Discussion Paper ERU/2015 – 02
September, 2015

Gender Bias In Education In West Bengal

Amita Majumder and Chayanika Mitra
Economic Research Unit, Indian Statistical Institute
203 B. T. Road, Kolkata-700108, India

Abstract

This paper attempts to capture gender bias at two different levels of education, namely, below class-10 and above class-10 using NSSO 64th round education expenditure data on West Bengal. The analysis for the below class-10 level involves an intra household framework and Heckman's two step model. For above class-10 level gender bias has been captured through a multinomial logit model for selection of subjects across households.

1 Introduction

In India substantial gender bias exists in allocation of goods and services.[1] Evidences of such discrimination at national/regional level brings forth the issue of discrimination in household consumption pattern. In intra household consumption pattern, there exist literatures (e.g. Deaton(1989), Deaton and Subramanian(1991)) that support existence of pro male bias in allocation of food articles. In the field of education Gandhi (2013) notes that even though at both primary and upper primary stage there have been almost similar percentage increase in promotion rate between the two years for both boys and girls [Govt of India, 2013], still there exists sufficient gender bias within households, the reasons being differential spending on sons and daughters within a household and differential decision making in sending a girl child and a boy child to education institution. According to Gandhi and Dreze

[1] The report of Govt. of India (1988) has evidences of excess mortality in female infants.

(2001), 54% female children are out of school in rural India. The discontinuation rate is close to 50% in West Bengal in 2009-2010 [Govt. of India, 2013]. 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 doing work to earn money. Girls have been found to be more victimized as households consider investment on a girl child for education an additional monetary burden over and above the dowry and expenditure on marriage. On the other hand, investment on a boy child is considered more fruitful as he would turn out to be another bread earner for the household. All these studies on discrimination have, however, been based on boys and girls at an aggregate level, without any reference to the different levels of education. In this paper we distinguish two broad levels, viz., 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 alike education institution or not and second, given that the child will be sent to an educational institute, the amount spent on the child for education [Gandhi (2003, 2013); Deaton (1997)]. On the other hand, in the above class 10 level, the households' decision involves the issue of choice of specialization (e.g., Science, Arts, Commerce etc.). The fact that higher studies calls for higher cost, it is important to study the nature of gender bias at this level separately. 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.[2] The model specification is based on the wage differential model that was used by Liu, Zhnag and Chong (2004) to capture inter occupational and intra occupational wage differential between natives and immigrants using data of Hong Kong.

The plan of the paper is as follows. In Section 2 the models used in this paper i.e., Heckman's two step model (Section 2.1) and wage differential model (Section 2.2) have been discussed. Section 3 presents the results and discussion with a brief description of the data used (Section 3.1) and the methodology of the estimation procedure (Section 3.2). The paper concludes in Section 4.

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 modeled 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 educational institute, they decide the share of the monthly spending of a household

[2]The Brown et al. (1980) approach has been used in many studies of wage differentials by gender (e.g., Dolton and Kidd (1994), Kidd and Shannon (1994, 1996), Meng and Miller (1995)), by race (e.g., Gabriel and Schmitz (1989), Moll (1992)), by immigration status (e.g., Liu, Zhang and Chong (2004)) and by Indian social group (e.g., Madheswaran and Attewell (2007)).

to be spent on the child. 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. These models are discussed in Section 2.1 and Section 2.2.

2.1 Heckman's model for below class-10 level

Sample selection model is a model where the participants who are involved in the activity of interest are deliberately over sampled. 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 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.[3]

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

[3]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 amount. Since no information regarding this can be retrieved from the available data, we use Heckman's model.

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 if $Y_1^* > 0$

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

The participation equation is given by $Y_1 = X_1^{*/}\beta_1 + \epsilon_1$ .... (1)

where ϵ_1 is a standard normal variate[4] 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

$Y_2 = X_2^{*/}\beta_2 + \epsilon_2$ .... (2)

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

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

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

The OLS regression of Y_2 on X_2^* will be inconsistent when the error terms in equation (1) and equation (2) are correlated.

