partial\pi_{is}}{\partial\beta_{1s}}Cov(\mu_{i},\beta_{1s}) + 2\frac{\partial\pi_{is}}{\partial\mu_{i}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Cov(\mu_{i},\hat{\beta}_{1}) +$

$2\frac{\partial\pi_{is}}{\partial\mu_{i}}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\mu_{i},\beta_{1x}) + 2\frac{\partial\pi_{is}}{\partial\beta_{1s}}.\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}Cov(\beta_{1s},\hat{\beta}_{1}) +$

$2\frac{\partial\pi_{is}}{\partial\beta_{1s}}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\beta_{1s},\beta_{1x}) + 2\frac{\partial\pi_{is}}{\partial\hat{\beta}_{1}}.\frac{\partial\pi_{is}}{\partial\beta_{1x}}Cov(\hat{\beta}_{1},\beta_{1x})$

For the sake simplicity in the expressions of the partials of π ratios we have assumed that

$$\pi_{is} = \frac{A}{B},$$

where A and B are the numerator and the denominator of the expression of π_{is} obtained from equation (2.6).

Using this we get the following partials,

$$\frac{\partial\pi_{is}}{\partial\eta_i} = \frac{1}{Bx}$$

$$\frac{\partial\pi_{is}}{\partial\gamma_{is}} = \frac{1}{Bx}$$

$$\frac{\partial\pi_{is}}{\partial\gamma_{ij}} = \frac{-\frac{n_j}{n}}{Bx}$$

$$\frac{\partial\pi_{is}}{\partial\beta_{1i}} = \frac{-(A+B)^2}{B^2x}$$

$$\frac{\partial\pi_{is}}{\partial\beta_{2i}} = 2\ln\left(\frac{x}{n}\right)\frac{(A-B)^2}{B^2x}$$

$$\frac{\partial \pi_{is}}{\partial \mu_i} = \frac{\lambda(X_1^{*/}\hat{\beta}_1)\left[\lambda(X_1^{*}\hat{\beta}_1\right) + X_1^{*}\hat{\beta}_1](A.x.\widehat{\beta_{1x}} - B.n.\widehat{\beta_{1n})}}{B^2 x}$$

$$\frac{\partial \pi_{is}}{\partial \widehat{\beta_{1x}}} = \frac{-\mu_i.\lambda(X_1^{*/}\hat{\beta}_1)\left[\lambda(X_1^{*}\hat{\beta}_1\right) + X_1^{*}\hat{\beta}_1]}{B^2 x^2}$$

$$\frac{\partial \pi_{is}}{\partial \widehat{\beta_{1n}}} = \frac{-n\mu_i\lambda(X_1^{*/}\hat{\beta}_1)\left[\lambda(X_1^{*}\hat{\beta}_1\right) + X_1^{*}\hat{\beta}_1]}{Bx}$$

$$\frac{\partial \pi_{is}}{\partial \widehat{\beta_1}} = \frac{\lambda(X_1^{*/}\hat{\beta}_1)\left[\lambda(X_1^{*}\hat{\beta}_1\right)\{\lambda(X_1^{*/}\hat{\beta}_1) + X_1^{*}\hat{\beta}_1\}^2 + X_1 + \lambda(X_1^{*/}\hat{\beta}_1)\left[\lambda(X_1^{*}\hat{\beta}_1\right) + X_1^{*}\hat{\beta}_1](n.\mu_i\widehat{\beta_{1n}}.B - x.\mu_i.\widehat{\beta_{1x}}A)}{B^2 x^2}$$

Table A2.1: Items included under 'Education' by NSSO Consumption Expenditure Data

1. BOOKS FIRST HAND
2. BOOKS SECOND HAND
3. NEWSPAPER AND PERIODICALS
4. STATIONERY AND PHOTOCOPY CHARGES
5. TUITION FEES
6. PRIVATE COACHING
7. OTHER EXPENDITURE WHICH INCLUDES FEES FOR WEB BASED TRAINING.

Table A2.2: Items included under 'Education' by NSSO Education Expenditure Data <u>(Majumder and Mitra (2017))</u>

1. BOOKS
2. STATIONERY AND PHOTOCOPY CHARGES
3. TUITION FEES
4. PRIVATE COACHING
5. TRANSPORT COST

Table A2.3: Age Group Categorisation

EDUCATION LEVELS	AGE GROUPS (years)
PRIMARY	5-9
SECONDARY	10-16
HIGH	17-29

Chapter 3

Gender Bias in Below Class-10 and Above Class 10 Levels: Application of Heckman's Two Step Model and Multinomial Logit Model using Individual Level Education Expenditure Data

3.1 Introduction

Studies on children education generally focus on child schooling involving children typically in the age group of below 16 years. We contribute to the existing literature by extending the scope to young adults pursuing higher studies. First we consider two different levels of the education system in India, viz., below class-10 level (involving school going children) and above class-10 level. These two levels are different, because at below class-10 level, the decision of a household involves two stages: first, whether to send the child to school (or similar educational institution) or not and second, given that the child will be sent to an educational institute, the amount to be spent on the child for education [Kingdon (2005, 2013); Deaton (1997)]. On the other hand, in the above class 10 levels, the households' decision involves the issue of choice of specialization (e.g., Science, Arts, and Commerce). Given the fact that higher studies call for higher costs and inter-specialization differences in cost are rather high, gender bias may have some role in the choice of specializations. Hence, it is important to study the nature of gender bias at this level separately. According to the 64^{th} round (2007-2008) data on education expenditure by NSSO, out of a total of 8524 households, about 16% households report zero expenditure on education (dropout) at the secondary level (below Class 10) whereas the percentage of households in the higher secondary level is only about 2%. Within the below class 10 group, we differentiate between students belonging to classes 1-8 and students belonging to classes 9-10. The reason is the following: according to the Right to Education (RTE) act (2005), every child from 6-14 years should be provided free elementary schooling by Central and State government bodies. The government education bodies will not charge any fee like tuition fee, capitation fee etc. from the students belonging to this age group. However, the facility is with respect to exemption of tuition fees in schools for students aged 6-14 years. For sending a child to school, a household has to incur other related costs. Our study attempts to find the reason why 16% households (which include 6% households with children of 6-14 years age group) do not want to spend this additional amount on education, or in other words, why a household decides not to send the child to school at all. However, since, the decision of a household will still involve two stages described above; we have used two separate, but similar models for these two different sections, i.e., class 1-8 and class 9-10 in the below class-10 level. The two-stage analysis for the below class-10 level has been done using Heckman's (1979) two step method.

To capture gender bias in inter-stream and intra-stream allocation of educational expenditure for the above class-10 level students, we follow the approach used by Brown et al. (1980). This approach enables us to assess the relative importance of within and between subject expenditure allocation effects.[20] The model specification is based on the wage differential model that was used by Liu, Zhang and Chong (2004) to capture inter occupational and intra occupational wage differential between natives and immigrants using data of Hong Kong. In this paper, we focus our attention on the education scenario of the rural and urban sectors of the state of West Bengal.

The plan of the chapter is as follows. In Section 3.2, the models used in this chapter i.e., Heckman's two step model (Section 3.2.1) and wage differential model (Section 3.2.2) have been discussed. Section 3.3 presents the results and discussion with a brief description of the data

[20]The Brown et al. (1980) approach has been applied in many studies of wage differentials for various classifications of the population, namely, by gender in Dolton and Kidd (1994), Kidd and Shannon (1994, 1996) and Meng and Miller (1995), by race in Gabriel and Schmitz (1989), Moll (1991), by immigration status in Liu, Zhang and Chong (2004) and by Indian social group in Madheswaran and Attewell (2007).

used (Section 3.3.1) and the methodology of the estimation procedure (Section 3.3.2). The chapter concludes in Section 3.4.

3.2 Model specification

The two levels of education i.e., below class-10 and above class-10 level, are analyzed using two different models. The first level (below class-10) has been modelled by a sample selection model, as at this level the parents first decide whether to send the child to any educational institute or not, and then given that the child will be sent to an educational institute, they decide the share of the monthly spending of a household to be spent on the child. Within the first level, we make a distinction between classes 1-8 and classes 9-10. This, as mentioned earlier, is because of the RTE act. However, since the basic model is same for both these sections, we describe the model only once in Section 3.2.1.

The second level (above class 10 level) involves selection of a subject by the student for further studies. The choice of the subject and the sex of the child are likely to be important factors for differences in the amount to be spent across subjects. A multinomial logit framework has been used to formulate the model. This model is discussed in Section 3.2.2.

3.2.1 Heckman's model for below class-10 level[21]

Two appropriate models for the simple two-stage decision process for the below class-10 level are Heckman's two step model (Heckman, 1979) and the Double hurdle model (Cragg, 1971). Both are based on the same framework: they address the probability of zero or positive spending

[21] This model has been described in Chapter 2. However, for completeness we are repeating it here.

first using a Probit model (participation equation) and then given that the probability is positive they study the factors that influence the amount of spending using Ordinary Least Squares (OLS) method (outcome equation). Thus, both models have two equations: i.e., participation equation and outcome equation.[22]

There are two latent variables Y_1* and Y_2*. The first latent variable i.e., Y_1* is a variable which depends on factors that determine whether a household would spend on education or not, but Y_1* cannot be observed. The decision is represented by a binary variable Y_1, where $Y_1=1$ means that Y_1* has a strictly positive value and the household is willing to send their child to an educational institution.

$$
\begin{aligned}
Y_1 &= 1 \quad \text{if} \quad Y_1*>0 \\
&= 0 \quad \text{if} \quad Y_1* \leq 0.
\end{aligned}
$$

The *participation equation* is given by

$$Y_1 = X_1^{*/}\beta_1 + \varepsilon_1 \tag{3.1}$$

where ε_1 is a standard normal variate and X_1^* is a vector of independent variables that determine whether a household would send the child to educational institute or not.

The second latent variable Y_2^* (share of monthly spending of the household for education) can only be observed when the household is spending on education and cannot be observed for those who do not spend ($Y_1 = 0$). The outcome equation represents the observed value of Y_2^* dependent on a vector of independent factors X_2^*.

Thus, the *outcome equation* is given by

[22] Zero expenditure on education may be a result of the decision at the first stage and a corner solution at the second stage where households end up incurring zero expenditure. The Double Hurdle model distinguishes between households who decide not to spend in the first stage and those who decide to spend, but end up spending zero amounts. Since no information regarding this can be retrieved from the available data, we use Heckman's model.

$$Y_2^* = X_2^{*/}\beta_2 + \varepsilon_2 \quad (3.2)$$

$$Y_2 = Y_2^* \;\; if \;\; Y_1^*>0$$

$$= 0 \;\; if \;\; Y_1^* \leq 0,$$

where, ε_2 is a normal variate with mean 0 and variance σ^2.

The OLS regression of Y_2 on X_2^* will be inconsistent when the error terms in equation (3.1) and equation (3.2) are correlated.

The conditional censored mean in this framework will be as follows

$$E[Y_2 \,|X_1^*, X_2^*, Y_1^*] = E\left[X_2^{*/}\beta_2 + \varepsilon_2 \,\middle|X_1^{*/}\beta_1 + \varepsilon_1 \geq 0\right]$$

$$= X_2^{*/}\beta_2 + E[\varepsilon_2 \,|\varepsilon_1 \geq -X_1^{*/}\beta_1]. \quad (3.3)$$

To obtain $E[\varepsilon_2 \,|\varepsilon_1 \geq -X_1^{*/}\beta_1]$, Heckman (1979) assumed joint normality of the error terms and arrived at the following expression:

$$E[Y_2 \,|X_1^*, X_2^*, Y_1^*] = X_2^{*/}\beta_2 + \sigma_{12}\,\lambda(X_1^{*/}\beta_1),$$

where, $\lambda\left(X_1^{*/}\beta_1\right) = \phi(X_1^{*/}\beta_1)/\Phi(X_1^{*/}\beta_1)$, which is the Inverse Mill's Ratio[23] and σ_{12} is the correlation between the two error terms.

Hence, the estimating equation is given by

$$Y_2 = X_2^{*/}\beta_2 + \sigma_{12}\,\lambda(X_1^{*/}\beta_1) + u, \quad (3.4)$$

where u is the error term.

The estimate of $\widehat{\beta_1}$ is obtained by a Probit regression and estimate of $\widehat{\beta_2}$ is obtained by applying OLS to equation (3.4). Gender bias is detemined through appropriate parameters in the parameter vectors $\widehat{\beta_1}$ and $\widehat{\beta_2}$.

[23]The inverse Mill's ratio takes account of the possible selection bias, arising out of a censored dependent variable, which causes a concentration of observations at zero values. Ignoring this leads to biased OLS estimates.

3.2.2 Model for above class-10 level

For the above class-10 level, we use a model analogous to the wage differential model of Brown et al. (1980). Suppose there are *J* subject streams and let the stream-specific expenditure function for an individual *l* be:

$$\text{Male:} \quad ln\left[Y_j^l\right] = \alpha_j^m + X_j^{l/}\beta_j^m + \varepsilon_j^{ml} \tag{3.5}$$

$$\text{Female:} \quad ln\left[Y_j^l\right] = \alpha_j^f + X_j^{l/}\beta_j^f + \varepsilon_j^{fl}, \tag{3.6}$$

Here, subscript $j = 1, 2 \ldots\ldots\ldots.J$, denotes subject stream, superscripts *m* and *f* denote males and females, respectively, ε is the error term, *Y* is expenditure on education and *X* denotes the vector of variables assumed to determine the educational expenditure. β is a vector of unknown parameters.

Let the sample proportions of males and females in each subject category be denoted by P_j^m and P_j^f, respectively. In this model, parameters of equations (3.5) and (3.6) are estimated separately for men and women based on a set of personal characteristics. In early models, discrimination was measured as the difference between the intercepts in the two regressions. However, Blinder (1973) and others argued that the β coefficients also contain information about discrimination. Since fitted regressions pass through the means of the data, the raw mean expenditure differential (*R*) in each subject group can be decomposed as follows[24].

$$\underbrace{\overline{ln\,[Y_j^m]} - \overline{ln[Y_j^f]}}_{R} = \underbrace{(\widehat{\alpha_j^m} - \widehat{\alpha_j^f})}_{U} + \underbrace{(\overline{X_j^m}^{/} - \overline{X_j^f}^{/})}_{E}\widehat{\beta_j^m} + \underbrace{\overline{X_j^f}^{/}(\widehat{\beta_j^m} - \widehat{\beta_j^f})}_{C}, \tag{3.7}$$

where $\hat{\alpha}$'s are the estimated values of the intercepts.

Thus, *E* is the portion of the differential due to endowments, *C* is the portion attributable to differing coefficients, and *U* is the unexplained portion. Blinder (1973) defines $D = C+U$ as a

[24] This type of decomposition technique has been used in Meng and Miller (1995).

measure of the portion of the total differential attributable to discrimination. However, this specification of discrimination takes no account of differences in educational attainment. If the same characteristics that determine educational expenditure also determine the choice of subject, then this approach would be sufficient. But, there are likely to be other determinants of choice of subject streams - some from childhood influences, some from personal characteristics, and some from discriminatory constraints on subject choice. An approach that incorporates a separate model of educational attainment into the analysis of expenditure differentials is the explicit inclusion of the probability of being able to choose a certain subject. Taking advantage of the fact that α and β can be estimated using least squares method, we can restate and further decompose the total expenditure differential from (3.7) as follows:

$$\overline{ln\ Y^m} - \overline{ln\ Y^f} = \sum_j \left(P_j^m \overline{lnY_j^m} - P_j^f \overline{lnY_j^f}\right)$$

$$=\sum_j (P_j^m \hat{\alpha}_j^m - P_j^f \hat{\alpha}_j^f) + \sum_j (P_j^m \overline{X_j^m}' \hat{\beta}_j^m - P_j^f \overline{X_j^f}' \hat{\beta}_j^f)$$

$$= \underbrace{\underbrace{\sum_j P_j^f \left(\hat{\alpha}_j^m - \hat{\alpha}_j^f\right) + \sum_j P_j^f \overline{X_j^f}' (\hat{\beta}_j^m - \hat{\beta}_j^f)}_{WU} + \underbrace{\sum_j P_j^f (\overline{X_j^m}' - \overline{X_j^f}') \hat{\beta}_j^m}_{WE}}_{Due\ to\ expenditure} + \underbrace{\sum_j \left(\overline{X_j^m}' \hat{\beta}_j^m + \hat{\alpha}_j^m\right)(P_j^m - P_j^f)}_{Due\ to\ subject}$$

(3.8)

Here the expenditure differential has been further decomposed into the portion attributable to differences in coefficients of (3.5) and (3.6) between male and female (*WU*) and the portion due to differences in the characteristics of male and female (*WE*). The effect of subject differences on the expenditure differential can also be further decomposed into a portion attributable to differences in qualifications for the subject of choice (*BE*) and that due to the structure of educational attainment between male and female (*BU*). Consequently, the full model becomes:

$$\overline{ln\ Y^m} - \overline{ln\,Y^f} = WU + WE + \underbrace{\textstyle\sum_j \overline{lnY_J^m}\left(P_j^m - \widehat{P_J^f}\right)}_{BE} + \underbrace{\textstyle\sum_j \overline{lnY_J^m}\left(\widehat{P_J^f} - P_j^f\right)}_{BU}\ , \tag{3.9}$$

where $\widehat{P_J^f}$ is the proportion of female in the sample who would be in subject *j* if they were to face the same educational structure as males (counter factual proportion). Thus, in the full model *WE* and *BE* capture the expenditure differentials due to differences in measured characteristics between males and females, while those that are unexplained (*WU* and *BU*) reflect differential returns to the measured characteristics and may be the result of discrimination. The estimation of this model requires within-subject expenditure regressions and a technique for predicting the educational distribution of females assuming that they are subject to the same structure of determination as males ($\widehat{P_J^f}$). Since $\widehat{P_J^f}$s are unobserved, we estimate them through a model to predict educational attainment for males on the basis of a number of personal characteristics. The parameters for the male sample are then combined with female characteristics to estimate female educational distribution. This technique is described in detail below.

3.2.2.1 Predicting Educational Distribution

Multinomial Logit Model of Educational Attainment (estimation of $\widehat{P_J^f}$):

An individual's educational attainment is a function of availability of the colleges and of the individual's desire for a specific subject. Willingness of college authorities to admit an individual depends on personal qualifications such as education, training etc. The individual's desire for a particular subject can be expressed by the utility function. The interactions of these supply and demand factors lead to an individual's admission in a particular subject. Owing to constraints of our data set, the interaction is summarized in terms of a reduced form multinomial logit model. The multinomial logit model captures how variables affect the probability of an individual studying a subject, treating the subject choice as endogenously determined. This probability may be defined as

$$P_{lj} = P(T_l = j^{th} stream) = \frac{e^{Z_j^{'}\gamma_j}}{\sum_j e^{Z_j^{'}\gamma_j}}. \tag{3.10}$$

Z_j represents a vector of exogenous variables assumed to be influential in selection and availability of stream j, and γ_j is a vector of coefficients for these variables corresponding to the stream. The qualitative dependent variable T_l can take any of the J possible values, each corresponding to a different stream. Since each individual must select one stream, only $J-1$ sets of coefficients are uniquely defined. We will normalize by setting the coefficients for the J^{th} subject category to zero. The parameters of the model are estimated by the maximum likelihood method. Estimates of the parameters of this model are obtained for male observations, and female data are substituted into the estimated equations producing for each woman a vector of predicted probabilities of belonging to each of the J streams. These predicted probabilities of being in each stream are summed over observations to produce the predicted distribution of females across streams, i.e., $\widehat{P_j^f}$.