The conditional censored mean in this framework will be as follows

[4]See section 14.4.1 in Cameron and Trivedi (2005).

$E[Y_2|X_1^*, X_2^*, Y_1^*] = E[X_2^{*/}\beta_2 + \epsilon_2|X_1^{*/}\beta_1 + \epsilon_1 \geq 0]$

$= X_2^{*/}\beta_2 + E[\epsilon_2|\epsilon_1 \geq -X_1^{*/}\beta_1]$ .... (3)

To obtain $E[\epsilon_2|\epsilon_1 \geq -X_1^{*/}\beta_1]$, 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, $\lambda(X_1^{*/}\beta_1) = \phi(X_1^{*/}\beta_1)/\Phi(X_1^{*/}\beta_1)$, which is the inverse Mill's ratio and σ_{12} is the correlation between two error terms.

Hence, the estimating equation is given by

$Y_2 = X_2^{*/}\beta_2 + \sigma_{12}\lambda(X_1^{*/}\hat{\beta}_1) + \epsilon_2$ .... (4)

The estimate of $\hat{\beta}_1$ is obtained by a probit regression and estimate of $\hat{\beta}_2$ is obtained by applying OLS to equation (4). Gender bias is detemined through appropriate parameters in the parameter vectors β_1 and β_2.

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 will be:

Male: $\ln[Y_j^l] = \alpha_j^m + X_j^{l/}\beta_j^m + \epsilon_j^{ml}$ .... (5)

Female: $\ln[Y_j^l] = \alpha_j^f + X_j^{l/}\beta_j^f + \epsilon_j^{fl}$ .... (6)

$j = 1, 2, \cdots J,$

where superscripts m and f denote males and females, respectively, subscript j denotes subject stream, 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 (5) and (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 argue 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.[5]

$$\underbrace{\overline{\ln[Y_j^m]} - \overline{\ln[Y_j^f]}}_{\text{R}} = \underbrace{(\hat{\alpha_j^m} - \hat{\alpha_j^f})}_{\text{U}} + \underbrace{(\overline{X_j^m} - \overline{X_j^f})\hat{\beta_j^m}}_{\text{E}} + \underbrace{\overline{X_J^f}(\hat{\beta_j^m} - \hat{\beta_j^f})}_{\text{C}} \ldots\ldots\ldots\ldots\ldots (7)$$

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 defines D = C+U as a 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 explicit inclusion of the probability of being able to choose a certain

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

chosen 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 (7) as follows:

$$\begin{aligned}
&\overline{\ln(Y^m)} - \overline{\ln(Y^f)} = \textstyle\sum_j [P_j^m \overline{\ln Y_j^m} - P_j^f \overline{\ln Y_j^f}] \\
&= \textstyle\sum_j (P_j^m \hat{\alpha_j^m} - P_j^f \hat{\alpha_j^f}) + \sum_j (P_j^m \overline{X_j^m} \hat{\beta_j^m} - P_j^f \overline{X_j^f} \hat{\beta_j^f}) \\
&= \underbrace{\underbrace{\sum_j P_j^f (\hat{\alpha_j^m} - \hat{\alpha_j^f}) + \sum_j P_j^f \overline{X_j^f} (\hat{\beta_j^m} - \hat{\beta_j^f})}_{\text{(WU)}} + \underbrace{\sum_j P_j^f (\overline{X_j^m} - \overline{X_j^f}) \hat{\beta_j^m}}_{\text{(WE)}}}_{\text{(Due to expenditure)}} + \underbrace{\sum_j (\overline{X_j^m} \hat{\beta_j^m} + \hat{\alpha_j^m})(P_j^m - P_j^f)}_{\text{(Due to subject)}}
\end{aligned} \quad \ldots\ldots (8)$$