3.2.2.2 Incorporation of Selection Bias

Since stream attainment is determined by the interaction between demand factors and supply factors, the samples of individuals observed in each stream may not be random. We collect the samples only from the individuals who are studying in that stream. Hence, we must use the information obtained from equation (3.10) to adjust the educational expenditure equations for potential effects of selection bias. Following Lee (1983), the estimating expenditure equations may be modified to consider the effect of this sample selection bias. Thus, the expenditure equations conditional on stream *j* being chosen is:

$$lnY_j = X'_j\beta_j - \sigma_j\rho_j \frac{\phi[\tau(Z'_j\gamma_j)]}{F(Z'_j\gamma_j)} + \varepsilon_j \tag{3.11}$$

where, ϕ is the standard normal density function, σ is the standard error of the disturbance term in the expenditure equation, ρ is the correlation between stream attainment equation and expenditure equation, τ is a function which is strictly increasing transformation that transforms a random variable associated with stream attainment equation $Z'_j\gamma_j$ into a standard normal variate.

Hence, $\tau = \Phi^{-1}(F)$, where *F* is the distribution of multinomial logit model.

Hence, the transformed expenditure function will look as follows:

$$lnY_j = X'_j\beta_j + \widehat{\lambda_j}\theta_j + \varepsilon_j \tag{3.12}$$

with $\widehat{\lambda_j} = -\frac{\phi[\tau(Z'_j\gamma_j)]}{F(Z'_j\gamma_j)}$ and $\theta_j = -\sigma_j\rho_j$.[25]

3.2.2.3 Decomposition of the Overall Mean Expenditure Differential

Finally, the overall expenditure differential is decomposed into the portion attributable to stream attainment and that attributable to expenditure discrimination, as shown in equation (3.9).

[25] The specification of multinomial logit model is guided by the literature (e.g., Schmidt and Strauss 1975; Miller, 1987; Meng and Miller, 1995). A similar analysis has been performed by Liu, Zhang and Chong (2004) in studying the earnings differentials between the immigrants and the natives of Hong Kong.

3.3 Results and Discussion

3.3.1 Description of sample data and estimation process

The data set used here is the 64th round data on participation and expenditure on education conducted by the National Sample Survey Office (NSSO), Govt. of India, for the state of West Bengal. The span of the data set is July 2007-June 2008. The survey has been conducted over 29 states and 6 UTs. This is the latest data set available that contains detailed information on education expenditure[26]. From this data set a complete profile of household education expenditure pattern at the individual level can also be obtained. The expenditure categories on education and the corresponding NSSO codes are given in Appendix III, Table A3.1. These categories have been merged to form three categories, viz., 'Science', 'Arts' and 'Commerce' (see Appendix III, Table A3.2).

In West Bengal, out of a total of 8524 households, there are 7020 (82.35%) households who have 29442 school or college going children aged 5-29 years. There are 8.28% of the households reporting zero spending on education in urban areas and 9.35% households reporting zero spending on education in rural areas.

But, strikingly, households that spend on education spend approximately 33% and 26% of their monthly per capita spending in urban and rural areas, respectively.

Table 3.1 presents the distribution of students by age. Overall, the percentage of boys attending school is higher than that of girls both in the below class-10 level (age 5-16 years) and in the

[26] Data that are more recent became available after we completed this work. The most recent data set has been used in Chapter 5.

above class 10 level in both sectors. It is very clear that the percentage of students in class 9-10 is much less compared to that in below class 8 in both sectors. The difference between the two percentages is more prominent in the rural sector than in the urban sector. This points to high dropout rate in the higher classes in the rural sector. On the other hand, an examination of the break up by subject stream in above class-10 level reveals that there is concentration of girl students in Arts stream in the rural sector and the percentage of girls studying Arts is higher than that of boys.

The reason behind the overall concentration of students (84.60%) in Arts in the rural sector could be the supply side constraints such as lack of laboratory facilities and lack of availability of other facilities which make students unable to go for science stream.

3.3.2 Methodology

3.3.2.1 The estimation procedure for below class 10 level

In order to capture intra household gender bias, we use the technique adopted by Subramanian and Deaton (1991) and Azam and Kingdon (2013). Four broad age groups for children have been categorized as 0-4 years, 5-9 years, 10-14 years and 15-16 years. $\frac{n_{si}}{n_i}$ is the fraction of household members for household i in the s^{th} age group. As the fractions $\frac{n_{si}}{n_i}$ will add up to 1, the age group 0-4 years has been dropped. According to the average age of children going to different classes below class-10 in West Bengal, children aged 5-9 years are likely to study in primary level (class-1 to class-4), 10-14 years are likely to study in middle level (class-5 to class-8) and 15-16 years are likely to study in secondary level (class-9 to class-10). Each age group is further categorized into male (m) and female (f). The proportions are denoted as

$[\frac{n_{Pi}}{n_i}]_m$ for male aged 5-9 years (*P* denotes primary).

$[\frac{n_{Pi}}{n_i}]_f$ for female aged 5-9 years .

$[\frac{n_{Mi}}{n_i}]_m$ for male aged 10-14 years (*M* denotes middle).

$[\frac{n_{Mi}}{n_i}]_f$ for female aged 10-14 years.

$[\frac{n_{Si}}{n_i}]_m$ for male aged 15-16 years (*S* denotes secondary).

$[\frac{n_{Si}}{n_i}]_f$ for female aged 15-16 years.

First, a Probit regression has been run to address the question whether a household will spend on children education $(w_i > 0)$ or not $(w_i = 0)$, w_i being the share of education expenditure for household i. The explanatory variables are: education level of household head (*hdedu*)[27], household size (n_i), logarithm of monthly per capita consumption expenditure (*LnMPCE*), square of logarithm of monthly per capita consumption expenditure (*Lsq*), $[\frac{n_{Pi}}{n_i}]_m$, $[\frac{n_{Pi}}{n_i}]_f$, $[\frac{n_{Mi}}{n_i}]_m$, $[\frac{n_{Mi}}{n_i}]_f$, $[\frac{n_{Si}}{n_i}]_m$, $[\frac{n_{Si}}{n_i}]_f$ and distance from residence to educational institute (*DISTANCE*).

The second stage equation is given by,

$$w_i = \alpha + \beta_1\, lnx_i + \beta_2[lnx_i]^2 + \sum_{s=P,M,S} \theta_{sm}\, [\frac{n_{si}}{n_i}]_m + \sum_{s=P,M,S} \theta_{sf}\, [\frac{n_{si}}{n_i}]_f + \eta z_i + \mu\, lnn_i + \delta\hat{\lambda} + \varepsilon_i. \tag{3.13}$$

Here, z_i denotes amount of land possessed (*LAND*), lnn_i (*lnhhsize*) allows for individual scale effect, θ_s reflects the effect of changing household composition without changing the household size (e.g., replacing a boy by a girl in a given age group). Testing gender difference reduces to

[27] Here education level of the household head is the years of schooling of the head.

testing $\theta_{sm} = \theta_{sf}$, i.e, $\theta_{sm} - \theta_{sf} = 0$, where s = *P*, *M*, *S* in equation (3.13) and similar tests for the corresponding coefficients in the Probit model. Since the decision of a household will still involve two stages described in introduction, the same model has been used separately for the two different sections i.e., below class 8 and class 9-10 in below class-10 level.

In order to assess the presence of gender bias properly, we have chosen the households that have both male and female children going to schools at the below class 10 level. In West Bengal, there are 966 households with the composition of husband, wife, at least one male and at least one female child who belong to the age group of 5-16 years (below class 10 level). Moreover, we have done the estimation for below and above the average *MPCE* separately for both sectors. However, in the urban sector the number of households reporting zero spending for education is quite small. Hence, the estimation of the budget share of the education expenditure for i^{th} household (w_i) is done by the OLS regression (instead of Heckman's two-step model). In the rural sector, the number of households reporting zero spending for education is very small in the income range of above the mean of MPCE. Hence, the estimation of w_i is done by the OLS regression for this section of the population.

Here, the explanatory variables are similar like those used in equation (3.13)[28]. However, for sake of simplicity, the four broad age group categories are not taken into consideration. Instead, the fractions of household size belonging to the age group 5-16 years (below class 10 level) for both sexes (*Child_m* and *Child_f*) are taken as explanatory variables. Like before, the testing of gender bias boils down to testing of equality of coefficients of the fractions for these two age-sex demographic groups.

[28] lnn_i is not taken as an explanatory variable to avoid multicollinearity problem as $lnn_i$is likely to be correlated with *Child_m* and *Chlid_f*.

3.3.2.2 The estimation procedure for above class 10 level

The estimation is carried out in two stages. First, estimates of the coefficients vector $\widehat{\gamma_j}$ in the multinomial logit model for equation (3.10) are obtained by maximum likelihood method for male students to obtain the counterfactual probabilities for female students using the estimated vector $\widehat{\gamma_j}$. In the multinomial logit model, we choose Commerce as the base outcome. The dependent variables are the log of odds ratio of being in Arts (denoted by A) vs. Commerce (denoted by C) and Science (denoted by S) vs. Commerce. Thus, for Arts and Science, the dependent variables are log $[\frac{Prob(T_l=A)}{Prob\ (T_l=C)}]$ and log $[\frac{Prob(T_l=S)}{Prob\ (T_l=C)}]$, respectively, for l^{th} individual. Here, Z_j represents a vector of exogenous variables that influence selection and availability of the stream. This includes *LnMPCE, Lsq,* age of the student at entry to school (AGE), D_{1i} is a dummy representing whether education is free or not (where D_{1i} =1 if free and 0 for otherwise),[29] log of per capita expenditure on education *(ln(eduexp)),* number of courses studied by the student before (NUMBER)(more courses means more competence in education), *hdedu,* n_i, D_{2i} is a dummy variable representing whether the household is female headed or not (D_{2i}=1 if household head is female and 0 otherwise).

Next, coefficients ($\widehat{\gamma_j}$) are used to calculate $\hat{\lambda}$ and then to run an OLS regression for equation (3.12) for each stream and each sex for the urban sector.[30] Here, X_j is a vector of explanatory variables which includes *LnMPCE, Lsq, DISTANCE, hhsize* , OTHEREXP[31], D_{2i} and $\hat{\lambda}$.

---

[29] It is often found that students, after their class-10 board exam get some financial help from state government via accessibility to certain facilities.

[30] The number of observations on girls studying Science and Commerce in the rural sector is too small to run meaningful regression.

[31] OTHEREXP = (Total expenditure - Education expenditure).

3.3.2.3 Discussion of results for below class-10 level

In this subsection, results of below class-10 level have been reported. The empirical results for the Probit model and second stage regression are reported in Table 3.2a and Table 3.2b for below class 8 section and for class 9-10 section, respectively.

All the marginal effects are calculated at mean values of the explanatory variables.

For the Probit model, in the rural sector, the sign of marginal effect of *LnMPCE* is positive, significant and that of *Lsq* is negative, significant for both below class 8 section and class 9-10 section. In the urban sector, the marginal effects of *LnMPCE* and *Lsq* are non-significant for below class 8 section, but significant with positive coefficient of *LnMPCE* and negative coefficient of *Lsq* for class 9-10 section. Thus, for households in both rural and urban sectors, income is a determining factor for sending children belonging to class 9-10 section to school. At the below class 8 level, while income plays a significant role in the rural sector, its role is non-significant in the urban sector. This clearly brings in the issue of the differential costs that rural and urban households face to access primary education possibly because of lack of awareness of the RTE act on the part of the rural households.

The household head's education level (*hdedu*) has positive and significant impact on the decision to spend on education for children studying in class 9-10 in both sectors, but is non-significant for children studying in below class 8 section in both sectors.

The marginal effect of household size (n_i) has a significant negative impact in both sectors for children studying in both sections (i.e., below class 8 and class 9-10).

For households with children studying in below class 8 section, the marginal effects of the proportions of male and female students aged 5-9 years and 10-14 years are positive and significant in both sectors. For the age group of 15-16 years, in both the sectors, the effect is negative and significant. The increase in proportion of children in the age groups 5-9 years and 10-14 years makes the household better off due to the economies of scale (Kingdon, 2005). On the other hand, for households with children studying in class 9-10 section, marginal effects of the proportions of male and female students aged 5-9 years, 10-14 years and 15-16 years, are negative and significant in both sectors. In all cases, however, the magnitudes are smaller for females.

The effect of distance from residence to educational institute (*DISTANCE*) has significant negative impact on the decision to spend for education for both sections of children for the urban sector, possibly owing to higher transportation cost. In the rural sector, on the other hand, where children mostly walk to school, distance is a matter of concern for the parents of the younger group (the effect is negative and significant), while the effect is non-significant for children in class 9-10.

For the second stage conditional OLS, for below class 8 section, coefficient of *LnMPCE* is positive for both the urban and the rural sectors but is non-significant in the rural sector. The coefficient of *Lsq* is negative for both sectors but non-significant for the rural sector. Thus, for rural households, whose children are studying in below class 8 section, once the decision to send their child to school is taken, income becomes irrelevant. On the contrary, at this stage for urban households income is a significant factor possibly because of various choices of types of schools. In case of class 9-10 section, the coefficient of *LnMPCE* is positive and significant for both the urban and the rural sector, coefficient of *Lsq* is negative and significant for both sectors. Thus, for the households whose children are studying in class 9-10 section, income of the household is playing a significant role when the amount to be spent is under consideration as the amount of free services provided under RTE act is not applicable for children studying in class 9-10 section (as they are above 14 years).

The coefficient of log of household size (*lnhhsize*) is negative and significant for both below class 8 and class 9-10 sections in the urban sector, but in the rural sector in the below class 8 section, the coefficient is positive and significant while that in class 9-10 section, the coefficient is negative and significant. Thus, in the rural sector, the positive effect of economies of scale defined in Kingdon (2005) is corroborated.

In below class 8 section, the proportions of children belonging to age group of 5-9 years and 10-14 years in a household have significant positive impacts on w_i for male and significant negative impact for female in the urban sector. In the rural sector, the effects are positive for male, but non-significant for female. This is in line with school dropout at the primary level in the rural sector, where children, especially, girls are sent off for other kinds of works. The positive coefficients for proportion of males in this age group in both sectors establish that they are not discouraged from education, while the negative impact for female children points to possibility of discrimination. The coefficients of the fractions of children in 15-16 years are negative and significant in both sectors for male and female, with smaller coefficients for females.

In class 9-10 section, the proportions of children in 5-9 years and 10-14 years have negative and significant effects on w_i in both sectors for male and female. But the proportion of male child in 15-16 years has positive and significant impact on w_i in both sectors, while the proportion of female child of this age group has negative and significant impact on w_i in both sectors.

Amount of land possessed (*LAND*) is non-significant in the urban sector for both sections but has significant and positive impact on share of spending on education for both sections in the rural sector, which is quite plausible. The coefficients of λ are significant for both sectors. This means that there is a correlation between the decision to spend on education and the actual amount to be spent for education, which implies that conventional OLS regression would have led to inconsistent results.

Table 3.3 reports the results of testing the restrictions $\theta_{Pm} - \theta_{Pf} = 0$, $\theta_{Mm} - \theta_{Mf} = 0$, $\theta_{Sm} - \theta_{Sf} = 0$ for the two sections below the secondary level. For both sections, in case of deciding whether to spend on education or not and the amount to be spent on education, the values of $\theta_{Pm} - \theta_{Pf}$, $\theta_{Mm} - \theta_{Mf}$, $\theta_{Sm} - \theta_{Sf}$ are all positive in both sectors. However, while the F-values are non-significant for the age group 5-9 years, the F-values for age groups 10-14 years and 15-16 years are significant. This means that at the decision making stage as well as in terms of spending on education, there exist discrimination against girls at the Middle and Secondary level in both sectors, but not at the primary (5-9 years) level. The discrimination is apparent in the estimates of the second stage parameters, but not quite obvious from the Probit regression.

In Table 3.4, the second stage OLS regression results for the urban sector are reported for both income ranges (i.e. below mean MPCE and above mean MPCE). It can be seen that the signs and significance of the coefficients are similar to that in Table 3.2a and Table 3.2b. The coefficients of lnx_i and its squared term are positive and negative, respectively. However, the coefficient of *LAND* is positive, significant in the below MPCE income range. The signs of coefficients of *Child_m* and *Child_f* are same (both positive) in the above mean of MPCE, and opposite (positive for *Child_m* and negative for *Child_f*) in the below mean of MPCE. This supports the findings of the presence of male encouragement in education in below mean of MPCE income range in the urban sector. Further, in this context, the results of the hypotheses testing reported in Table 3.5, supports the above mentioned finding. In Table 3.6, the results of

Heckman's two step model for below mean of MPCE income range are reported in the rural sector. The coefficients of lnx_i, its squared term and *LAND* are significant and hence, usual explanation follows. The signs of coefficients of *Child_m* and *Child_f* are same in the Probit regression and opposite in the OLS regression, which indicates the presence of gender bias in the expenditure stage. These findings are supported by similar pattern of signs in Table 3.7 for the OLS regression in the above mean of MPCE income range in the rural sector. In Table 3.8, the reported results of the hypotheses testing indicate rejection of H_0 in favour of H_1. The presence of gender bias is detected in both stages (decision making and expenditure) in the rural sector.

3.3.2.4 Discussion of results for above class-10 level

The empirical results of the multinomial logit model are reported in Table 3.9. The multinomial logit model has been fitted using Commerce as base outcome.

The coefficients of log of monthly expenditure *(LnMPCE)* are negative and significant for Arts and positive and significant for Science for both sectors. It implies that with the increase in monthly income, the probability of choosing Arts over Commerce falls and probability of choosing Science over Commerce rises. This is also evident from the average monthly per capita spending by households on these streams: Science (Rs. 1577.89), Commerce (Rs. 1011.67), Arts (Rs. 525.73), and from the coefficient of (*ln(eduexp)*). The coefficients of *Lsq* are positive for Arts and negative for Science in both sectors, which implies that as the spending rises, the probability of choosing Arts over Commerce falls at an increasing rate and the probability of choosing Science over Commerce rises at a decreasing rate. The explanatory variable age at entry to school (AGE) is negative and significant for choosing Science over Commerce in the rural sector implying that as the students start schools late, they choose Commerce over Science. The explanatory variable-whether education is free or not (denoted by a dummy variable D_{1i}) has similar signs like *LnMPCE* which implies that making education free leads to higher probability of choosing Science over Commerce and Commerce over Arts. However, this variable is not significant for the rural sector. The number of courses studied (NUMBER) is non-significant for both subjects in the urban sector and for Arts in the rural sector, but positive and significant for Science in the rural sector. This implies that, as the number of courses studied rises, the students choose Science over Commerce. In other words, more courses studied leads to higher eligibility for studying Science. Household size (n_i) is found to have positive and significant effect for Arts in both sectors and negative and significant effect for Science in the urban sector. This means that as household size rises, students choose Arts over Commerce in both sectors and Commerce over Science in the urban sector. The underlying reason could be that studying Arts is less expensive than studying Commerce, which in turn is less expensive than studying Science. The effect of household head's education level (*hdedu*) is significant for both subjects for both sectors. While the coefficient is negative for Arts, it is positive for Science, thus implying that with higher education of the household head, inclination is lesser towards Arts and higher towards Science compared to Commerce. The coefficient of D_{2i} is non-significant for both subjects for both sectors. Thus, whether the household is female headed or not does not have any effect.

The predicted and observed probabilities of the multinomial logit model are reported in Table 3.10. The reported values in the last row $\left(\widehat{P_f}\right)$ are obtained by estimating the predicted probabilities if the female students were given same privileges and conditions like male students (using data for female and estimated coefficients obtained from male). As is evident from the table, the observed proportion of females is higher than that of males in ‘Arts’, where the average monthly per capita spending by households is the least among the three streams. The observed proportions of females are lower in ‘Science’ and ‘Commerce’ in both sectors.[32] A comparison between P_f and $\widehat{P_f}$ shows that if female students were given similar opportunities like male students there would be a shift from ‘Arts’ and a substantial increase in proportion of female students in ‘Science’ and ‘Commerce’. The actual(P_f) and counter factual probabilities ($\widehat{P_f}$) of choosing a subject stream are provided in Table 3.10. $\widehat{P_f}$ is the probability of choosing a stream by a female child if she were treated like a male child. It can be seen that there is a substantial shift to Science and Commerce from Arts in both sectors. However, the percentage of female students who actually *prefer* to study Arts can be calculated by $P_f - \widehat{P_f}$. This is the percentage of female students who study Arts even after getting the opportunities to study other streams. These percentages are 18% and 42% in the rural and the urban sector, respectively. We may mention that Gautam (2015) found female students opting for Arts stream in colleges even after studying Science in their plus 10+2 school levels in India.

After estimating the predicted probabilities for each stream for both sexes, an OLS regression of equation (3.12) has been run.

As already noted, the number of observations on girls studying Science and Commerce in the rural sector is very small and hence does not permit estimation of coefficients of all explanatory

[32]Given that we are addressing the issue of gender discrimination in terms of cost incurred by a household on education, we look at the issue from the perspective of female participation, given that the males are already concentrated in streams that involve higher cost.

variables in each individual stream. We, therefore, concentrate on the urban sector, the results of which are reported in Table 3.11.

The coefficient of *LnMPCE* has a positive impact on education expenditure irrespective of sexes and streams, which means that with the increase in income, expenditure on education also rises, but at a decreasing rate, as the coefficient of *Lsq* is negative and significant. The distance from educational institute has positive and significant effect as more spending is required for transportation. The explanatory variable- other expenditure (OTHEREXP) is negative and significant for all subjects for both sexes except for Science for female students. Household size (n_i) has a mixed effect. An interesting feature regarding the variable D_{2i} is that it is significant and positive only for Science in the urban sector for female students. This implies that female students are encouraged to study Science in female headed households in the urban sector. The significant inverse mills ratio λ indicates that there exists sufficient sample selection bias ignoring which would have led to inconsistent results. For the female students for Science and Commerce, the estimates of λ turn out to be non-significant. This implies that choice of these subjects is largely random in the female data.

The results of the decomposition in equation (3.9) are presented in Table 3.12 for the urban sector only.

The percentages of the total differential for *WE*, *WU*, *BE*, *BU* are 35.23%, 28.86%, 30.42% and 5.49%, respectively. Hence, 35.91% of total differential comes from between stream component and 64.06% from within stream component. Of the 35.91% in the between component, *BU* contributes 83.71%, and of the 64.06% in the within component *WU* contributes 53.09%. Thus, the unexplained part, which reflects differential returns to measured characteristics of males and females, has a larger contribution and hence points to discrimination, as noted earlier. To determine the exact nature of discrimination, we look at the individual terms of *WU* and *BU* for each stream in equations (3.8) and (3.9). From Table 3.6 it can be calculated that the term $\left(\hat{\alpha}_j^m - \hat{\alpha}_j^f\right) + \overline{X_j^f}^{/}(\hat{\beta}_j^m - \hat{\beta}_j^f)$ in *WU* are -0.991, 15.158, 12.569 for Arts, Science and Commerce, respectively. Since the term is positive for Science and Commerce and negative for Arts, it means that in the expenditure allocation for studying the same subject there is pro male bias in Science and Commerce, and pro female bias in Arts.

On the other hand, from Table 3.10 it can be seen that the stream specific term $\widehat{P_j^f} - P_j^f$ is negative for Arts and positive for Science and Commerce. This means that there is pro male bias in choice of subject streams Science and Commerce.

3.4 Conclusion

In this chapter, we have addressed the issue of gender discrimination differently from the existing literatures by applying two different models for two different levels of education i.e., below secondary and above secondary level. While the analysis for below the secondary level is in line with the that used in the existing literature, this study adds a further dimension by recognizing the Right to Education act (2005), which applies to the below class 8 level. The gender discrimination study for the above secondary level is based on a multinomial logit model that takes into account the preferences of students and their implications. This study establishes that there exists pro male bias in the decision to send the child to school and spending on education in the below class-10 level in both the rural and the urban sectors for children belonging to the age groups 10-14 years and 15-16 years, which corroborates the findings of Kingdon (2005). For the above class-10 level, the detailed analysis indicates evidence of discrimination against girl students at the level of selection of subject streams such as Science and Commerce and also in the allocation of expenditure within a subject stream in the urban areas in the state of West Bengal during 2007-08.[33]

[33] The data did not permit examination of discrimination in the rural sector in our framework at the above class-10 level.

Table 3.1: The Distribution (Per cent) of Students Attending Educational Institutions

SEX	AGE													
	5-16 years						17-29 years							
	RURAL			URBAN			RURAL				URBAN			
	CLASS		TOTAL	CLASS		TOTAL	STREAM			TOTAL	STREAM			TOTAL
	1-8	9-10		1-8	9-10		ARTS	SCIENCE	COMMERCE		ARTS	SCIENCE	COMMERCE	
MALE	45.86	6.10	51.96	31.30	23.09	54.39	52.22	13.12	2.02	67.36	19.79	11.46	13.02	54.27
FEMALE	42.80	5.24	48.04	25.05	20.56	45.61	32.38	0.26	0.00	32.64	32.29	9.80	3.64	45.73
TOTAL	88.66	11.34	100	56.35	43.65	100	84.60	13.38	2.02	100	52.08	21.26	16.66	100

Table 3.2a: The Results for Below Class 8 Level

VARIABLES	SECTORS			
	RURAL		URBAN	
	PROBIT (w_i>0 or =0)[$$]	OLS (w_i)	PROBIT (w_i> 0 or =0)	OLS (w_i)
LnMPCE	6.483 (0.000)***	0.036 (0.000)***	2.306 (0.137)	0.451 (0.007)***
Lsq	-0.409 (0.000)***	-0.006 (0.000)***	-0.132 (0.840)	-0.027 (0.001)***
n_i	-0.064 (0.000)***		-0.048 (0.010)**	
lnhhsize		0.048 (0.000)***		-0.049 (0.000)***
hdedu	0.109 (0.118)		-0.067 (0.102)	
DISTANCE	-0.276 (0.000)***		-0.238 (0.008)***	
LAND		0.008 (0.000)***		0.011 (0.122)
Λ		0.226 (0.000)***		0.237* (0.000)
Primary_m	4.548 (0.000)***	0.288 (0.000)***	6.764 (0.000)***	-0. 772 (0.000)***
Primary_f	4.452 (0.000)***	0 .301 (0.201)	6.168 (0.000)***	-0.838 (0.000)***
Middle_m	1.797 (0.000)***	0.914 (0.000)***	3.297 (0.000)***	0.977 (0.000)***
Middle_f	1.458 (0.000)***	0.819 (0.101)	2.231 (0.000)***	0.857 (0.000)***
Secondary_m	-2.421 (0.000)***	-0.329 (0.000)***	-2. 501 (0.000)***	-0.202 (0.000)***
Secondary_f	-3.243 (0.000)***	-0 .396 (0.000)***	-2.626 (0.000)***	-0.505 (0.000)***
CONSTANT	-26.122 (0.000)***	0.007 (0.000)***	-10.443 (0.009)***	-2.026 (0.000)***

Figures in parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

[$$]The expressions in the parentheses denote the dependent variables.

Table 3.2b: The Results for Class 9-10 Level

VARIABLES	SECTORS			
	RURAL		URBAN	
	PROBIT (w_i>0 or =0)[$$]	OLS (w_i)	PROBIT (w_i> 0 or =0)	OLS (w_i)
LnMPCE	2.768 (0.000)***	1.109 (0.000)***	3.777 (0.008)***	1.493 (0.001)***
Lsq	-0.136 (0.000)***	-0.055 (0.000)***	-0.201 (0.014)**	-0.021 (0.000)***
n_i	-0.061 (0.000)***		-0.039 (0.001)***	
lnhhsize		-0.003 (0.000)***		-0.065 (0.002)***
hdedu	0.098 (0.008)***		0.178 (0.001)***	
DISTANCE	0.041 (0.137)		-0.172 (0.000)***	
LAND		0.015 (0.000)***		-0.023 (0.175)
Λ		9.007 (0.000)***		0.201 (0.000)***
Primary_m	-3.097 (0.000)***	-0.132 (0.000)***	-3.057 (0.000)***	-0.059 (0.000)***
Primary_f	-3.141 (0.000)***	-0. 158 (0.000)***	-3.279 (0.000)***	-0. 142 (0.000)***
Middle_m	-1.253 (0.000)***	-0.366 (0.009)***	-0.388 (0.000)***	-0.070 (0.001)***
Middle_f	-1.845 (0.000)***	-0.445 (0.001)***	-0. 509 (0.000)***	-0. 197 (0.000)***
Secondary_m	-3.151 (0.000)***	0.067 (0.000)***	-3.024 (0.000)***	0.510 (0.001)***
Secondary_f	-3.789 (0.000)***	-0.124 (0.000)***	-3.324 (0.000)***	-0.658 (0.001)***
CONSTANT	-14.412 (0.001)***	1.087 (0.000)***	-18.340 (0.001)**	2.782 (0.001)***

Figures in parentheses are the p-values.

***: Significant at 1% level; **:   Significant at 5% level; *: Significant at 10% level.

[$$]The expressions in the parentheses denote the dependent variables.

Table 3.3: Values of F-Statistic (Testing $\theta_{sm} - \theta_{sf} = 0$, $s = P, M, S$) for Below Class 10 Level

Education (Age)	Sections In Below Class 10 Level	Sectors			
		Rural		Urban	
		Probit	OLS	Probit	OLS
PRIMARY (5-9 YEARS)	**CLASS 1-8**	0.069 (0.198)	0.785 (0.102)	0.580 (0.445)	0.780 (0.113)
	CLASS 9-10	0.010 (0.098)*	0.801 (0.101)	0.070 (0.708)	0.050 (0.709)
MIDDLE (10-14 YEARS)	**CLASS 1-8**	14.511 (0.001)***	4.920 (0.026)***	18.720 (0.004)***	12.720 (0.007)***
	CLASS 9-10	12.350 (0.001)***	10.209 (0.006)***	10.510 (0.009)***	11.060 (0.007)***
SECONDARY (15-16 YEARS)	**CLASS 1-8**	13.46 (0.001)***	11.498 (0.002)***	10.040 (0.004)***	13.420 (0.009)***
	CLASS 9-10	11.89 (0.009)***	17.145 (0.002)***	15.770 (0.005)***	23.300 (0.001)***

Figures in parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 3.4: The Second Stage Regression for the Urban Sector (w_i)

Explanatory Variables	Below Mean	Above Mean
LnMPCE	0.504 (0.056)**	0.138 (0.006)***
Lsq	-0.028 (0.006)***	-0.007 (0.001)***
***Child_m* (5-16 years)**	0.044 (0.009)***	0.088 (0.021)***
***Child_f* (5-16 years)**	-0.007 (0.021)***	0.022 (0.011)***
LAND	0.041 (0.011)***	0.068 (0.233)
Constant	0.188 (0.001)***	0.509 (0.001)***
No. of observations	146	72
R^2	0.1759***	0.1987***

*Figures in the parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 3.5: P-values of the Hypotheses Testing to Detect Gender Bias (Urban Sector)

Hypotheses	Below Mean	Above Mean
$H_0: \theta_m - \theta_f = 0$ $H_1: \theta_m - \theta_f > 0$	0.023	0.108
Decision	Rejection of H_0	Non rejection of H_0

Table 3.6: Heckman's Estimation for the Rural Sector (Below Mean of MPCE)

Explanatory variables	Below Mean of MPCE	
	PROBIT (w_i>0 or =0)[$$]	**OLS (w_i)**
LnMPCE	1.225 (0.058)*	1.053 (0.001)***
Lsq	-0.658 (0.045)**	-0.068 (0.012)**
***Child_m* (5-16 years)**	0.033 (0.001)***	0.025 (0.002)***
***Child_f* (5-16 years)**	0.137 (0.001)***	-0.003 (0.002)***
Hdedu	0.126 (0.011)**	
DISTANCE	-0.048 (0.002)***	
LAND		0.706 (0.039)**
Λ		0.032 (0.001)***
Constant	5.111 (0.000)***	0.547 (0.001)***
No. of observations	461	

*Figures in the parentheses are the p-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 3.7: OLS Estimation (w_i) for the Rural Sector (Above Mean of MPCE)

Explanatory Variables	Coefficients
LnMPCE	0.006 (0.001)***
Lsq	-0.107 (0.011)**
***Child_m* (5-16 years)**	0.303 (0.005)***
***Child_f* (5-16 years)**	-0.142 (0.047)**
LAND	0.164 (0.046)**
No. of observations	287
$\mathbf{R^2}$	0.1689

*Figures in the parentheses are the p-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 3.8: P-values of the Hypotheses Testing to Detect Gender Bias (Rural Sector)

Hypotheses	**Below Mean**		**Above Mean**
	Probit	**OLS**	
$H_0: \theta_m - \theta_f = 0$ $H_1: \theta_m - \theta_f > 0$	0.012	0.043	0.029
Decision	Rejection of H_0	Rejection of H_0	Rejection of H_0

Table 3.9: The MLogit Marginal Effects for Above Class 10 Level

Variables	SECTORS			
	RURAL		URBAN	
	ARTS $(\frac{Prob\ (T=A)}{Prob\ (T=C)})^{\$\$}$	SCIENCE $(\frac{Prob\ (T=S)}{Prob\ (T=C)})$	ARTS $(\frac{Prob\ (T=A)}{Prob\ (T=C)})$	SCIENCE $(\frac{Prob\ (T=S)}{Prob\ (T=C)})$
$LnMPCE$	-0.00039 (0.000)***	0.00001 (0.009)***	-0.001 (0.000)***	0.0002 (0.000)***
Lsq	0.00005 (0.045)**	-0.0003 (0.001)***	0.0002 (0.001)***	-0.0001 (0.001)***
AGE	0.303 (0.285)	-0.633 (0.046)**	0.845 (0.978)	0.010 (0.577)
D_{1i}	0.079 (0.876)	0.236 (0.987)	-1.021 (0.030)**	0.904 (0.001)***
ln(eduexp)	-2.464 (0.011)**	0.752 (0.030)**	-1.765 (0.001)***	0.564 (0.001)***
NUMBER	-1.67452 (0.856)	1.4007 (0.001)***	-1.835 (0.071)*	-17.435 (0.994)
n_i	0.046 (0.010)**	-7.956 (0.994)	0.073 (0.001)***	-0.192 (0.007)***
hdedu	-13.626 (0.001)***	12.665 (0.004)***	-11.745 (0.005)***	9.230 (0.030)**
D_{2i}	-150.541 (0.996)	-138.027 (0.997)	90.532 (0.998)	101.601 (0.996)
Constant	5.308 (0.010)***	3.052 (0.000)***	3.201 (0.002)***	4.707 (0.001)***

Figures in parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

$$ The expression in parentheses is the dependent variable.

Note: Commerce is the base outcome.

Table 3.10: Observed and Predicted Probabilities from MLogit

PROBABILITIES		SECTORS					
		RURAL			URBAN		
		ARTS	SCIENCE	COMMERCE	ARTS	SCIENCE	COMMERCE
OBSERVED	P_m	0.8566	0.1189	0.0246	0.5690	0.2310	0.2000
	P_f	0.9524	0.0476	0.0000	0.7369	0.1657	0.0947
PREDICTED	$\widehat{P_m}$	0.7981	0.1602	0.0415	0.6043	0.2314	0.1641
	$\widehat{P_f}$*	0.6718	0.3895	0.0385	0.3186	0.2241	0.3571

*These are counterfactual probabilities.

Table 3.11: The OLS Regression for Above Class 10 Level for the Urban Sector
(Dependent variable: *ln (eduexp)*)

VARIABLES	SUBJECTS					
	ARTS		SCIENCE		COMMERCE	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
$LnMPCE$	2.486 (0.000)***	3.314 (0.000)***	1.970 (0.000)***	1.325 (0.000)***	1.054 (0.000)***	2.742 (0.001)***
Lsq	-0.169 (0.000)***	-0.220 (0.000)***	-0.134 (0.000)***	-0.451 (0.000)***	-0.006 (0.001)***	-0.168 (0.001)***
DISTANCE	0.005 (0.038)**	-0.0006 (0.992)	0.002 (0.446)	0.003 (0.804)	0.051 (0.001)***	0.058 (0.002)***
OTHEREXP	-0.00008 (0.000)***	-0.00005 (0.000)***	-0.00007 (0.000)***	0.040 (0.997)	-0.0005 (0.016)**	-0.000025 (0.016)**
n_i	0.013 (0.000)***	0.009 (0.058)*	-0.008 (0.033)**	0.012 (0.203)	-0.025 (0.006)***	0.053 (0.152)
D_{2i}	NA$^{\$\$}$ (NA)	-0.005 (0.998)	NA (NA)	0.488 (0.000)***	NA (NA)	NA (NA)
λ	0.083 (0.001)***	-0.0004 (0.001)***	-0.023 (0.046)**	-7.459 (0.770)	0.004 (0.001)***	0.037 (0.166)
CONSTANT	-5.681 (0.000)***	-4.683 (0.000)***	-3.915 (0.000)***	-2.459 (0.000)***	-9.007 (0.000)***	-7.802 (0.010)***

Figures in parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

$$ The regression coefficients and the p-values cannot be obtained due to lack of observations.

Table 3.12: The Expenditure Decomposition (Urban Sector)

Total	Within		Between	
$\overline{ln\ Y^m} - \overline{ln\ Y^f}$	**WE** $\sum_j P_j^f (\overline{X_j^m} - \overline{X_j^f})\hat{\beta}_j^m$	**WU** $\sum_j P_j^f \left(\hat{\alpha}_j^m - \hat{\alpha}_j^f\right) + \sum_j P_j^f \overline{X_j^f}(\hat{\beta}_j^m - \hat{\beta}_j^f)$	**BE** $\sum_j \overline{lnY_j^m}\left(P_j^m - \widehat{P_j^f}\right)$	**BU** $\sum_j \overline{lnY_j^m}\left(P_j^f - \widehat{P_j^f}\right)$
8.4335	2.4344	2.9718	0.4618	2.5655
% contribution in total (100)	64.06%		35.91%	
% contribution in Within (100) and Between (100)	46.91%	53.09%	16.29%	83.71%

Appendix III

Table A3.1: THE SUBJECT CODES PROVIDED BY NSSO

Subjects	Codes
General courses (up to class-10)	01
Arts and Humanities	02
Science	03
Commerce	04
Medicine	05
Engineering	06
Agriculture	07
Law	08
Management	10
Education	11
CA and similar courses	12

Table A3.2: THE MODIFIED SUBJECT STREAMS FOR ANALYSIS

STREAMS	SUBJECTS
Science (S)	Science, Medicine, Engineering and Agriculture
Arts (A)	Arts, Law and Education
Commerce (C)	Commerce, Management and CA and similar courses

Chapter 4

Gender Bias in Education in West Bengal: An Application of Intra-Household Collective Model

4.1 Introduction

The recent literature on household demand models focuses on intra-household models. In intra-household models, each individual household member is characterized by his/her own rational preferences. This is in contrast with the conventional unitary models, where a household as a whole is assumed to be a single decision maker. A variety of intra-household models has been developed in the literature along two major directions: *Collective models* (Pareto efficient models) [see Becker (1973, 1981), Chiappori (1988, 1992), Chiappori and Browning (1998), Arias, Atella and Perali (2004), Lewbel and Pendakur (2008), Lisa and Seitz (2011), Browning, Chiappori and Lewbel (2013), Chiappori, Lacroix and Fortin (2002), Cherchye et al. (2015) and the references therein] and *Cooperative/ non-cooperative* bargaining models (McElroy and Hurney, 1981; Lundberg and Pollack, 1993). Pareto efficient models assume that husband and wife jointly choose an efficient allocation of resources through a 'sharing rule', although they have separate utility functions. The *Cooperative/ non-cooperative* bargaining framework involves 'Nash-bargaining game' and 'separate spheres' models. This chapter focuses on a *collective model* in the context of India, in particular, the state of West Bengal.