Here the expenditure differential has been further decomposed into the portion attributable to differences in coefficients of (5) and (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:

$$\overline{\ln(Y_j^m)} - \overline{\ln(Y_j^f)} = WU + WE + \underbrace{\sum_j \overline{\ln Y_j^m} (P_j^m - \hat{P_j^f})}_{\text{(BE)}} + \underbrace{\sum_j \overline{\ln Y_j^m} (\hat{P_j^f} - P_j^f)}_{\text{(BU)}} \quad \ldots\ldots (9)$$

where $\hat{P_j^f}$ 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 ($\hat{P}_j^f$). Since $\hat{P}_j^f$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.

2.2.1 PREDICTING EDUCATIONAL DISTRIBUTION

Multinomial Logit Model of Educational Attainment (estimation of $\hat{P}_j^f$):

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

$$P_{lj} = P(T = j^{th} stream) = \frac{e^{Z_j^l \gamma_j}}{\sum e^{Z_j^l \gamma_j}} \ldots\ldots (10)$$

Z_j represents a vector of exogenous variables assumed to be influential in selection and availability of the stream, and γ_k is a vector of coefficients for these variables corresponding to the k^{th} stream. The qualitative dependent variable T can take on any of the J possible values, each corresponding to different streams. 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 were 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., $\hat{P_j^f}$.

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 (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:

$$\ln Y_j = X_j\beta_j - \sigma_j\rho_j\frac{\phi[(\tau(Z_j\gamma_j))]}{F(Z_j\gamma_j)} + \epsilon_j \qquad (11)$$

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, $\tau = \Phi^{-1}(F)$

F is the distribution of multinomial logit model.

Hence, the transformed expenditure function will look like,

$$\ln Y_j = X_j\beta_j + \hat{\lambda}_j\theta_j + \epsilon_j \qquad (12)$$

$\hat{\lambda}_j = -\frac{\phi[\tau(Z_j\hat{\gamma}_j)]}{F(Z_j\hat{\gamma}_j)}$

$\theta_j = \sigma_j\rho_j$.[6]

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 (9). 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 implicit when stream is ignored or incorrectly included. The second calculation examines the separate effects of within- and between-stream expenditure differences.

[6] 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 Results and discussion

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. Moreover, from this data set a complete profile of household education expenditure pattern at the individual level can also be obtained. The analysis has been done in two parts: below class-10 level and above class-10 level. The expenditure categories on education are given in Appendix Table A1. For the above class-10 level the different codes for different subjects are given in Appendix Tables A2 and A3. In West Bengal, there are 29442 households who have school or college going children aged 5-29 years. 31.97% of the households have zero spending on education in urban area and 41.36% households have zero spending on education in rural areas. But, strikingly, those households who spend on education, spend approximately 33% and 26% of their monthly per capita spending in urban and rural areas, respectively. Table 1 presents the distribution of students by age. Overall, the percentage of boys attending school is higher than that of girls in the below class-10 level (age 5-16 years) and in the above class 10 level in both sectors. However, 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 high 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.

3.2 Methodology

The estimation has been carried out at two different levels of education using two different methodologies described below.

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 Deaton and Subramanian (1991) and Gandhi (2003). Four broad age groups for children have been categorized as 0-4 year, 5-9 years, 10-14 years, 15-16 years and n_{ji}/n_i is the fraction of household members for household-i in the j^{th} age group. As the fractions n_{ji}/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 yeas 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

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

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

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

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

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

$[n_{Si}/n_i]_f$ 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 (s_i>0) or not ($s_i = 0$), s_i being the share of education expenditure for household i. The explanatory variables are education level of household head (HHHEAD'S EDU) [7], household size (n_i), logarithm of total expenditure ($\log(x_i)$), square of logarithm of total expenditure ($[\log(x_i)]^2$), $[n_{Pi}/n_i]_m$, $[n_{Pi}/n_i]_f$, $[n_{Mi}/n_i]_m$, $[n_{Mi}/n_i]_f$, $[n_{Si}/n_i]_m$, $[n_{Si}/n_i]_f$ and distance from residence to educational institute (DISTANCE).