A large number of papers (Lundberg, Pollak and Wales, 1997; Phipps and Burton, 1998; Bobonis, 2009) have rejected the income pooling restriction (which is the idea that the household's demands depend only on its total income and not on the sources of income), used in the unitary model. Lundberg, Pollak and Wales (1997), Phipps and Burton (1998) and Bobonis (2009) found that the expenditure on the children's clothing, health and education increases when the mother's income improves in a household. Reggio (2011), using data from Mexico, pointed out that the daughter's hours of work as labourer decreases with the increase in the mother's bargaining power. Attanasio, Battinstin and Mesnard (2012) showed that in Colombia, a conditional cash transfer scheme (PROGRESA), where the money is received by the mother (which enhances the mother's bargaining power), leads to a more than proportionate increase in the share of food than that by an increase in the total consumption.

In the context of India, Calvi, Lewbel and Tommasi (2017) pointed out that the women's bargaining power has positive impact on her and children's health outcomes (i.e., they are less likely to be anaemic or underweight). In another study on India, Fuwa et al. (2006) estimated intra-household resource allocation rules, incidence of child labour, and the effects of credit constraints on time allocation among household members in Andhra Pradesh, India. However, the data set was collected exclusively for this purpose. They collected time use data, premarital and parental generation data and also collected information on extra-household linkages, such as existence or absence of parents of the couple, to remedy the problem of potential endogeneity of "bargaining power" proxy variables (Quisumbing and Maluccio (2000)).

'Sharing rule' of allocation of resources is often interpreted as an indicator of the bargaining power and governs the within-household distribution (apportionment) of household resources. An attractive feature of the resource-sharing rule is that it is expressed in monetary terms. The sharing rule is also useful for recovering information about the economic well-being of household members [Cherchye et al. (2015)]. Identification and estimation of sharing rule empirically is not, however, straightforward. Typically, consumption expenditure surveys report data at the household level and hence individual level information is not available. Various methods have so far been proposed to resolve this problem (see next section for a discussion on this).

In this chapter, we attempt to examine gender bias in the allocation of resources to children's education within a household in the rural and the urban sectors of West Bengal through a collective household model. We use the 68th round (2011-12) household level consumption expenditure data provided by the Indian National Sample Survey Office (NSSO). First, we estimate the sharing rule parameters (that determine the apportionment of resources between members) within a household. Next, we use the sharing rule information to categorise two types of households where in one type the husband has a higher share of resources (higher bargaining power) and the ones in which the wife has a higher share of resources. Finally, we estimate the budget share of education expenditure for each group of households and examine gender bias through testing of equality of allocation of expenditure for male and female children. To the best of our knowledge, this is the first attempt to study the presence of gender bias in the education expenditure using the collective model and estimated resource share (bargaining power) of the husband and wife in the Indian context.

The contribution of this chapter is twofold. First, we suggest a modification of the Quadratic Almost Ideal Demand System (QUAIDS) (Banks, Blundell, and Lewbel (1997)) functional from to retrieve the information of sharing rule.[34] Second, we assess gender bias in education expenditure in West Bengal applying the obtained information of sharing rule. While in most of the studies the effect of mother's bargaining power is generally assessed with respect to children's welfare, in this chapter, we also include young college going adults along with the

[34]Use of QUAIDS in the collective set up can also be found in Browning, Chiappori and Lewbel (2013) and Attanasio, Battinstin and Mesnard (2012).

children in the assessed group, by considering the age group of 5-29 years as 'students', as has been done in the previous chapter.

The plan of the chapter is as follows: Section 4.2 provides a brief account of the relevant literature; Section 4.3 specifies the model; Section 4.4 describes the data and methodology; Section 4.5 presents the results and discussion; and finally Section 4.6 concludes the paper.

4.2 A Brief Account of Relevant Literature

Three approaches have been suggested in the literature for identification of sharing rules:

(i) Collect extensive intra-household (assignable) consumption data (Cherchye, De Rock and Vermeulen (2011), Menon, Pendakur and Perali (2012)) to obtain the sharing rule. However, this method has a problem in dealing with measurement of the fraction of shared goods that are consumed by each individual, and fails to deal with potential externalities within a household;

(ii) Make additional identifying assumptions to retrieve the sharing rule from household level data. These assumptions include **(ii-a)** restricting individual preferences across households of different compositions (Browning Chiappori, Lewbel (2013), Couprie (2007), Lewbel and Pendakur (2008), Bargain and Donni (2009, 2012), Couprie, Peluso and Trannoy (2010), Lise and Seitz (2011)), **(ii-b)** restricting sharing rules and assuming preference similarities (Dunbar, Lewbel, and Pendakur (2014)),**(ii-c)** restricting sharing rules and using distribution factors (Dunbar, Lewbel, and Pendakur (2017)); and **(ii-d)** assuming that the unobservable sharing rule interacts with individual incomes *a la* Barten (Barten 1964, Perali 2008) and that at least one assignable good, or, equivalently, two exclusive goods[35] are observable (Aria, Atella and Peralli (2004)); and

(iii) Obtain bounds on shares (Cherchye, De Rock and Vermeulen (2011), Cherchye, De Rock, Lewbel and Vermeulen (2015) to retrieve the sharing rule from household level data.

In the present chapter we follow the line along (ii-d) to estimate the sharing rule in the state of West Bengal, India. The underlying assumption is that parents first agree on the sharing rule, determining the direction and amount of transfer between members and then they maximise their individual egoistic utility functions subject to their private budget constraint. After having estimated the sharing rule and the budget shares, the contribution of parents (individual) to expenditure on education of children is determined.

4.3 Model Specification

We proceed along the line of the collective household model of Chiappori (1988, 1992) and Dunbar, Lewbel and Pendakur (2014), which assumes that intra-household decisions are Pareto efficient.

We consider a family with two adults (husband (m), wife (f)) and students (c). Adults are the earning members of the household, they take decisions. Since students do not earn, they have to depend on adults to fulfill their needs and hence, we are not explicitly assuming any utility function for the students separately.

The household consumes K types of goods, the consumption vector is denoted by $x = (x_1,$ $x_2 \dots\dots\dots. x_K)$ and the corresponding market price vector is denoted by $p = (p_1,\ p_2 \dots\dots\dots. p_K)$.

[35]A good is exclusive when one identifiable member of the household only consumes a strictly private good. A good is assignable when each member of the household consumes a strictly private good in observed proportions.

The consumption bundle x takes into account the economies of scale (private and joint consumption) of each item. As the minimum informational requirement for identification of sharing rule is the observability of at least one assignable good or equivalently two exclusive goods (Vermeulen, 2002), we have considered 'clothing' as one assignable good here. While 'clothing' is an assignable good, 'male clothing' and 'female clothing' separately are exclusive goods. So, x consists of male clothing (x_m), female clothing (x_f), and other items which are clubbed together as composite good (q) (following Arias, Atella and Perali (2004)), which consists of public goods, private non-assignable goods (food) and education.[36]

A Pareto optimal allocation between the two individuals can be obtained by the following maximisation problem:

$$\text{Max}_{x_m, x_f, q}\, \mu(p, y, d_{dis}) V^m(x_m, x_f, q, g_{dem}) + (1 - \mu(p, y, d_{dis})) V^f(x_m, x_f, q, g_{dem})$$

$$\text{s.t. } p'x = y \quad \text{and} \quad x = (x_m, x_f, q), \tag{4.1}$$

where, y is the total expenditure of the household. $\mu(p, y, d_{dis})$ is the Pareto weight (bargaining power of the husband *m*). It is a function of the factors (d_{dis}) that do not affect the preferences and the budget constraint of a household. These factors are called *distribution factors* (Bourguignon, Chiappori and Browning, 2009)[37] and are independent of the demographic factors g_{dem}. The position of the maximising solution on the Pareto frontier is determined by the bargaining power or the Pareto weight ($\mu(.)$), it does not have any effect on the preferences of the household members. Further, the Pareto weight is a function of prices (p) and total expenditure (y) of the household as these factors also determine the power dynamics within the household. Each adult's preference depends on the consumption of own and other partner's exclusive good,

[36] The list of goods considered as male and female exclusive goods are given in Appendix IV, Table A4.1.

[37] The distribution factors are also called Extra Environmental Factors by McElroy (1990).

and consumption of the composite good. Here, q appears in both adults' preferences, as expenditure on the composite good q is mostly done for joint consumptions in the household. We now assume egoistic preferences $V^m(x_m, g_{dem})$ and $V^f(x_f, g_{dem})$, in which an individual member's preference depends on his/her own consumption.[38]

The Pareto efficient household allocation problem (4.1) can now be shown to be equivalent to the existence of a function $\phi(y, p, d_{dis})$, such that x_m and x_f are solutions to the maximisation problem (Chiappori, Fortin and Lacroix, 2002):

$$Max_{x_m} V^m(x_m, g_{dem}) \text{ s.t. } P_m'.x_m \leq \phi(y, p, d_{dis}) \tag{4.2}$$

$$\text{and } Max_{x_f} V^f\left(x_f, g_{dem}\right) \text{ s.t. } P_f'.x_f \leq y - \phi(y, p, d_{dis}), \tag{4.3}$$

where $\phi(y, p, d_{dis})$ is the husband's resource allocation, $(y - \phi(y, p, d_{dis}))$ is the wife's resource allocation and P_m, P_f are the price vectors corresponding to the male and female exclusive goods, respectively,

Solving (4.2) and (4.3) we get the following budget share functions (s_m and s_f) of the exclusive goods, under the condition that the individual is living alone and maximises utility subject to his own share of income.

$$\left.\begin{array}{l} s_m = \frac{x_m(P_m, \phi(y,p,d_{dis}), g_{dem})}{\phi(y,p,d_{dis})}: \text{ budget share for male exclusive goods} \\ s_f = \frac{x_f(P_f, y-\phi(y,p,d_{dis}), g_{dem})}{y-\phi(y,p,d_{dis})}: \text{ budget share for female exclusive goods} \end{array}\right\} \tag{4.4}$$

To estimate these equations we make the following three assumptions.

[38] Donni (2009) pointed out that in case of private goods, the Pareto weight is an equivalent representation of the distribution of power within the household, but the presence of public good consumption leads to more critical relationship between Pareto weight and the distribution of power within the household. We, therefore, assume egoistic preference. Note, however that $\phi(y, p, d_{dis})$ is an apportionment of total expenditure, which includes all items.

ASSUMPTION 1:

The estimation of equation (4.5) requires the information on price variations, as the household budget share functions (s_m and s_f) involve household shadow price vector. The household consumption expenditure data provided by NSSO do not have any price information. This problem is mitigated in two ways. First, the budget share functions may be considered in the Engel curve form, which means that the price variation is not exploited.[39] Second, the market prices of the private goods are fixed, and two individuals buying exactly the same private goods are facing a single market price which cannot be altered by their demand. Hence, the market price of a private good would not affect the resource share function as such as the consumers are price takers.

Following this, we have estimated the resource share function in the Engel curve form (underlying the QUAIDS), that is, it is independent of the price variation, but dependent on the household total expenditure.

Modification of QUAIDS for collective household framework

The general functional form of the demographic variable augmented QUAIDS (Ray, 1983) in the unitary framework is as follows

$$s_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij}.\ln(p_j) + \left(\beta_i + \acute{\eta}_i g_{dem}\right)\ln\left(\frac{y}{\bar{m}_0(g_{dem})a(p)}\right) + \frac{\lambda_i}{b(p)c(p,g)}[\ln\left(\frac{y}{\bar{m}_0(g_{dem})a(p)}\right)]^2, \tag{4.6}$$

where, g_{dem} is the vector of demographic characteristics,

$$c(p, g_{dem}) = \prod_{j=1}^{n} p_j^{\acute{\rho}_j g_{dem}}$$

[39] Dunbar, Lewbel and Pendakur (2014), Donni(2009) and Lewbel, Pendakur (2008) have used this technique.

$$ln(a(p)) = \alpha_0 + \sum_{i=1}^{n} \alpha_i . ln\ (p_j) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} lnp_i . lnp_j$$

$$b(p) = \prod_{i=1}^{n} p_i{}^{\beta_i}$$

$$\lambda(p) = \sum_{i=1}^{k} \lambda_i ln\ (p_i),$$

and $\bar{m}_0(g_{dem})$ is a function of g_{dem}.

The underlying budget shares without any price variation is given by

$$s_i = \alpha_i + (\beta_i + \acute{\eta_i} g_{dem}) \ln\left(\frac{y}{m_0(g_{dem})}\right) + \lambda_i[\ln\left(\frac{y}{m_0(g_{dem})}\right)]^2. \quad (4.7)$$

Here, for $m_0(g_{dem})$, which can be interpreted as the household equivalence scale[40], we take the 'Square root scale' (square root of the household size) published by OECD 2011, 2008 (Organisation for Economic Co-operation and Development).

According to Ray (1983), $c(p, g)$ would be chosen to be unity if we consider a single budget share function. Hence, the coefficient of the quadratic part ($[\ln\left(\frac{y}{m_0(g_{dem})}\right)]^2$) is taken as λ_i.

To formulate the collective model, we assume that the husband and the wife individually has a QUAIDS budget share structure. The demand functions for exclusive goods are given by

$$\left.\begin{aligned} S_m &= \phi(y, d_{dis})\left(\alpha_m + (\beta_m + \acute{\eta_m} g_{dem})\ln\left(\frac{\phi(y,d_{dis})}{\sqrt{hhsize}}\right) + \lambda_m\left(\ln\left(\frac{\phi(y,d_{dis})}{\sqrt{hhsize}}\right)\right)^2\right) \\ S_f &= (y - \phi(y, d_{dis}))\left(\alpha_f + (\beta_f + \acute{\eta_f} g_{dem})\ln\left(\frac{y-\phi(y,d_{dis})}{\sqrt{hhsize}}\right) + \lambda_f\left(\ln\left(\frac{y-\phi(y,d_{dis})}{\sqrt{hhsize}}\right)\right)^2\right) \end{aligned}\right\}. \quad (4.8)$$

It may be noted that direct estimation of budget share functions defined above is not possible unless we have identified the sharing rule. To address this problem, we assume a separability property (see ASSUMPTION 2 below) adopted by Lise and Seitz (2011), Lewbel and Pendakur

[40] Ray (1983) used a linear function of the form $\bar{m}_0(g) = 1 + \rho' g$

(2008), Bargain and Donni (2012), Cherchye, De Rock, Lewbel and Vermeulen (2015) and Menon, Pendakur and Perali (2012).

ASSUPMTION 2:

$$\phi(y, d_{dis}) = y.\psi(d_{dis}). \tag{4.9}$$

This assumption (separability condition) states that the resource share depends on total expenditure but the scaling function $\psi(d_{dis})$ does not. In this method, $\psi(d_{dis})$ is considered to be a scaling function that determines the size and direction of the symmetric transfer occurring between husband and wife in the household. Hence, the wife gets $y(1-\psi(d_{dis}))$.

Using this assumption, we obtain the demand functions of male and female exclusive goods as

$$\left.\begin{aligned} S_m &= S_m(y.\psi(d_{dis}), g_{dem}) \\ S_f &= S_f\big(y\big(1-\psi(d_{dis})\big), g_{dem}\big). \end{aligned}\right\} \tag{4.10}$$

ASSUMPTION 3:

The functional form of the scaling function $\psi(d_{dis})$ is assumed to have a logistic form.

$$\psi(d_{dis}) = \frac{e^{d_{dis}{}'\gamma}}{1+e^{d_{dis}\gamma}}. \tag{4.11}$$

Given the above, the demand functions of male and female exclusive goods are functions of the share of total expenditure for male member and female member, demographic factors and distribution factors. The functional form is as follows,

$$\left.\begin{aligned} S_m &= \big(y.\psi(d_{dis})\big)\left(\alpha_m + (\beta_m + \eta_m' g_{dem})\ln\left(\frac{y.\psi(d_{dis})}{\sqrt{hhsize}}\right) + \lambda_m\left(\ln\left(\frac{y.\psi(d_{dis})}{\sqrt{hhsize}}\right)\right)^2\right) + \varepsilon_1 \\ S_f &= y.(1-\psi(d_{dis}))\left(\alpha_f + (\beta_f + \eta_f' g_{dem})\ln\left(\frac{y(1-.\psi(d_{dis}))}{\sqrt{hhsize}}\right) + \lambda_f\left(\ln\left(\frac{y(1-\psi(d_{dis}))}{\sqrt{hhsize}}\right)\right)^2\right) + \varepsilon_2 \end{aligned}\right\}, \tag{4.12}$$

The functional form of $\psi(d_{dis})$ is chosen such that $\psi(d_{dis})$ remains bounded between 0 and 1. This functional form has also been used by other authors (Browning, Chiappori and Lewbel, 2013).

4.4 Data and Methodology

This section consists of three parts. First, a brief description of the data set used in this chapter is provided in Section 4.4.1. Second, the estimation procedure is illustrated in Section 4.4.2. Third, the endogeneity problem is addressed in Section 4.4.3.

4.4.1 Data Description

For our analysis we have used the 68^{th} round (July 2011-June 2012) household level consumption expenditure data provided by the Indian National Sample Survey Office (NSSO). The 68^{th} round consumption expenditure data provides information on household level expenditure on different items like food, tobacco, alcohol, fuel, medical, education, recreation etc. with some specific household and individual demographic features for all states for both rural and urban sectors. For some items, e.g., for clothing, male and female items are separately reported.

As mentioned earlier, we focus on the state of West Bengal, where information on 6324 households (3575 in rural, 2749 in urban) are reported. For our analysis, we only use data with the following family structure: husband, wife, at least one male and one female offspring belonging to the age group 5-29 years. This leaves us with a sample size of 1458 households (973 in rural, 485 in urban). The descriptive statistics are provided in Table 4.1.

4.4.2 Empirical Analysis

The empirical analysis has three parts:

1. Estimation of the sharing rule: Since expenditure on public goods does not play a role in 'sharing' resources, we start with the estimation of equation (4.12) which is a system of demand functions of male and female clothing. The estimation strategy is nonlinear Seemingly Unrelated Regression (SUR) model as the error terms ε_1 and ε_2 may vary across these two equations. The estimates of $\psi(d_{dis})$ are obtained from this estimation.
2. The predicted values of scaling functions are used to categorise the households into two groups: one where husband's share of resources (bargaining power) is higher and the other where wife's bargaining power is higher.
3. To capture gender bias, an NLS regression of budget share for 'education expenditure' is done separately for two types of households. The differences between the estimated coefficients of fraction of household size for male children and that of female children are tested for the direction of gender bias for two types of households mentioned above.

In our data the expenditures on clothing and education are reported in terms of expenditure in last 365 days. These are adjusted to monthly data to match with total expenditure, which is reported as monthly per capita total expenditure. Since we are not considering any price variation, we have clubbed the male clothing items into one group and female clothing items into another group.

The demographic vector g_{dem} contains the following information: proportion of male and female household members belonging to age group 5-29 years (*Child_m* and *Child_f*)[41]. For the scaling function $\psi(d_{dis})$, d_{dis} contains the distribution factors- Age gap of husband and wife (*Age_gap* = Age of husband-Age of wife), Education levels of husband and wife (*Edu_m* and *Edu_f*).[42]

Now, after estimation of S_m and S_f, the estimated coefficients of distribution factors ($\hat{\gamma}$) are used to obtain the values and the standard errors of the scaling function using the functional form (4.11). The standard errors are obtained by Delta method.

Since $\mu(p, X, d_{dis})$ and $\phi(y, p, d_{dis})$ are bijective, these functions are equivalent representation of distribution of decision making power of the household members in case of pure private good scenario (Donni, 2009). Hence, after obtaining the estimated values of $\psi(d_{dis})$ we can use the information to categorise the households in two types where in one type husband would be enjoying higher bargaining power ($\widehat{\psi(d_{dis})} > 1 - \widehat{\psi(d_{dis})}$) and in the other type wife would be enjoying higher bargaining power($\widehat{\psi(d_{dis})} < 1 - \widehat{\psi(d_{dis})}$).[43]

We now examine the direction and magnitude of gender bias in education expenditure for these two types of households separately. Note that as education is a part of public good, the budget share here depends on total expenditure *y*. The functional form of the budget share of education expenditure (s_e) is formulated in the QUAIDS framework as:

---

[41]The young adults in this group are college going students and those pursuing higher studies [Majumder and Mitra (2017)]. Note that with the increase in average age at marriage, the probability of selection bias owing to daughters getting married and leaving home is expected to be small. In the rural sector, there are 51(5.24%) households with the students (male and female) belonging to the age group 20-29 years. In the urban sector, the number is 97(20%) with similar specifications.

[42] The education level is measured by years of schooling.

[43]Equality of $\widehat{\psi(d_{dis})}$ and $1 - \widehat{\psi(d_{dis})}$was not found in our data.

$$s_e = \alpha_e + \left(\beta_e + \theta_{em}Child_m + \theta_{ef}Child_f + \eta'_e g_e\right)\ln\left(\frac{y}{\sqrt{hhsize}}\right) + \lambda_e\left[\ln\left(\frac{y}{\sqrt{hhsize}}\right)\right]^2 + \varepsilon_3, \quad (4.13)$$

where g_e includes *Edu_m*, *Edu_f* and *hdage* (age of the father). After having estimated the above equation using NLS, the direction of gender bias is tested using the hypotheses.

$$\left.\begin{array}{l} H_0 : \theta_{em} - \theta_{ef} = 0, \\ H_1 : \theta_{em} - \theta_{ef} > 0 \end{array}\right\}. \quad (4.14)$$

Non-rejection of H_0 would imply that the effects of increase in the numbers of male and female children on education expenditure are similar. That is, there is no significant difference due to the sex of the child. However, rejection of H_0 means that there is significant gender bias in favour of male children in households.

4.4.3 Endogeneity Problem

There is likely to be an endogeneity problem as the error terms in equation (4.12) includes taste shocks, error in recall period[44], which are correlated with the total expenditure (*y*). Further, total expenditure itself involves measurement error problem. Attanasio and Lechene (2014), Dunbar, Lewbel and Pendakur (2014), Browning and Chiappori (1998) noted this problem and suggested using traditional instruments like wage income, net income and wealth measures. In our dataset the only relevant instrument is 'land holding' as a measure of wealth. We have adopted a method suggested by Davidson and Mckinnon (1993). Here, the total expenditure is regressed on all exogenous variables which includes the demographic (g_{dem}) and distribution factors (d_{dis}) and land holding (*LAND*). Then, the residual of this regression (*resi*) is used as a separate regressor in SUR estimation of both equations in equation (4.11). We have repeated the estimation procedure

[44] The data are collected by asking the household members to recall what they spent on specific items.

explained in Section 4.2 after this correction. The results are reported in Appendix IV, Tables A4.2-A4.7.

4.5 Results and Discussion

The estimation of SUR model for the demand functions of the male and female exclusive goods is carried out for both sectors. The results of the rural and the urban sectors are reported in Table 4.2 and Table 4.3, respectively. The values of R^2 are 0.2849 and 0.3456 for the rural sector, and these values are 0.3011 and 0.3547 for the urban sector. We first discuss the results for the rural sector (Table 4.2), which will be followed by a discussion for the urban sector (Table 4.3).

In the rural sector, the coefficients of *ln(share_m)*[45] and its squared term are significant but positive and negative, respectively, for the male clothing. Similar pattern is observed for *ln(share_f)* and its squared term for female exclusive goods. The variables included as the demographic factors including the demographic constant (β_i) have significant coefficients. For male clothing, the coefficient of *Child_f* is negative (non-significant), and that of *Child_m* is positive. Hence, with increase in the number of male child the expenditure on male clothing increases, but increase in the number of female child has no significant effect. For female clothing, interestingly, both coefficients are significant, with a positive sign for *Child_f* and a negative sign for *Child_m*, implying a cut down on the budget of female clothing with increase in number of male child. The variables considered as distribution factors are all significant. It is interesting to note that an educated husband has a significant negative impact on the expenditure of female clothing and an educated wife has a similar impact on male clothing, while the coefficients are positive for their own expenditures. The points to the fact that a more educated

[45] *share_m*= $y.\psi(d_{dis})$ and *share_f*= $y.(1-\psi(d_{dis}))$.

partner, while retaining his/her budget, can influence a budget cut for the other partner. The age gap between husband and wife (*Age_gap*) has negative coefficient for male exclusive goods and positive coefficient for female clothing. Thus, wives with older husbands have more influence on budget allocation. This finding is corroborated by the following studies.A study by Li, L., & Wu, X. (2011) finds that the wife's decision becomes less important when she becomes as old as her husband. Mabsout and Van (2010), using Confirmatory Categorical Factor Analysis (CCFA) in Ethiopia, observed a positive impact of higher age gap between spouses on the bargaining power of the wife. Further, Calvi, Lewbel and Tommasi (2017) also found that higher age gap leads to higher bargaining power of wife in the household.

In the urban sector, the 'resource share' coefficients are of similar nature as in case of the rural sector. All the demographic factors, except for *Child_f* also have similar impacts. Here the coefficient of *Child_f* is positive and significant for male clothing. A possible reason could be that male clothing includes certain unisex items (like t-shirts, trousers etc.), which can also be consumed by females in the urban sector unlike in the rural sectors of India. In case of distribution factors, the education level of husband has a positive impact on the budget of both male and female clothing, but the education level of wife has significant (positive) effect only on female clothing. This means that while an educated husband encourages the female members to spend on their requirements unlike that in the rural sector, an educated wife is concerned only with her expenditure. The pattern of signs of *Age_gap* is similar to those in the rural sector.

From the estimated coefficients of the distribution factors in the demand function, the predicted values of the scaling function $\left(\psi(d_{dis})\right)$ are obtained. The average values of the scaling function for the two sectors are reported in Table 4.4. The values reveal that on an average, 68.14% and 63.98% of the total expenditure are enjoyed by the husband in the rural and the urban sectors, respectively. The scaling function is statistically significant, as is evident from the computed standard errors.

Using the information on household level scaling functions, the households are categorized into two groups for each sector. The first type is where the husband has a higher bargaining power, and the second type is where the wife has a higher bargaining power. In Table 4.5 we have reported the percentage of the sample of households in these two categories. It is evident from the table that higher bargaining power is enjoyed by the husband in 87.72% of the sample of households in the rural sector and in 69.01% of the sample of households in the urban sector.

The estimates of the parameters of the budget share on education expenditure are reported in Table 4.6 for both sectors. The values of R^2 range between 0.1577 and 0.2011. It can be seen that all the coefficients of the explanatory variables are statistically significant. The coefficients of *ln(y)* and its squared variable are positive and negative, respectively, in both types of households for both sectors. The coefficients of *Child_m* and *Child_f* are positive and negative, respectively, when the husband has a higher bargaining power in both sectors. However, the coefficients of these two explanatory variables are positive when wife has higher bargaining power in both sectors. This implies that with an increase in the numbers of children, irrespective of whether it is a male or a female child, the expenditure on education will rise when the wife is the decision maker in the household. But when the husband has a higher bargaining power, the expenditure on education will rise only with an increase in the number of male child. This clearly points to presence of gender bias towards male children for these households. The coefficients of *Edu_m* are positive and significant in both types of households in the rural sector and in father-dominated households in the urban sector. On the other hand, the coefficients of *Edu_f* are positive and significant only for mother-dominated households in the rural sector and father-dominated households in the urban sector. Interestingly, none of the education coefficients is significant in mother-dominated households in the urban sector. *hdage* (age of the father) has significant positive effect in both types of households in the rural sector and in father-dominated households in the urban sector.

To statistically establish the presence of gender bias, the *p*-values for the test of hypothesis (4.14) are reported in Table 4.7. Clearly, when the husband has higher bargaining power, the difference between the coefficients of *Child*_m and *Child_f* is positive and the *p*-values are much smaller. Hence the null hypothesis is rejected at 1% level of significance in both sectors. That is, in both sectors there is a pro- male bias in the household where the husband takes the decisions. On the other hand, when the wife has higher bargaining power, the *p*-values are much higher and the null hypothesis cannot be rejected. This implies that the difference between the coefficients of *Child_m* and *Child_f* is not significantly different from zero. In other words, there is no gender bias in education expenditure in households where wife has higher bargaining power.

After correction for endogeneity (as discussed in Section 4.4.3), the results are reported in Appendix IV, Tables A4.2 – A4.7. Except for some minor changes, the results remain qualitatively the same. It is evident from Table A4.2 and Table A4.3 that the values of R^2 are improved than before in both sectors. Further, the coefficients of the variable *resi* are significant and positive which corroborates the existence of the endogeneity problem. However, from Table A4.4, the resource shares are underestimated compared to Table 4.4. 64.83% and 60.18% of the total expenditure are enjoyed by the husband in the rural and the urban sectors, respectively. Hence, the percentages of households with the husband having higher bargaining power, is increased in both sectors (see Table A4.5). In Table A4.6, the regression results of budget share of education expenditure are reported. Here also, the values of R^2 are improved from before in both sectors. The coefficient of *Edu_f* is significant and positive in the rural sector for the households with the husband having higher bargaining power. Further, the coefficients of the variable *resi* are significant and positive. In Table A4.7, the results of the hypotheses testing are reported where again it is established that there exists a pro-male gender bias in the households where the husband is the decision maker and there is no gender bias where wife is the decision maker.

4.6 Conclusion

In this chapter, we have tried to estimate the sharing rule of total household expenditure between couples using a modified version of the QUAIDS and the 68th round of household level consumer expenditure data provided by the NSSO. Using the sharing rule information, we have examined gender bias in spending on education for children in the households of West Bengal. We have also addressed the endogeneity problem.

The findings that emerge are: pro male bias in education expenditure is observed in households where husbands enjoy higher bargaining power. However, when the wife is the decision maker, no gender bias in education expenditure is observed. Even after addressing the endogeneity problem, the findings remain unaltered. This corroborates the findings of Balatchandirane (2003).

Table 4.1: Descriptive Statistics of Observed Variables

Variables	Variable Names	Rural		Urban	
		Mean	Standard deviation	Mean	Standard deviation
Total expenditure (Rs.)	$\boldsymbol{Y}$	5123.958	5323.971	10752.07	13544.07
Expenditure on clothing (male) (Rs.)	$\boldsymbol{S_m}$	988.52	184.92	1345.09	289.09
Expenditure for clothing (female) (Rs.)	$\boldsymbol{S_f}$	778.32	302.67	1289.49	577.88
Share of expenditure for education	$\boldsymbol{s_e}$	0.119	0.039	0.178	0.085
Household size	***Hhsize***	6.082	9.089	4.011	5.056
Proportion of members belonging to age group 5-29 (male)	***Child_m***	0.651	0.604	0.506	0.103
Proportion of members belonging to age group 5-29 (female)	***Child_f***	0.252	0.809	0.457	0.156
Education level of husband (years of schooling)	***Edu_m***	10.19	6.807	15.079	3.072
Education level of wife (years of schooling)	***Edu_f***	7.98	6.233	10.193	5.403
Age gap between husband and wife (years)	***Age_gap***	11.69	14.901	6.343	10.908
Age of the husband (years)	***Hdage***	36.04	12.322	45.035	7.609

Table 4.2: SUR Regression Results of Male and Female Clothing (Rural Sector)

	Explanatory Variables	Male Clothing (S_m)	Female Clothing (S_f)
Demographic Factors	***Child_m***	0.007 (0.001)***	-0.078 (0.005)***
	Child_f	-0.032 (0.037)**	0.002 (0.000)***
	Demographic Constant (β_i)	0.056 (0.000)***	0.216 (0.000)***
Resource share for male	***ln(share_m)***	0.055 (0.000)***	NA
	[ln(share_m)]^2	-0.002 (0.000)***	NA
Resource share for female	***ln(share_f)***	NA	0.021 (0.001)***
	[ln(share_f)]^2	NA	-0.013 (0.000)***
Distribution factors	***Edu_m***	0.204 (0.001)***	-0.338 (0.067)**
	Edu_f	-0.131 (0.005)***	0.063 (0.034)***
	Age_gap	-0.035 (0.072)*	0.042 (0.001)***
	Constant	0.275 (0.001)***	0.435 (0.000)***
	R^2	0.2849	0.3456
	No. of observations	973	973

Figures in the parentheses are the *p*-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Note: *share_m*= $y.\psi(d_{dis})$ and *share_f*= $y.(1-\psi(d_{dis}))$.

Table 4.3: SUR Regression Results of Male and Female Clothing (Urban Sector)

	Explanatory Variables	Male Clothing (S_m)	Female Clothing (S_f)
Demographic Factors	***Child_m***	0.004 (0.001)***	-0.029 (0.003)***
	Child_f	0.107 (0.025)**	0.059 (0.000)***
	Demographic Constant (β_i)	0.030 (0.017)**	0.029 (0.011)**
Resource share for male	***ln(share_m)***	0.029 (0.014)**	NA
	[ln(share_m)]^2	-0.001 (0.000)***	NA
Resource share for female	***ln(share_f)***	NA	0.205 (0.061)*
	[ln(share_f)]^2	NA	-0.012 (0.000)***
Distribution factors	***Edu_m***	0.172 (0.031)**	0.021 (0.001)***
	Edu_f	-0.003 (0.089)*	0.192 (0.005)***
	Age_gap	-0.084 (0.002)***	0.049 (0.003)***
	Constant	0.156 (0.001)***	0.198 (0.000)***
	R^2	0.3011	0.3547
	No. of observations	485	485

Figures in the parentheses are the *p*-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 4.4: Descriptive Statistics of the Estimated Values of the Resource Share of the Husband: $\psi(\widehat{d_{\iota s}})$

Statistics	**Rural Sector**	**Urban Sector**
Mean	0.6814	0.6398
Standard error	0.0950	0.0930
Minimum	0.3986	0.4011
Maximum	0.9568	0.8896

Table 4.5: Comparison of Contributions by Parents

SECTOR	Percentage (%) of Sample Households	
	Father dominated $\boldsymbol{\psi(\widehat{d_{ıs}}) > 1 - \psi(\widehat{d_{ıs}})}$	**Mother dominated** $\boldsymbol{\psi(\widehat{d_{ıs}}) < 1 - \psi(\widehat{d_{ıs}})}$
RURAL	87.72	12.28
URBAN	69.01	30.99

Table 4.6: Regression Result of Share of Education Expenditure (s_e)

Explanatory Variables	Rural		Urban	
	Father dominated	**Mother Dominated**	**Father dominated**	**Mother dominated**
ln(y)	0.611 (0.058)*	0.244 (0.001)***	1.002 (0.033)**	0.209 (0.001)***
[ln(y)]²	-0.028 (0.012)**	-0.065 (0.015)**	-0.031 (0.051)*	-0.102 (0.023)**
Child_m	0.045 (0.008)***	0.029 (0.019)**	0.049 (0.002)***	0.089 (0.069)*
Child_f	-0.039 (0.011)**	0.019 (0.003)***	-0.028 (0.001)***	0.039 (0.028)**
Edu_m	0.051 (0.024)**	0.092 (0.067)*	0.019 (0.002)***	0.059 (0.002)***
Edu_f	0.093 (0.205)	0.029 (0.061)*	0.003 (0.002)***	0.078 (0.008)***
hdage	0.022 (0.001)***	0.023 (0.005)***	0.042 (0.015)**	-0.037 (0.209)
Constant	0.911 (0.001)***	1.302 (0.003)***	-0.992 (0.011)**	0.706 (0.002)***
R²	0.1577	0.1744	0.1863	0.2011
No. of observations	853	120	334	151

Figures in the parentheses are the *p*-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 4.7: The *p*-values for Testing the Hypotheses of Gender Bias

Hypotheses Testing	p-values (Rural)		p-values (Urban)	
$\boldsymbol{H_0: \theta_{em} - \theta_{ef} = 0}$	**Father dominated**	**Mother dominated**	**Father dominated**	**Mother dominated**
$\boldsymbol{H_1: \theta_{em} - \theta_{ef}} > 0$	0.002	0.359	0.011	0.194
Decision	Rejecting H_0	Non-rejection H_0	Rejecting H_0	Non-rejectionH_0

Note: Level of significance: 5%

APPENDIX IV

Table A4.1: List of Male and Female Exclusive Goods

MALE EXCLUSIVE GOODS	1. DHOTI 2. SCHOOL/COLLEGE UNIFORM, MALES 3. KURTA/PAJAMA, MALES 4. SHIRTS,T-SHIRTS 5. SHORTS,TROUSERS, BERMUDA 6. LUNGI
FEMALE EXCLUSIVE GOODS	1. SAREE 2. SCHOOL/COLLEGE UNIFORM, FEMALES 3. KURTA/PAJAMA, FEMALES 4. PAJAMA, SALWAR 5. FROCKS, SKIRTS 6. BLOUSE, DUPATTA, SCARF

Table A4.2: Endogeneity Corrected SUR Regression Results of Male and Female Clothing (Rural Sector)

	Explanatory Variables	Male Clothing (S_m)	Female Clothing (S_f)
Demographic Factors	***Child_m***	0.036 (0.000)***	-0.001 (0.001)***
	Child_f	-0.002 (0.001)***	0.044 (0.023)**
	Demographic Constant (β_i)	-0.021 (0.151)	0.062 (0.097)*
Resource share for male	***ln(share_m)***	0.019 (0.014)**	NA
	[ln(share_m)]^2	-0.007 (0.001)***	NA
Resource share for female	***ln(share_f)***	NA	0.061 (0.001)***
	[ln(share_f)]^2	NA	-0.073 (0.025)**
Distribution factors	***Edu_m***	0.166 (0.002)***	-0.038 (0.278)
	Edu_f	-0.013 (0.005)***	0.180 (0.001)***
	Age_gap	-0.077 (0.002)***	0.043 (0.000)***
Endogeneity Check	***Resi***	0.006 (0.002)***	0.023 (0.000)***
	Constant	0.124 (0.001)***	1.009 (0.000)***
	R^2	0.4179	0.5140
	No. of observations	973	973

Figures in the parentheses are the p-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table A4.3: Endogeneity Corrected SUR Regression Results of Male and Female Clothing (Urban Sector)

	Explanatory Variables	Male Clothing (S_m)	Female Clothing (S_f)
Demographic Factors	***Child_m***	0.073 (0.000)***	-0.004 (0.096)*
	Child_f	0.029 (0.009)***	0.019 (0.001)***
	Demographic Constant (β_i)	0.051 (0.162)	-0.148 (0.001)***
Resource share for male	***ln(share_m)***	0.054 (0.000)***	NA
	$[ln(share\_m)]^2$	-0.002 (0.051)*	NA
Resource share for female	***ln(share_f)***	NA	0.146 (0.001)***
	$[ln(share\_f)]^2$	NA	-0.072 (0.000)***
Distribution factors	***Edu_m***	0.207 (0.023)**	0.252 (0.000)***
	Edu_f	-0.133 (0.028)**	0.049 (0.011)**
	Age_gap	-0.037 (0.017)**	0.025 (0.001)***
Endogeneity Check	***Resi***	0.029 (0.021)**	0.048 (0.015)**
	Constant	0.258 (0.001)***	0.379 (0.000)***
	R^2	0.4863	0.4483
	No. of observations	485	485

Figures in the parentheses are the p-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table A4.4: Descriptive Statistics of the Estimated Values of the Resource Share of the Husband (After Endogeneity correction)

Statistics	Rural Sector	Urban Sector
Mean	0.6483	0.6018
Standard error	0.0530	0.0930
Minimum	0.3029	0.3968
Maximum	0.9937	0.8064

Table A4.5: Comparison of Contributions by Parents

SECTOR	Percentage (%) of Households	
	Father dominated $\psi(\widehat{d_{\imath s}}) > 1 - \psi(\widehat{d_{\imath s}})$	Mother dominated $\psi(\widehat{d_{\imath s}}) < 1 - \psi(\widehat{d_{\imath s}})$
RURAL	79.34	20.66
URBAN	65.82	34.18

Table A4.6: Regression Result of Share of Education Expenditure (s_e)

Explanatory Variables	Rural		Urban	
	Father dominated	Mother Dominated	Father dominated	Mother dominated
ln(y)	0.024 (0.000)***	0.046 (0.011)**	0.037 (0.034)**	0.067 (0.001)***
[ln(y)]^2	-0.011 (0.047)*	-0.023 (0.032)**	-0.017 (0.007)***	-0.015 (0.024)**
Child_m	0.034 (0.079)*	0.204 (0.022)**	0.058 (0.007)**	0.039 (0.005)***
Child_f	-0.043 (0.046)*	0.057 (0.031)**	-0.095 (0.076)*	0.109 (0.019)**
Edu_m	0.008 (0.001)***	0.028 (0.002)***	0.307 (0.011)**	0.205 (0.089)*
Edu_f	0.101 (0.016)***	0.067 (0.003)***	0.077 (0.003)***	0.119 (0.001)***
Hdage	0.029 (0.011)***	0.019 (0.015)**	0.055 (0.001)***	0.048 (0.003)***
Resi	0.004 (0.012)**	0.072 (0.021)**	0.022 (0.031)**	0.038 (0.021)**
Constant	0.711 (0.001)***	0.603 (0.005)***	-0.221 (0.000)***	0.502 (0.000)***
R^2	0.1803	0.1906	0.1785	0.1847
No. of observations	772	201	319	166

Figures in the parentheses are the *p*-values.
***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table A4.7: The p-values for Testing the Hypotheses of Gender Bias

Hypotheses Testing	p-values (Rural)		p-values (Urban)	
$H_0: \theta_{em} - \theta_{ef} = 0$	Father dominated	Mother dominated	Father dominated	Mother dominated
$H_1: \theta_{em} - \theta_{ef} > 0$	0.021	0.409	0.019	0.324
Decision	Rejecting H_0	Non-rejection H_0	Rejecting H_0	Non-rejection H_0

Note: Level of significance: 5%.