The second stage equation is given by,

$$s_i = \alpha + \beta_1 \log(x_i) + \beta_2[\log(x_i)]^2 + \sum_{j=P,M,S} \theta_{jm}(n_{ji}/n_i)_m + \sum_{j=P,M,S} \theta_{jf}(n_{ji}/n_i)_f + \eta z_i + \gamma \ln(n_i) + \delta\hat{\lambda} + \epsilon_i \qquad (13)$$

Here, z_i denotes amount of land possessed, $\ln n_i$ allows for individual scale effect, θ_j reflects the effect of changing household composition without changing the household size, like- replacing a boy by a girl in a given age group. Testing gender difference reduces to testing $\theta_{jm} = \theta_{jf}$, i.e., $\theta_{jm} - \theta_{jf} = 0$, where j= P, M, S in equation (13) and similar test for the corresponding coefficients in the probit model.

3.2.2 The estimation procedure for above class 10 level

The estimation is carried out into two stages. First, estimates of the coefficients vector $\hat{\gamma}_j$ in the multinomial logit model for equation (10) are obtained by maximum

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

likelihood method for male students to obtain the counterfactual probabilities for female students using the esimated vector $\hat{\gamma_j}$. In the multinomial logit model we choose Commerce as the base outcome the dependent variable is 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 $\log(x_i)$, $[\log(x_i)]^2$, 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,[8] log of per capita expenditure on education ($\log(eduexp)$), number of courses studied by the student before (NUMBER)(more courses means more competence in education), HHHEAD'S EDU, n_i, D_{2i} is a dummy variable representing whether household is female headed or not (D_{2i}=1 if household head is female and 0 otherwise).

Next, coefficients ($\hat{\gamma_k}$) are used to calculate $\hat{\lambda}$ and then to run an OLS regression for equation (12) for each stream, each sex for the urban sector.[9]. Here, X_j is a vector of explanatory variables which includes $\log(x_i)$, $[\log(x_i)]^2$, DISTANCE, n_i, OTHEREXP (total expenditure - education expenditure), D_{2i} and $\hat{\lambda}$.

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

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

3.3 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 2.

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

For the probit model, in the urban and the rural sector both, the sign of marginal effect of $\log(x_i)$ is positive and that of $[\log(x_i)]^2$ is negative. However, for the urban sector the effects are nonsignificant. Thus for urban households income is not a determining factor for sending children to school. The household head's education level (HHHEAD'S EDU) has positive and significant (at 10% level) impact on the decision to spend on education in the urban sector, but is nonsignificant in the rural sector.[10] The marginal effect of household size (n_i) has a signifcant negative impact in both sectors (significant at 6% level in the rural sector). The marginal effects of the proportions of males and female students aged 10-14 years, 15-16 years are negative and significant. For the age group of 5-9 years in both sectors the effect is positive and significant. The increase in proportion of children in this age group makes the household better off due to the economies of scale (Gandhi, 2003). The effect of distance from residence to educational institute (DISTANCE) has significant negative impact on the decision to spend for education as higher distance leads to higher transportation cost for a household. Age of the student at entry to school (AGE) has a significant negative impact. Since higher age of the child makes him/her

[10] The average education level of household head is lesser in the rural sector (7.67) than that of the urban sector (11.87). The percentage of educated household is higher in the urban sector (24.36%) than the rural sector (4.32%), which is possibly why the household head's education is playing insignificant role in the decision to spend on education in the rural sector.

more suitable for household jobs or other works, if at low age a child is not sent to school, he/she is more likely not to go to school in future also.