Chapter 5

Dropout Behaviour of Children: The Case of West Bengal

5.1 Introduction

There are a number of studies of empirical and theoretical nature that address the dropout problem of India from different aspects (Chugh (2011), Nithiya and Kundupuzhakkal (2017), Gouda and Sekher (2014), Prakash et al. (2017), Mishra and Chatterjee (2017)).[46] Chugh (2011) pointed out that family background (lack of financial resources and lack of interest in studies) are major reasons for drop out of children living in the slum areas in Delhi. She further noted that the dropout problem becomes acute at the below higher-secondary level. Nithiya and Kundupuzhakkal (2017) studied the current situation of education system in the light of gender disparity. They observed that the dropout problem is more acute among girl children because of marriage at early age, confined social structure and lack of suitable facilities for female students. Gouda and Sekher (2014) observed that dropout was high among children belonging to Muslim, Scheduled Caste and Scheduled Tribe families. They also noted that dropouts among children belonging to illiterate parents were four times higher than that of the literate parents and that, if parents were not working; the possibility of dropout among their children was relatively high. As in the case of 64th round (2007-2008) data on education expenditure (Chapter 3), according to the 71st round (January-June, 2014) data on education expenditure by NSSO, the dropout rate is 17% in the secondary level (below Class 10) and 2% in the higher secondary level. Further, Choudhury (2006) stated that the dropout rate becomes 2.7 times higher when a child makes a transition from the primary level to higher level in schools in India. Gouda and Sekher (2014) using NFHS-3 (2005-2006) data, observed that the primary level dropout rate is highest in West

[46] Christen L. Bradley and Renzulli (2011), Rumberger (2001), Le Thuc Duc and Tran Ngo Minh Tam (2013) have theoretically and empirically illustrated the dropout problem for US and Vietnam, respectively.

Bengal (15.5%) among all states in India. Hence, these studies indicate the fact that the dropout problem in acute when the child is studying in below class 10 level. On the supply side, in the context of lack of suitable facilities for female students, Prakash et al. (2017) provided a vivid picture based on villages in Karnataka where the vulnerability of adolescent girl children is another reason for discontinuing school. In case of West Bengal, Hati and Majumder (2012) found that 19.6% dropout rate in primary schools for boys and 15.4% for girls are higher than the overall dropout rate in India (6.4%) in 2009-2010. Further, due to financial constraints of the family, the dropped out children are often engaged in jobs in some unorganized sectors.

In this chapter, we address the dropout problem of school going children (below class 10 level) from two different perspectives. First, keeping in mind that the socio-economic and socio-demographic backgrounds of a household are influential factors for the dropout decision, we use a Probit regression to identify those household factors which prevent the child from going to school, using household level information available in our dataset. Second, in line with the study by Christen L. Bradley and Renzulli (2011), we classify the reasons, as stated by the students; into certain groups of reasons for dropout to look into the supply side. A Multinomial Logit regression is used to identify the specific factors that influence the dropped out child's particular choice. For this analysis, we use individual specific information available in our dataset. The latest (71st round, January 2014 - June 2014) education expenditure data, provided by the National Sample Survey Office (NSSO), for the state of West Bengal are used in this analysis. The analysis is done separately for boys and girls. In most of the empirical works, the *incidence* of dropout is analysed taking the dependent variable to be a dichotomous variable. It is, however, more important from policy perspective to track the paths for the chosen reasons for the child to drop out (Christen, Bradley and Renzulli (2011)).

The plan of the chapter is as follows. In Section 5.2, a brief description of empirical models used in this chapter is provided. The methodology is discussed in Section 5.3. Section 5.4 describes the data. In Section 5.5, the results of the Probit regression and Multinomial Logit regression are provided. Finally, Section 5.6 concludes the chapter with some remarks.

5.2 Model Specification

The dropout problem is complex and multidimensional. The related research works have found that apart from individual specific reasons (no interest in studies, low grade points etc.), the socio-economic background of the family and financial constraints are also reasons that influence a child to stop going to school. Keeping this in mind, the household specific reasons and individual reasons are analysed as follows:

Household level Analysis:

A dichotomous variable Y is defined as

$$Y = 1, \text{if the household has at least one dropped out child}$$

$$= 0 \text{ if, otherwise}$$

The Probit regression equation is the following

$$Prob(Y = 1) = \Phi(X'\beta), \tag{5.1}$$

where Φ is the cumulative density function of standard normal distribution. X is a vector of explanatory variables, which contain household specific variables that influence a child to drop out and β is the vector of coefficients for the explanatory variables. β is estimated using maximum likelihood method.

Individual level Analysis:

We now look at the dropout issue from an alternative angle. Here the students come up with reasons for not going to school. The possible reasons, as stated by the students, are summarized below.

According to Bradley and Renzulli (2011), the "push" factor group refers to the supply side factors that discourage a student from going to school. These factors are generally related to facilities provided by school authority: the environment of classes, medium of instruction, unable to cope up with studies, unfriendly atmosphere, quality of teachers etc. In case of India, as per the recent District Information System for Education (DISE) report (2014), 30% of primary and 15% of upper primary schools have PTRs (Pupil Teacher Ratio) higher than 30:1 and 35:1, respectively. Some private schools in urban areas even have just one teacher for 65-70 students. The percentages clearly indicate that such "push" factors would discourage many students from going to school and consequently there will be dropouts.

The "pull" factors refer to the reasons for dropout as a cost-benefit analysis (Stearns and Glennie 2006). In the context of India, almost every household has more than one child. Normally, in middle class and lower middle class families, the elder child in the household takes more responsibilities for the household. For example, a student may have to work with family members in farming, take care of younger siblings or may have to go out to work for earning money for the family. Consequently, he is unable to devote the required time and effort needed to succeed in studies. This leads to failure and discontinuity in studies. As per the latest Census report (2011), there are 33 million child labourers in the age group of 5-18 years in India.

Clearly, this implies that financial constraints, engagements in domestic and economic activities lead students to discontinue study.

"Opted out" factors refer to individual specific factors. Apart from push and pull factors a student may lose interest in studies or may think that a certain level of achievement in education is sufficient for him/ her.

"Other reasons" refer to the reasons that do not fall into the above three categories. For example, it might be a tradition in the family of a student not to study further.

For a boy child, the reasons for dropping out, as stated by the child, are categorized into the above-mentioned four groups-push factor, pull factor, opted out factor and other factors. For a girl child, another category is added to these four categories, which is "female-specific reasons". Prakash et al. (2017) noted that in addition to the four reasons, there are some hurdles that only a girl child has to face in India. It might so happen that due to reasons like early marriage, lack of female teachers, improper toilet facilities and distance between school and residence, the girl child has to leave school. These reasons become particularly relevant when the girl child attends adolescence[47].

A Multinomial Logit model, described below, is used for linking the choices with the actual variables of demand and supply side factors. The analysis is performed for boys and girls separately.

---

[47] Appendix V, Table A5.1 lists the group wise reasons obtained from our data.

Let the reason for dropping out for a student *l* be denoted by T_l. Here, a boy child has four categories of reasons ("push" factor, "pull" factor, "opted out" and "other reasons") and a girl child has five categories of reasons ("push" factor, "pull" factor, "opted out", "female-specific" and "other reasons"). The "other reasons" category is taken as base outcome for both boy and girl child. The probability of choosing the j^{th} reason category for l^{th} child (P_{lj}) is denoted by

$$P_{lj} = Prob(T_l = j^{th} reason) = \frac{e^{Z'_{lj}\gamma_j}}{\sum_j e^{Z'_{lj}\gamma_j}}, \; j=1 \ldots J. \tag{5.2}$$

Here, Z_{lj} represents a vector of individual specific information and some household characteristics that influence a child to choose a certain reason over the "other reasons" category and γ_j is the vector of coefficients corresponding to j^{th} reason. The qualitative dependent variable T_l can take any of the *J* possible values, each corresponding to a different reason category. Since each individual must select one reason category, only *J-1* sets of coefficients are uniquely defined. We will normalize by setting the coefficients for the J^{th} reason category to zero. The parameters of the model are estimated by the maximum likelihood method.

5.3 Methodology

There are three parts of the estimation procedure in this paper.

First, the Probit model (equation (5.1)) is estimated to analyze the important household specific features that cause child drop out. Here, the explanatory variables[48] are: (i) the log of monthly per capita expenditure (*Log(MPCE)*) as the household's financial profile, (iii) education level[49] of the father (*Father_edu*) and (iv) education level of the mother (*Mother_edu*) as the

[48] A brief description of explanatory variables is given in Appendix Table A5.2.

[49]Here, years of schooling is the indicator of education level.

household's educational background, (v) household size (*hhsize*), (vi) proportion of male children of below 16 years (*Child_m*) and (vii) proportion of female children of below 16 years (*Child_f*),(viii) age of the household head (*hdage*) as the household's demographic information, (ix) distances from primary, upper primary and secondary schools to the residence (*Distance1, Distance2* and *Distance3*),respectively, and (x) a dummy (d_1) which takes value 1 if any member of the household has computer[50] and takes value 0 if, otherwise.

Next, the Multinomial Logit model (equation (5.2)) for a male child is estimated. As already mentioned, here "other reasons" category is chosen as the base outcome. In this case, the dependent variables are the log of odds ratio of choosing "push factor" (denoted by *push*) vs. "other reasons" (denoted by *other)*, "pull factor" (denoted by *pull*) vs. "other reasons" and "opted out" (denoted by *opted*) vs. "other reasons". Thus, for "push factor", "pull factor" and "opted out", the dependent variables are $\log\left[\frac{Prob(T_l=push)}{Prob\ (T_l=other)}\right]$, $\log\left[\frac{Prob(T_l=pull)}{Prob\ (T_l=other)}\right]$ and $\log\left[\frac{Prob(T_l=opted)}{Prob\ (T_l=other)}\right]$, respectively, for l^{th} male child.

Here the explanatory variables are (i) the log of monthly per capita expenditure (*Log(MPCE))*, (iii) education level of the father (*Father_edu*) and (iv) education level of the mother (*Mother_edu*), (v) proportion of male children of below 16 years *(Child_m*) and (vi) proportion of female children of below 16 years (*Child_f*), (vii) a dummy (d_2), which takes value 1 if the child has access to computer, 0 otherwise, (viii) a dummy (*d_institute*), which takes value 1 if the school is a Government school, 0 otherwise, (ix) a dummy (*d_language*) which takes value 1 if the language used in the school and that at home are same, 0 otherwise, (x) a dummy (*d_midday*)

[50]Computer includes desktop, laptop, palmtop, notebook, netbook, smartphone, tablets etc.

which takes value 1 if the school provides midday meal, 0 otherwise, (xi) age at the time of drop out (*dropout_age*) and (xii) grade when dropped out (*dropout_grade*).

Finally, the Multinomial Logit model (equation (5.2)) is estimated for a female child. Here, there is an additional option for the dependent variable, viz., "female-specific factors" (denoted by *fem*) vs. "other reasons". Thus, for "push factor", "pull factor", "opted out" and "female-specific"the dependent variables are log $\left[\frac{Prob(T_l=push)}{Prob\ (T_l=other)}\right]$, log $\left[\frac{Prob(T_l=pull)}{Prob\ (T_l=other)}\right]$,log $\left[\frac{Prob(T_l=opted)}{Prob\ (T_l=other)}\right]$ and log $\left[\frac{Prob(T_l=fem)}{Prob\ (T_l=other)}\right]$, respectively, for l^{th} female child. The explanatory variables remain the same as in the case of a male child.

5.4 Description of the Data

The data set used here is the 71st round data on participation and expenditure on education conducted by the National Sample Survey Office (NSSO), Govt. of India, for the state of West Bengal. The span of the data set is January 2014 - June 2014. The survey has been conducted over 29 states and 6 UTs. This is the latest data set available that contains detailed information on education expenditure. Moreover, from this data set a complete profile of household education expenditure pattern at the individual level can be obtained. For West Bengal, the sample size is 2536 households in the rural sector and 2304 households in the urban sector. Of these, 1268 households in the rural sector and 936 households in the urban sector have at least one dropped out child. These constitute our data set.

Table 5.1 presents the distribution of male and female students (below class 10 level) by educational status. While the dropout rate for male children is 17.45% in the rural sector and

13.51% in the urban sector, the corresponding figures for the female children are 17.84% in the rural sector and 15.29% in the urban sector.

Table 5.2 presents the distribution of dropped out children by reason for dropping out, average dropout age and average grade of dropout separately for male and female students. Clearly, the "pull factor" plays the major role from children's' perspective, followed by "opted out" for males and "female specific" reasons for females. Hence, as per the child's perception, financial constraints, engagement in economic and domestic activities seem to be the major reasons for dropping out for both male and female children in both sectors. It is also evident that on an average the children drop out at the 7th grade across all categories.[51]

5.5 Results and Discussion

Table 5.3 gives the results of the Probit regression for households with at least one dropped out child. The values of R^2 are 0.1238 and 0.1564 for the rural and the urban sector, respectively. The coefficients of *Distance1*, *Distance2* and *Distance3* are non-significant in both sectors, which implies that the distance from the residence to the schools are not important factors for drop out of a child. The coefficients of *LnMPCE* are negative, significant which implies that children belonging to the households with better financial status are less likely to drop out. The coefficients of *hhsize* are positive, significant in both sectors implying that children from larger households (in terms of household size) are more likely to drop out compared to smaller households. This is further supported by the positive, significant coefficients of *Child_m* and *Child_f* in both sectors. In other words, belonging to a large family and having many younger siblings negatively affects school participation, which corroborates the findings of earlier studies.

[51] It may be pointed out that the results are specific to this particular grouping. The sensitivity of results with respect to alternative grouping of reasons could not be checked owing to lack of data.

This is possibly because children belonging to the households with more children (male and female) get less care in terms of education. The coefficients of *Father_edu*, *Mother_edu* and *hdage* are negative, significant in both sectors. This implies that children of the households with educated parents and households with older head (with experience and wisdom) are less likely to drop out. The coefficients of the dummy d_1 are negative, significant in both sectors, indicating that children belonging to the households having computer are less likely to discontinue studies. Children belonging to households owning a computer generally belong to affluent society and they become used to such facility. Hence, the probability for such a child to drop out, especially for 'pull' factor is likely to be low.

In Table 5.4, the results of the Multinomial logit regression for reasons to drop out for the male children in both sectors are provided. The salient points that emerge from this table are as follows:

(i) For the "push factor", all coefficients except "*Child_*f" in the rural sector and "*Child_f*" and "*Child_m*" in the urban sector are significant with meaningful signs. For a male child in the rural sector, belonging to an affluent household with larger number of male siblings and educated parents have negative impacts on the probability of choosing this option. In the urban sector, while the other implications remain the same, the number of siblings does not play any role. On the supply side, access to computers, going to non-public schools that provide midday meals, same language as the medium of instruction and that spoken at home, have negative impacts on the probability of choosing "push factor" as a reason for dropout compared to "other factors". The coefficients of "*dropout_age*" and "*dropout_grade*" are positive. That is, male children who drop out at a higher age or higher grade, have higher probability of choosing "push factor" as a reason for dropout.

(ii) For the "pull factor", all coefficients relating to household characteristics are significant in both sectors. On the supply side, except for d_2 (computer access) in the rural sector and "*d_language*" in both sectors other coefficients are significant. The coefficients of "*dropout_age*" and "*dropout_grade*" are positive in both sectors. Thus, a male child, belonging to an affluent household with smaller number of siblings and educated parents is less likely to choose this option in the rural sector. In the urban sector, a male child, belonging to an affluent household with smaller number of male siblings, higher number of female siblings and educated parents is less likely to choose this option. This is possibly because the female siblings take care of the household activities. On the supply side, language does not play any role in either sector. Access to computers has no role in the rural sector, while in the urban sector male children having computer access are less likely to choose this option. Finally, probability of choosing this option is higher for children going to public schools with no midday meal in both sectors. Thus, for children choosing the "pull factor" as a reason for dropout, the school infrastructure is not adequate to motivate them to study. They would rather prefer to be engaged in domestic activities.

(iii) For the "opted out" option, in the rural sector, the significant variables are household income, parents' education (with plausible negative coefficients) on the household side and most of the supply side variables (except "*language*"). In other words, for a male child, going to public schools that provide midday meals, have negative impacts on the probability of choosing "opted out factor" as a reason for dropout compared to "other factors". In the urban sector, in addition to the above household factors, having larger number of male siblings has a negative impact. On the supply side, computer access and provision of midday meal have negative impacts on the probability of choosing this option, while going to public schools has a positive impact in the rural sector. In the urban sector, in addition to the above factors, different language for medium of instruction at school and that spoken at home has a positive impact.

In Table 5.5 the results of the multinomial logit regression for reasons to drop out for female children in both sectors are provided. The values of the Pseudo R^2 are 0.1925 and 0.1894 in the rural and the urban sector, respectively. From this table the following observations are in order.

(iv) For the "push factor", all coefficients are significant with meaningful signs in both sectors. To be specific, for a female child, belonging to an affluent household with large number of siblings and educated parents have negative impacts on the probability of choosing this option. On the supply side, access to computers, going to public schools that provide midday meals, same language as the medium of instruction and that spoken at home, have negative impacts on the probability of choosing "push factor" as a reason for dropout compared to "other factors". The coefficient of "*dropout_age*" is positive, while that for "*dropout_grade*" is negative. That is, female children who drop out at a higher age, have higher probability of choosing "push factor" as a reason for dropout, but female children who drop out at a higher grade have lower probability of choosing "push factor" as a reason for dropout.

(v) For the “pull factor”, all coefficients are significant in both sectors, except for d_2 (computer access) and “*dropout_age*” in the urban sector. Except for “*Child_m*”, “*Child_f*” and “*dropout_grade*’, all other variables have similar impact on the probability of choosing this option as in case of the “push factor”. Here larger number of siblings has a positive impact. It is also interesting to note that female children who drop out at a higher grade have higher probability of choosing “pull factor” (as opposed to the case in “push factor”) as a reason for dropout. These point to the fact that female children are forced to drop out not because they are incapable or are unable to cope with studies, but because they have to be engaged in domestic/economic activities and take care of the siblings.

(vi) For the “opted out” option, in the rural sector, the significant variables are parents’ education (with plausible negative coefficients) on the household side and most of the supply side variables. In other words, going to public schools that provide midday meals, same language as the medium of instruction and that spoken at home, have negative impacts on the probability of choosing “opted out factor” as a reason for dropout compared to “other factors”. In the urban sector, for a female child, belonging to an affluent household with large number of male siblings and educated father have negative impacts on the probability of choosing this option. However, lager number of female siblings and mother’s education has a positive impact. One explanation could be that with a larger number of female children the mother decides that the girl child has had enough education. The positive impacts of “*dropout_age*” and “*dropout_grade*” in both sectors seem plausible for the choice for this option.

(vii) For “female specific factors”, on the household side in the rural sector, for a female child, belonging to an affluent household with smaller number of siblings and educated parents have negative impacts on the probability of choosing this option. On the supply side, only the fact that the child goes to a public school, has a positive impact on the probability of choosing this option. This points to the fact that the schools possibly suffer from improper toilet facilities or lack of female teachers. In the urban sector, belonging to an affluent household with small number of female siblings and educated parents have negative impacts on the probability of choosing this option. On the supply side, computer access has a negative impact and public school (providing midday meals, but not having adequate female specific infrastructure) have positive impact on probability of choosing this option. The positive impacts of “*dropout_age*” and “*dropout_grade*” in both sectors seem plausible for the choice for this option. Thus, ‘inadequate female specific infrastructure’ seems to turn out as a major factor that contributes to female children dropout and this applies generally to children from poorer families, as children from richer families do not go to such schools.

5.6 Concluding Remarks

In this chapter, the drop out behaviour of children (male and female) is studied both from the point of view of households and from that of children. The 71st round (2014) data on education expenditure, provided by the NSSO, for the state of West Bengal is used in this study.

In terms of incidence of dropout, children from affluent families and educated parents are less likely to drop out in both the rural and the urban sector. However, belonging to a large family and having many younger siblings negatively affect school participation in the rural sector, while for urban children these variables do not make any difference.

From the children's perspective, for boys who drop out, reporting about "push factor", "pull factor" or "opted out" are less probable compared to "others" if they belong to affluent families and have educated parents. They are also less likely to report these factors if the school is a non-public school providing midday meals.

The same implications hold for girls with similar family background, in terms of the above choices. The additional choice, viz., "female specific reasons" is also less likely, compared to "other" factors, as these families are generally liberal (in terms of marrying off a girl at an early age) and these girls generally go to schools with adequate infrastructure. This is corroborated by the fact that "going to a public school" has a positive impact on the probability of choosing this option, and children from poorer section of the society go to public schools, in general. It is interesting to note that mid-day meal plays no significant role in reporting "female specific reasons". Thus, public school infrastructure emerges as a major factor contributing to female dropout because of 'female specific reasons'. Given that infrastructure includes 'adequate toilet facilities', which are lacking in most public schools[52], this study contributes to the justification for the demand of proper toilet facilities for female children at school.

[52]According to a report (2014) prepared by the Department of School Education and Literacy under the Union Human Resource Development, West Bengal features in the second spot in government schools lacking girls' toilets.

Table 5.1: Distribution of Students (below class 10 level) by Educational Status

Sex of the child	Educational Status	Sector	
		Rural	**Urban**
Male	Dropped out	996 (17.45%)	689 (13.51%)
	Studying	4,787 (82.55%)	4,506 (86.49%)
	Total	**5783 (100%)**	**5195 (100%)**
Female	Dropped out	972 (17.84%)	743 (15.29%)
	Studying	4,564 (82.16%)	4,223 (84.71%)
	Total	**5536 (100%)**	**4966 (100%)**

Table 5.2: Descriptive Statistics of Dropped out Students

Reasons/ factors For Dropping out	Male (1685)		Female (1715)	
	Rural (996)	**Urban (689)**	**Rural (972)**	**Urban (743)**
Push	45 (4.52%)	22 (3.19%)	40 (4.12%)	19 (2.56%)
Pull	605 (60.74%)	374 (54.28%)	520 (53.50%)	337 (45.36%)
Opted out	281 (28.21%)	246 (35.70%)	176 (18.11%)	186 (25.03%)
Female Specific			180 (18.52%)	152 (20.46%)
Others	65 (6.53%)	47 (6.82%)	56 (5.76%)	49 (6.59%)
Average dropout age (years)	12	14	14	16
Average dropout grade	7	7	7	7

Table 5.3: Probit Regression Results for Households with at least One Dropped out Child

Explanatory variables	**Rural**	**Urban**
LnMPCE	-0.239** (0.031)	-0.580*** (0.000)
Hhsize	0.371*** (0.000)	0.302*** (0.000)
Child_m	0.990*** (0.001)	0.625*** (0.000)
Child_f	0.034*** (0.002)	0.906*** (0.001)
Father_edu	-0.027*** (0.003)	-0.048** (0.028)
Mother_edu	-0.037*** (0.001)	-0.042*** (0.001)
Hdage	-0.028** (0.021)	-0.032*** (0.002)
DISTANCE1(distance of residence from primary school)	-0.048 (0.119)	-0.147 (0.201)
DISTANCE2(distance of residence from upper primary school)	0.061 (0.102)	0.053 (0.134)
Distance3(distance of residence from secondary school)	-0.033 (0.502)	-0.059 (0.147)
d_1(Dummy =1 if household has computer access, 0 otherwise)	-0.250*** (0.002)	-0.244*** (0.002)
Constant	0.668*** (0.000)	0.847*** (0.000)
Number of observations	2536	2304
Pseudo R^2	0.1238	0.1564

Figures in the parentheses are the *p*-values.

***: Significant at 1% level; **: Significant at 5% level

Table 5.4: Reasons for Male Children Dropout (Multinomial Logit Model)

Explanatory Variables	Reasons (Factors) for dropping out (base outcome: Other)					
	Rural			Urban		
	Push	Pull	Opted out	Push	Pull	Opted out
LnMPCE	-1.961*** (0.006)	-0.958** (0.012)	-1.531*** (0.009)	-0.575** (0.025)	-0.014** (0.012)	-0.041*** (0.005)
Child_m	-0.092*** (0.005)	0.182** (0.022)	-0.140 (0.301)	0.211 (0.145)	0.380*** (0.001)	-0.592*** (0.001)
Child_f	0.101 (0.127)	0.093*** (0.002)	0.183 (0.711)	-0.594 (0.102)	-0.495*** (0.008)	0.478 (0.109)
Father_edu	-0.084*** (0.001)	-0.057** (0.031)	-0.012** (0.031)	-0.001*** (0.007)	-0.018*** (0.002)	-0.133*** (0.001)
Mother_edu	-0.035** (0.011)	-0.079*** (0.001)	-0.078*** (0.001)	-0.058*** (0.002)	-0.102*** (0.001)	-0.028*** (0.001)
d_2 (=1 if child has computer access,0 otherwise)	-0.217*** (0.005)	-0.109 (0.101)	-0.331*** (0.001)	-1.201** (0.021)	-0.336** (0.015)	-0.875** (0.021)
d_Inst (=1 if public school, 0otherwise)	0.279*** (0.001)	1.753* (0.081)	2.805*** (0.001)	1.096*** (0.002)	0.864*** (0.002)	0.687*** (0.001)
d_language (=1 if same language in school and home, 0 otherwise)	-0.732*** (0.002)	0.203 (0.501)	0.084 (0.778)	-1.647*** (0.001)	-0.090 (0.154)	-1.092*** (0.009)
d_midday (=1 if midday meal provided, 0 otherwise)	-0.485** (0.011)	-0.885** (0.011)	-1.091* (0.098)	-1.819*** (0.002)	-0.136** (0.011)	-0.732*** (0.001)
dropout_age	0.147*** (0.001)	0.075*** (0.001)	0.022*** (0.005)	0.292** (0.021)	0.051*** (0.002)	-0.167*** (0.001)
dropout_grade	0.140*** (0.001)	0.119*** (0.001)	0.064*** (0.001)	0.432*** (0.003)	0.024*** (0.002)	-0.041*** (0.001)
Constant	1.664*** (0.001)	-0.615*** (0.001)	1.044*** (0.001)	-0.125*** (0.031)	0.029*** (0.001)	0.274*** (0.001)
Number of observations	996			689		
Pseudo R^2	0.1764			0.1545		

Figures in the parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

Table 5.5: Reasons for Female Children Dropout (Multinomial Logit Model)

Explanatory Variables	Reasons (Factors) for dropping out (base outcome: Other)							
	Rural				Urban			
	Push	Pull	Opted out	Female specific	Push	Pull	Opted out	Female specific
LnMPCE	-0.488*** (0.001)	-0.056*** (0.009)	0.007 (0.129)	-0.885*** (0.000)	-0.595** (0.011)	-0.156*** (0.001)	-0.307*** (0.005)	-0.665** (0.012)
Child_m	-0.454*** (0.003)	0.349*** (0.001)	-0.532 (0.801)	0.404*** (0.008)	-1.048*** (0.001)	0.462** (0.034)	-0.414*** (0.001)	-0.417 (0.608)
Child_f	-0.509*** (0.001)	0.503*** (0.006)	0.503 (0.651)	0.532*** (0.001)	-0.667*** (0.000)	0.474** (0.011)	0.451** (0.018)	0.370*** (0.001)
Father_edu	-0.005** (0.029)	-0.008** (0.011)	-0.027*** (0.001)	-0.011** (0.031)	-0.151*** (0.000)	-0.054** (0.012)	-0.036** (0.013)	-0.001*** (0.001)
Mother_edu	-0.097*** (0.009)	-0.112**(0.021)	-0.106*** (0.001)	-0.177*** (0.006)	-0.146*** (0.005)	-0.096*** (0.004)	0.042*** (0.001)	-0.069*** (0.005)
d_2 (=1 if child has computer access,0 otherwise)	-0.076*** (0.002)	-0.098*** (0.007)	-0.652 (0.809)	1.009 (0.109)	-0.807*** (0.001)	0.089 (0.801)	0.093 (0.301)	-1.237*** (0.002)
d_Inst (=1 if public school, 0 otherwise)	0.008*** (0.009)	0.235*** (0.001)	0.045* (0.076)	1.205*** (0.001)	0.103*** (0.002)	0.054*** (0.001)	0.706*** (0.001)	0.034*** (0.001)
d_language (=1 if same language in school and home, 0 otherwise)	-0.514*** (0.008)	-0.175*** (0.001)	-0.394*** (0.001)	-0.487 (0.269)	-0.550*** (0.003)	-0.362*** (0.002)	-0.858*** (0.001)	-0.819 (0.507)
d_midday (=1 if midday meal provided, 0 otherwise)	-0.074*** (0.001)	-0.705** (0.011)	-0.183*** (0.001)	0.694 (0.451)	-0.114** (0.010)	-0.317** (0.015)	0.704 (0.191)	0.564* (0.091)
dropout_age	0.231*** (0.005)	0.054*** (0.002)	0.083*** (0.008)	0.272*** (0.002)	0.292*** (0.004)	-0.001 (0.998)	0.226** (0.026)	0.134*** (0.001)
dropout_grade	-0.137* (0.091)	0.023*** (0.001)	0.017** (0.031)	0.095*** (0.001)	0.113*** (0.001)	0.078** (0.021)	0.259*** (0.001)	0.341*** (0.001)
Constant	-0.905*** (0.001)	1.004*** (0.001)	1.067*** (0.001)	0.092*** (0.001)	1.055*** (0.001)	1.009** (0.023)	0.089** (0.043)	1.053*** (0.001)
No. of observations	972				743			
Pseudo R^2	0.1925				0.1894			

*Figures in the parentheses are the p-values.

***: Significant at 1% level; **: Significant at 5% level; *: Significant at 10% level.

APPENDIX V

Table A5.1: Grouping of the Reasons Provided By 71st Round Schedule of NSSO

Categories	Reasons in 71st round schedule of NSSO
Push factor	1. Unable to cope with studies 2. Unfriendly atmosphere
Pull factor	1. Financial Constraint 2. Engaged in domestic activities. 3. Engaged in economic activities
Opted out	1. Not interested to study 2. Achieved desired level.
Other	1. Other reasons 2. Not in the tradition in family
Female specific	1. Marriage 2. Female teacher not available 3. Toilet unavailable

Table A5.2: Explanation of the Explanatory Variables

Explanatory Variables	Explanation of the variables
LnMPCE	Log of the monthly per capita expenditure
Lsq	Squared value of Log(MPCE)
Hhsize	Household size
Child_m	Proportion of household members of age less than 16 years (male)
Child_f	Proportion of household members of age less than 16 years (female)
Father_edu	Education level of Father (years of schooling)
Mother_edu	Education level of Mother (years of schooling)
Hdage	Age of the household head (years)
DISTANCE1	Distance from primary school to the residence (k.m.)
DISTANCE2	Distance from upper primary school to the residence (k.m.)
DISTANCE 3	Distance from secondary school to the residence (k.m.)
d_1	d_1=1, if the household has access to computer =0, if otherwise
d_2	d_2=1, if the child has access to computer =0, if otherwise
d_Inst	*d_Inst* =1 if the school is government school =0, if otherwise
d_language	*d_language* =1 if the languages used in the school and in the house are same, = 0, if otherwise
d_midday	*d_midday* =1 if the school provides midday meal = 0, if otherwise
dropout_age	Age (years) when the child dropped out
dropout_grade	Grade when the child dropped out

Chapter 6

Concluding Remarks and Future Works

This thesis is an endeavour to explore the issue of gender bias in the field of education for the state of West Bengal, India in the following directions.

(i) Gender bias among students at school and college levels, both in terms of participation in the education system, and expenditure allocation;

(ii) The role of mother's bargaining power in allocating educational expenditure to boys and girls; and

(iii) Dropout behaviour of children at the school level.

This chapter summarizes the findings and provides possible directions for future work.

Chapter 2 deals with identification of the educational items, which are used exclusively by 'students' (both school and college level) and examines the presence of gender discrimination in allocation of expenditure to male and female students on these items. From the analysis, it is found that all education items listed under NSSO data relate exclusively to students. Educational items like 'books', 'stationery and photocopy charges', 'tuition fees' and 'private coaching fees' are found to be students' items. 'Transport cost', 'Pan', 'Tobacco' and 'Alcohol' turn out to be adult goods. Evidence of pro male bias is found in expenditure allocation on educational items, the extent of bias being more in the urban sector compared to the rural sector.

Chapter 3 examines the issue of gender bias separately in below class-10 and above class 10 levels of the education system of West Bengal. This study establishes that there exists pro male bias in the decision to send the child to school and in the spending on education in the below class-10 level in both the rural and the urban sectors, particularly for children belonging to the age groups 10-14 years and 15-16 years. For the above class-10 level, the analysis indicates evidence of discrimination against girl students at the level of selection of subject streams such as ‘Science’ and ‘Commerce’ and also in the allocation of expenditure within a subject stream in the urban areas in the state of West Bengal.

Chapter 4 is concerned with gender bias in the context of intra household allocation. The findings that emerge are: pro male bias in allocation of expenditure on education is observed in households where husbands enjoy higher bargaining power. However, when the wife is the decision maker, no such bias is observed.

Chapter 5 analyzes the dropout behaviour of school children from two different perspectives, namely, the parents’ (household’s) perspective and child’s perspective. The findings from this study are: in terms of incidence of dropout, children from affluent families and educated parents are less likely to drop out in both the rural and the urban sector. However, belonging to a large family and having many younger siblings negatively affect school participation in the rural sector, while for urban children these variables do not make any difference.

From the childrens’ perspective, boys belonging to affluent families with educated parents are less likely to report “push factor”, “pull factor” or “opted out” options (compared to other reasons) as reasons for drop out. They are also less likely to report these factors if the school is a non-public school providing midday meals.

The same implications hold for girls with similar family background, in terms of the above choices. The additional choice, viz., “female specific reasons” is also less likely, compared to “other” factors, as these families are generally liberal (in terms of marrying off a girl at an early age) and send girls to schools with adequate infrastructure. Public school infrastructure emerges as a major factor contributing to female dropout because of ‘female specific reasons’. Given that infrastructure includes ‘adequate toilet facilities’, which are lacking in most public schools, this study justifies the demand of proper toilet facilities for female children at school.

Possible Extensions and Future Work

In this thesis, the focus is on the presence of gender bias in the field of education in West Bengal, India and the research work is oriented towards exploring the issue of gender discrimination in education from different angles in the state. However, these works can be extended in other directions also. Apart from extending these works to other states of India and thereby comparing the findings, there are other socio-economic policy issues, which can be addressed by incorporating and extending the findings of this thesis. In what follows, the highlights of some possible extensions are provided.

First, the presence of gender bias in below and above class 10 level has already been established for the overall population in West Bengal. However, the social structure of India is layered in multiple levels. Hence, the nature of gender bias in education expenditure is unlikely to be uniform across the various social (SC, ST and General) and religious groups (Hindu, Muslim and Others) in the rural and the urban sector. In order to explore this and to identify the major groups contributing to the discrimination, the overall male-female difference in mean expenditure on education can be decomposed by religious/social groups using Oaxaca-Blinder decomposition technique (modified to incorporate more than two groups).

Second, the resource share of an individual in a household is one of the important welfare indicators that helps to determine the bargaining power of the member, the access to resources in the household, intra household poverty, etc. Chapter 4 estimates the sharing rule only for the husband and the wife in a household under the assumption that children do not earn and they have to depend on adults to fulfil their needs. This structure can be extended by recognizing that a child's resource share can be considered as an indicator of the child's well-being (nutrition, health outcomes, etc and as a major tool for the policy designs targeted to the welfare of children. Hence, the child's resource share can be estimated separately through a utility function specific to the child as a function of the exclusive goods consumed by the child in the household. Further, the estimated resource share can be compared between a male child and a female child to detect gender discrimination in the household.

Another study of interest could be a comparison of the dropout scenario between the two periods of NSS rounds (i.e., the 64^{th} and 71^{st} rounds) using the education expenditure data.

We conclude by noting that there are some limitations of the studies undertaken in the thesis. The scope of extension and modification of the work are enormous. We plan to take these up in our future work.

Reference

1. Arias, C., Atella, V., Castagnini, R., and Perali, F. (2004). Estimation of the Sharing Rule Between Adults And Children and Related Equivalence Scales Within A Collective Consumption Framework. *Household Behaviour, Equivalence Scales, Welfare and Poverty* (pp. 129-161). Physica-Verlag.

2. Aslam, M., and Kingdon, G. G. (2008): Gender and Household Education Expenditure in Pakistan, *Applied Economics,* 40(20*)*, 2573-2591.

3. Attanasio, O. P., and Lechene, V. (2014): Efficient Responses to Targeted Cash Transfers, *Journal of Political Economy*, 122(1), 178-222.

4. Attanasio, O., Battistin, E., and Mesnard, A. (2012): Food and Cash Transfers: Evidence from Colombia. *The Economic Journal*, 122(559), 92-124.

5. Azam, M., and Kingdon, G. G. (2013): Are Girls the Fairer Sex in India? Revisiting Intra-Household Allocation of Education Expenditure. *World Development*, 42,September, 143-164.

6. Balatchandirane, G. (2003). Gender Discrimination in Education and Economic Development: A study of South Korea, China and India. *International Studies*, 40(4), 349-378.