For the second stage conditional OLS, coefficient of $\log(x_i)$ is positive for both the urban and the rural sector but nonsignificant in the rural sector. The coefficient of $[\log(x_i)]^2$ is negative for both sectors but nonsignificant for the rural sector. Thus, for rural households once they decide to send their child to school, income does not matter anymore. On the contrary, at this stage for urban households income is a significant factor possibly because of various choices of schools. The coefficient of log of household size ($\log(n_i)$) is negative for both sectors, but significant for the rural sector at 6.3% level of significance. The fraction of children belonging to age group of 5-9 years in a household has a nonsignificant negative effect on s_i for the urban sector for both males and females. In the rural sector the effect is negative for both sectors, but nonsignificant for males. 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 signs of the coefficients of the proportion of children belonging to age groups of 10-14 years, 15-16 years clearly imply that at this stage children are not discouraged from education. Amount of land possessed (LAND) is nonsignificant in the urban sector but has significant and positive impact on share of spending on education at 9.4% level of significance in the rural sector. 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 lead to inconsistent results.

In Table 3 the results of testing the restrictions $\theta_{Pm} - \theta_{Pf} = 0, \theta_{Mm} - \theta_{Mf} = 0, \theta_{Sm} - \theta_{Sf} = 0$ in the probit regression and the conditional OLS in the rural and the urban sectors are reported. In case of deciding whether to spend on education or not, the values of $\theta_{Pm} - \theta_{Pf}, \theta_{Mm} - \theta_{Mf}, \theta_{Sm} - \theta_{Sf}$ are positive and significant for age groups of 5-9 years, 10-14 years in the urban sector and for 5-9 years, 10-14 years, 15-16 years in the rural sector. This means that at the decision making stage there exists discrimination against girls at the below class-10 level in both sectors, except at the class 9-10 level in the urban sector. In case of the share of spending on education, the F-values are values positive but nonsignificant for both sectors, leading to the conclusion that the discrimination between two sexes exist at decision making stage, but not in the share of spending on education at the below class-10 level.

3.4 Discussion of results for above class-10 level

The empirical results of the multinomial logit model are reported in Table 4. The multinomial logit model has been fitted using Commerce as base outcome.

The coefficients of log of monthly expenditure ($\log(x_i)$) are negative and significant for Arts and positive and significant for Science for both sectors. It implies that with increase in monthly income the probability of preferring Arts over Commerce falls and probability of preferring 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 ($\log(eduexp)$). The coefficients of $[\log(x_i)]^2$ 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 prefer Commerce over Science. The explanatory variable-whether education is free or not (denoted by a dummy variable D_{1i}) has similar signs like $\log(x_i)$ 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 nonsignificant 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 number of courses studied rises the students prefer 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 prefer 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 (HHHEAD'S EDU) 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 less towards Arts and high towards Science compared to Commerce. The coefficient of D_{2i} is nonsignificant 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 5.

The reported values in the last row 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). It can be seen 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 studying Science and Commerce.

After estimating the predicted probabilities for each stream for both sexes, an OLS regression of equation (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 variables in each individual stream. We, therefore, concentrate on the urban sector, the results of which are reported in Table 6.

The coefficient of $\log(x_i)$ has a positive impact on education expenditure irrespective of sexes and streams, which means that with increase in income expenditure on education also rises, but at a decreasing rate, as the coefficient of $[\log(x_i)]^2$ 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 lead to inconsistent results. For the female students for Science and Commerce the estimates of λ turn out to be nonsignificant. This implies that choice of these subjects is largely random in the female data.

The results of the decomposition in equation (9) are presented in Table 7. The percentages of the total differential for WE, WU, BE, BU are 35.23%, 28.86%, 30.42% and 5.49%, respectively. Thus, 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, as noted in Section 2.2, there is pro male bias both at the level of choice of subjects, as well as in the expenditure allocation for studying the same subject.