7. Bandyopadhyay, M., and Subrahmanian, R. (2008). Gender Equity in Education: A Review of Trends and Factors. (see http://sro.sussex.ac.uk/1866/1/PTA_18.pdf)

8. Banks, J. Blundell, R. and Lewbel, A. (1997): Quadratic Engel Curves and Consumer Demand, *Review of Economics and Statistics*, 79(4), 527-539.

9. Beckert, W., and Blundell, R. (2004). Invertibility of Nonparametric Stochastic Demand Functions. *Working paper, Birkbeck Working Papers in Economics and Finance,BWPEF, 406*.
(see:http://www.bbk.ac.uk/ems/research/wp/PDF/BWPEF0406.pdf)

10. Bargain, O., and Donni, O. (2012): Targeting and Child Poverty, *Social Choice and Welfare*, 39(4), 783-808.

11. Bargain, O., & Donni, O. (2009). The measurement of child costs: a Rothbarth-type method consistent with scale economies. (see https://researchrepository.ucd.ie/bitstream/10197/2632/1/wp10.01.pdf)

12. Barten, A. P. (1964): Consumer Demand Functions under Conditions of Almost Additive Preferences. *Econometrica: Journal of the Econometric Society*, 32(1), 1-38.

13. Becker, G. S. (1973): A Theory of Marriage: Part I, *Journal of Political Economy*, 81(4), 813-846.

14. Becker, G. S. (1981): Altruism in the Family and Selfishness in the Market Place. *Economica*, 48(189), 1-15.

15. Becker, G. S., and Lewis, H. G. (1973): On the Interaction Between the Quantity and Quality of Children, *Journal of Political Economy*, 81(2, Part 2), S279-S288.

16. Blinder, A.S., (1973): Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources*, 8(4), *436 – 455*.

17. Bobonis, G. J. (2009): Is the Allocation of Resources Within the Household Efficient? New Evidence from A Randomized Experiment, *Journal of Political Economy*, 117(3), 453-503.

18. Bourguignon, F., Browning, M. and Chiappori, P. A. (2009): Efficient Intra-Household Allocations and Distribution Factors: Implications and Identification, *The Review of Economic Studies*, 76(2), 503-528.

19. Bradley, C. L. and Renzulli, L. A. (2011): The Complexity of Non-Completion: Being Pushed or Pulled to Drop Out of High School. *Social Forces*, 90(2), 521-545.

20. Brown, R.S., Moon, M. and Zoloth, S., (1980): Incorporating Occupational Attainment in Studies of Male– Female Differentials, *Journal of Human Resources* 15(1), 3 – 28.

21. Browning, M., and Chiappori, P. A. (1998). Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica: Journal of the Econometric Society*, November 1, 1241-1278.

22. Browning, M., Chiappori, P. A., and Lewbel, A. (2013): Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power, *Review of Economic Studies*, 80(4), 1267-1303.

23. Burgess, R., and Juzhong, Z. (2002): Modernization and Son Preference in People’s Republic of China. Economics and Research Department (ERD), WORKING PAPER SERIES NO. 20

24. Calvi, R., Lewbel, A. and Tommasi, D. (2017): LATE with Mismeasured or Misspecified Treatment: An Application to Women's Empowerment in India. ECARES (Free University of Brussels) working paper 2017-27.

25. Cameron, A. C. and Trivedi, P.K (2005): Microeconometrics: Methods and Applications, *Cambridge University Press.*

26. Chavan V.M. and Chavan G.M (2018): School Health Performance Score: A Comparative Study Between Rural And Urban School Performance, *International Journal of Community Medicine and Public Health*, 5(6), 2421-2424.

27. Cherchye, L., De Rock, B., and Vermeulen, F. (2011). The Revealed Preference Approach to Collective Consumption Behaviour: Testing and Sharing Rule Recovery. *The Review of Economic Studies*, 78(1), 176-198.

28. Cherchye, L., De Rock, B., Lewbel, A. and Vermeulen, F. (2015): Sharing Rule Identification for General Collective Consumption Models, *Econometrica: Journal of the Econometric Society*, 83(5), 2001-2041.

29. Chiappori, P. A. (1988): Nash-bargained households Decisions: A Comment, *International Economic Review*, 29(4), 791-796.

30. Chiappori, P. A. (1992): Collective Labor Supply and Welfare. *Journal of Political Economy*, 100(3), 437-467.

31. Chiappori, P. A., Fortin, B. and Lacroix, G. (2002): Marriage Market, Divorce Legislation and Household Labor Supply. *Journal of Political Economy*, 110(1), 37-72.

32. Choden, P. and Sarkar, D. (2013): Gender Bias in Schooling: The Case for Bhutan, *Journal of the Asia Pacific Economy*, 18(4), 513-528.

33. Choudhury, Amit (2006), “Revisiting Dropouts: Old Issues, Fresh Perspectives”, Economic and Political Weekly,December 16

34. Chugh, S. (2011): Dropout in Secondary Education: A Study of Children Living in Slums of Delhi. NUEPA Occasional paper 37, 1-53. National University of Educational Planning and Administration, New Delhi.

35. Couprie, H. (2007): Time Allocation Within the Family: Welfare Implications of Life in A Couple, *The Economic Journal*, 117(516), 287-305.

36. Couprie, H., Peluso, E. and Trannoy, A. (2010): Is Power More Evenly Balanced in Poor Households?, *Journal of Public Economics*, 94(7), 493-507.

37. Cragg, J. G. (1971): Some Statistical Models for Limited Dependent Variables With Application to The Demand for Durable Goods, *Econometrica: Journal of the Econometric Society*, September 1, 829-844.

38. Davidson, R. and MacKinnon, J. G. (1993): Estimation and Inference in Econometrics, *OUP Catalogue*.

39. Deaton, A. (1989): Looking for Boy-Girl Discrimination in Household Expenditures, *World Bank Economic Review*, 3(1), 1-15.

40. Deaton, A. and Subramanian, S. (1991): Gender bias in Indian Consumption Pattern, *Sarvekshana* 14(4), 1-12.

41. Deaton, A. (1997). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy, *World Bank Publications.*

42. Dolton, P.J. and Kidd, M.P. (1994): Occupational Access and Wage Discrimination". *Oxford Bulletin of Economics and Statistics,* 56(4), 457-474.

43. Dreze, J. and Kingdon, G. G. (2001): School Participation in Rural India, *Review of Development Economics*, 5(1), 1-24.

44. Donni, O. (2009). A simple approach to investigate intrahousehold allocation of private and public goods. *The Review of Economics and Statistics*, *91*(3), 617-628.

45. Duc, L. T. and Tam, T. N. M. (2013): Why Children in Vietnam Drop out of School and What They Do After That (Vol. 102). Working Paper: Oxford: Young Lives.(see: http://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-WP102_Le%20Thuc%20Duc_why%20do%20children%20leave%20school%20early.pdf).

46. Dunbar, G. R., Lewbel, A. and Pendakur, K. (2014): Identification of Random Resource Shares in Collective Households With An Application to Microcredit in Malawi, Mimeo, Boston University.

47. Dunbar, G. R., Lewbel, A. and Pendakur, K. (2017): Identification of Random Resource Shares in Collective Households Without Preference Similarity Restrictions, *Bank of Canada.*

48. Elborgh-Woytek, Newiak, M., Kochhar, K., Fabrizio, S., Kpodar, K., Wingender, P., Clements, B., and Schwartz, G. (2013). Women, Work, and the Economy: Macroeconomic Gains From Gender Equity, INTERNATIONAL MONETARY FUND, Strategy, Policy, and Review Department and Fiscal Affairs Department.

49. Fuwa, N. (2006): The Net Impact of The Female Secondary Stipend Program in Bangladesh. (see: http://ssrn.com/abstract=879245 [Google Scholar])

50. Gabriel, P. E. and Schmitz, S. (1989): The Effect of Occupational Segregation on the Relative Earnings of Black Males, *Economics Letters,* 30(1), 77-83.

51. Ghosh, B. and Kar, A. M. (2010). Child Marriage in Rural West Bengal: Status and Challenges. *Indian Journal of Development Research and Social Action*, 6(12), 1-23.

52. Gorman, W. M. (1976). 11 TRICK WITH UTILITY FUNCTIONS. *Essays in Economic Analysis: The Proceedings of the Association of University Teachers of Economics Sheffield 1975*, 211.

53. Gouda M, S. and Sekher, T.V. (2014), Factors Leading to School Dropouts in India: An Analysis of National Family Health Survey-3 Data, *IOSR Journal of Research and Method in Education (IOSR-JRME) e-ISSN: 2320–7388, p-ISSN: 2320–737X Volume 4, Issue 6 Ver. III* (Nov - Dec. 2014), 75-83.

54. Gautam, M. (2015). Gender, Subject Choice and Higher Education in India: Exploring 'Choices' and 'Constraints' of Women Students. *Contemporary Education Dialogue*, *12*(1), 31-58.

55. Government of India (2013): Survey for Assessment of Dropout Rates at Elementary Level in 21 States, Final Report, Edcil (India) Limited. A Government of India Enterprise.

56. Government of India (2014): School Education in India, National University of Educational Planning and Administration and Department of School Education and Literacy Ministry of Human Resource Development Government of India.

57. Government of India (2014): Annual Report, Department of School Education and Literacy under the Union Human Resource Development.

58. Government of India (2016). Educational Statistics at a Glance, Ministry of Human Resource Development Department of School Education and Literacy.

59. Hati, K. K. and Majumder, R. (2012): Proximate Determinants of School Dropout: A study on Rural West Bengal, MPRA Paper No. 45850 (see: http://mpra.ub.uni-muenchen.de/45850/).

60. Heckman, J. (1979): Sample Selection Bias as A Specification Error, *Econometrica: Journal of the Econometric Society,* 47(1), 153-161.

61. Himaz, R. (2010): Intra household Allocation of Education Expenditure: The Case of Sri Lanka, *Economic Development and Cultural Change,* 58(2), 231-258.

62. Hori, Masahiro, Mitsuyama, N. and Shimizutani, S. (2015): New Evidence on Intra Household Allocation of Resources in Japanese Households, ESRI (Economic and Social Research Institute) Discussion Paper Series No.321.

63. Kabeer, N. (2003): Gender Inequality in Educational Outcomes: A Household Perspective. *Paper commissioned for the Education for All Global Monitoring Report*, *4*.

64. Kambhampati, U. S., and Pal, S. (2001). Role of parental literacy in explaining gender difference: Evidence from child schooling in India. *The European Journal of Development Research*, 13(2), 97-119.

65. Kambhampati, U. S. (2008): Does Household Expenditure on Education in India Depend Upon the Returns to Education?, *Henley Business School University Whiteknights Reading RG6 6AA United Kingdom*.

66. Kidd, Michael P., and Shannon, M. (1996): The Gender Wage Gap: A Comparison of Australia and Canada. *Industrial and Labor Relations Review,* 49(4), 729-746.

67. Kidd, M. P., and Shannon, M. (1994). An Update And Extension of The Canadian Evidence on Gender Wage Differentials. *Canadian Journal of Economics*, 27(4), 918-938.

68. Kingdon, G. G. (2005): Where Has All The Bias Gone? Detecting Gender Bias in the Intrahousehold Allocation of Educational Expenditure, *Economic Development and Cultural Change*, 53(2), 409-451.

69. Kingdon, G. G. (1998): Does the Labour Market Explain Lower Female Schooling in India?, *Journal of Development Studies*, 35(1), 39-65.

70. Kingdon, G. G. (2002): The Gender Gap in Educational Attainment in India: How Much Can Be Explained?, *Journal of Development Studies*, 39(2), 25-53.

71. Lancaster, G., Maitra, P., and Ray, R. (2006): Endogenous Intra- household Balance of Power and its Impact on Expenditure Patterns: Evidence from India, *Economica*, 73(291), 435-460.

72. Lancaster, G., Maitra, P., and Ray, R. (2008). Household Expenditure Patterns And Gender Bias: Evidence From Selected Indian States. *Oxford Development Studies*, 36(2), 133-157.

73. Lee, L.F. (1983). Generalized Econometric Models With Selectivity. *Econometrica: Journal of the Econometric Society*, 51(2), 507-512.

74. Lewbel, A., and Pendakur, K. (2008). Estimation of Collective Household Models with Engel Curves. *Journal of Econometrics*, 147(2), 350-358.

75. Lewbel, A., and Pendakur, K. (2013). Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients. *Report, Boston College*.

76. Lise, J., and Seitz, S. (2011). Consumption Inequality and Intra-Household Allocations. *The Review of Economic Studies*, 78(1), 328-355.

77. Liu, P. W., Zhang, J., and Chong, S. C. (2004). Occupational Segregation and Wage Differentials between Natives and Immigrants: Evidence from Hong Kong. *Journal of Development Economics*, 73(1), 395-413.

78. Li, L., & Wu, X. (2011). Gender of children, bargaining power, and intrahousehold resource allocation in China. *Journal of Human Resources*, *46*(2), 295-316.

79. Lundberg, S. J., Pollak, R. A., and Wales, T. J. (1997). Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit. *Journal of Human resources*, 32(3), 463-480.

80. Lundberg, S., and Pollak, R. A. (1993). Separate Spheres Bargaining and the Marriage Market. *Journal of Political Economy*, 101(6), 988-1010.

81. Mabsout, R., & Van Staveren, I. (2010). Disentangling bargaining power from individual and household level to institutions: Evidence on women's position in Ethiopia. *World Development*, *38*(5), 783-796.

82. Madheswaran, S., and Attewell, P. (2007). Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey. *Economic and Political Weekly*, October 13, 4146-4153.

83. Majid, H. (2012). *Parental Decision-Making and Intra-Household Resource Allocation* (Doctoral thesis, The Ohio State University).

84. Majumder, A and Mitra, C. (2016). Gender Bias in Household Education Expenditure: The Case of West Bengal. *Indian Growth and Development Review*, 9(2), 129-150.

85. Majumder, A., and Mitra, C. (2017). Gender Bias in Education in West Bengal. *Journal of Quantitative Economics*, 15(1), 173-196.

86. McElroy, M. B. (1990). The Empirical Content of Nash-Bargained Household Behavior. *Journal of Human Resources*, October 1, 559-583.

87. McElroy, M. B., and Horney, M. J. (1981). Nash-Bargained Household Decisions: Toward A Generalization of the Theory Of Demand. *International Economic Review*, June 1, 333-349.

88. Meng, X., Miller, P. (1995): Occupational Segregation and Its Impact on Gender Discrimination in China's Rural Industrial Sector. *Oxford Economic Papers*, January 1, 136-155.

89. Menon, M., Pendakur, K., and Perali, F. (2012). On The Expenditure-Dependence of Children's Resource Shares. *Economics Letters*, 117(3), 739-742.

90. Miller, P. W. (1987). The Wage Effect of the Occupational Segregation of Women in Britain. *The Economic Journal*, 9*7*(388), 885-896.

91. Mishra, M.and Chatterjee, S. (2017). School Dropout Susceptibility Mapping With Fuzzy Logic–A Study in the District of Purulia, India. *Human Geographies*, 11(1), 111-128.

92. Moll, P. G. (1991). The Decline of Discrimination Against Colored People in South Africa, 1970 To 1980. *Journal of Development Economics*, *37*(12), 289-307.

93. Mohanty, I. (2006). Gender Discrimination in Child Schooling: Why Do We Observe The Gender Disparity?. *University of Sydney*. (see: https://pdfs.semanticscholar.org/5ce2/6029f7e06af419f804ce284a7058a6033469.pdf)

94. Nussbaum, M. C., Glover, J. and Chen, M. (1995). Women, Culture, and Development: A Study of Human Capabilities. Clarendon Press.

95. Nithiya Amirtham, S. and Kundupuzhakkal, S. (2017). Gender Issues and Dropout Rates in India: Major Barrier in Providing Education for All. *Educationia Confab*, 2(4), 226-233.

96. OECD (2008), Growing Unequal? Income Distribution and Poverty in OECD Countries, Paris.

97. OECD (2011), Divided We Stand – Why Inequality Keeps Rising, Paris.

98. Omolo, M. (2011). Intra Household Resource Allocation in Kenya. African Economic Research Consortium.
(see: https://www.issuelab.org/resources/21914/21914.pdf)

99. Pal, S. (2004). How Much of The Gender Difference in Child School Enrolment Can Be Explained? Evidence from Rural India. *Bulletin of Economic Research*, *56*(2), 133-158.

100. Parpiev, Z., and Yusupov, K. (2009). Intrahousehold Gender Based Discrimination among Children in Uzbekistan (No. 09/10e). EERC Research Network, Russia and CIS.

101. Perali, F. (2008). The Second Engel Law: Is It a Paradox? *European Economic Review, 52*(8), 1353-1377.

102. Phipps, S. A., and Burton, P. S. (1998). What's Mine Is Yours? The Influence of Male and Female Incomes on Patterns of Household Expenditure. *Economica*, 65(260), 599-613.

103. Prakash, R., Beattie, T., Javalkar, P., Bhattacharjee, P., Ramanaik, S., Thalinja, R. and Collumbien, M. (2017). Correlates of School Dropout and Absenteeism Among Adolescent Girls from Marginalized Community in North Karnataka, South India. *Journal of Adolescence*, 61, 64-76.

104. Quisumbing, A. R., and Maluccio, J. A. (2000). Intra household Allocation and Gender Relations: New Empirical Evidence from Four Developing Countries. Washington, DC: International Food Policy Research Institute.
(see: http://siteresources.worldbank.org/INTGENDER/Resources/wp2.pdf)

105. Ray, R. (1983). Measuring The Costs of Children: An Alternative Approach. *Journal of Public Economics*, 22(1), 89-102.

106. Reggio, I. (2011). The Influence of The Mother's Power on Her Child's Labor in Mexico. *Journal of Development Economics*, 96(1), 95-105.

107. Rothbarth, E. (1943). Note on A Method Of Determining Equivalent Income for Families of Different Composition. *Appendix IV in War-Time Pattern of Saving and Expenditure by Charles Madge, University Press, Cambridge*.

108. Rumberger, R. W. (2001). Why Students Drop out of School and What Can Be Done. Paper prepared for the Conference, "Dropouts in America: How Severe is the Problem? What Do We Know about Intervention and Prevention?" Harvard University, January 13, 2001.

109. Rustagi, P. (2005). Understanding Gender Inequalities in Wages And Incomes in India. *The Indian Journal of Labour Economics*, 48(2), 319-334.

110. Schmidt, P. and Strauss, R., (1975). The Prediction of Occupations Using Multiple Logit Models. *International Economic Review,* 16(2), 471-486.

111. Shahapur P.N. and Omprakash H.M. (2017), Rural Education in India: A Scenario, *Scholarly Research Journal for Interdisciplinary Studies*, 4(31), 5268-5272.

112. Stearns, E. and Glennie, E. J. (2006). When and Why Dropouts Leave High School, *Youth and Society*, 38(1), 29-57.

113. Subramanian, S., and Deaton, A. (1991). Gender Effects in Indian Consumption Patterns. *Sarvekshana*, 14(4), 1-12.

114. Tilak, J. B. (2002): Education Poverty in India, *Review of Development and Change*, 7(1), 1-44.

115. Tisdell, C. (2002). Gender Inequality in India: Evidence from a Survey in Rural West Bengal. *International Journal of Social Economics*, 29(9), 706-721.

116. Varkkey, B. and Korde, R. (2013). Gender Pay Gap in the Formal Sector: 2006 – 2013: Preliminary Evidences from Paycheck India Data, Wage Indicator Data Report, eSocial Sciences Working Papers.

117. Vermeulen, F. (2002). Collective Household Models: Principles and Main Results. *Journal of Economic Surveys*, 16(4), 533-564.

118. Zimmermann, L. (2012). Reconsidering Gender Bias in Intra-Household Allocation in India. *Journal of Development Studies*, 48(1), 151-163.