4 CONCLUSION

In spite of provision of free and compulsory education the educational backwardness of India even after 60 years of independence is quite clear. This analysis establishes that there exists pro male bias in the decision to send the child to school in the below class-10 level in both the rural and the urban sectors, but not at the level of allocating expenditure on education once households decide to send their children to school. For the above class-10 level the detailed analysis indicates evidence of discrimination against girl students at the level of selection of subject stream, and also in the expenditure incurred within a subject stream in the urban areas the state of West Bengal during 2007-08.[11]

References

Azam, Mehtabul, and Geeta Gandhi Kingdon (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development 42: 143-164.*

Blinder, A.S., (1973). Wage discrimination: reduced form and structural estimates.*Journal of Human Resources 8, 436 455.*

Brown, R.S., Moon, M., Zoloth, S., (1980). Incorporating occupational attainment in studies of male female differentials. *Journal of Human Resources 15, 3 28.*

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

Cragg, J.G (1971). Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica. Vol. 39. No.5 (pp 829-844).*

Deaton, A. (1989). Looking for Boy-Girl Discrimination in Household Expenditures , *World Bank Economic Review 3, no. 1: 1-15.*

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

Deaton, Angus (1997). The analysis of household surveys: a microeconometric approach to development policy. *World Bank Publications.*

Deaton, A. and Shankar Subramanian (1991). Gender bias in Indian consumption pattern, *Sarvekshana 14, no. 4: 1-12.*

Dolton, Peter J., and Michael P. Kidd (1994). Occupational access and wage discrimination. *Oxford Bulletin of Economics and Statistics 56.4: 457-474.*

Gabriel, Paul E., and Susanne Schmitz (1989). The efect of occupational segregation on the relative earnings of black males. *Economics Letters 30: 77-83.*

Government of India, (1988). Child mortality estimates of India, Census of India 1981, Occasional paper. Number 5 of 1988, Demography Division, Office of Registrar General, New Delhi.

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.

Heckman, James J (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society: 153-161.*

Kidd, Michael P., and Michael Shannon (1996). The gender wage gap: a comparison of Australia and Canada . *Industrial and Labor Relations Review 49.4: 729-746.*

Kingdon, Geeta Gandhi and Jean Dreze (2001). School Participation in Rural India. *Review of Development Economics.*

Lee, L., F., (1983). Generalized econometric models with selectivity. *Econometrica 51, 507-512.*

Liu , Pak-Wai.Zhang, Junsen.,Chong, Shu-Chuen., (2004). Occupational segregation and wage differentials between natives and immigrants: evidence from Hong Kong. *Journal of Development Economics,73: 395-413.*

Madheswaran, S., and Paul Attewell (2007). Caste discrimination in the Indian urban labour market: Evidence from the National Sample Survey. *Economic and Political Weekly: 4146-4153.*

Meng, X., Miller, P., (1995). Occupational segregation and its impact on gender wage discrimination in China's rural industrial sector. *Oxford Economic Papers 47, 136-155.*

Miller, p., (1987). The wage effect of the occupational segregation of women in Britain. *Economic Journal 97,885.*

Moll, P. G. (1992). The decline of discrimination against coloured people in South Africa, 1970 to 1980. *Journal of Development Economics 37, 298-307.*

Schmidt, P., Strauss, R., (1975). The prediction of occupations using multiple logit models. *International Economic Review 16, 471-486.*

Table 1: THE DISTRIBUTION OF STUDENTS ATTENDING EDUCATIONAL INSTITUTION

SEX	AGE									
	5-16 years							17-29 years		
				Rural		TOTAL		Urban		
	Rural	Urban	Arts	Science	Commerce		Arts	Science	Commerce	TOTAL
MALE	51.06	53.90	52.22	13.12	2.02	67.36	19.79	11.46	13.02	54.27
FEMALE	48.94	46.10	32.38	0.26	0.00	32.64	32.29	9.80	3.64	45.73
TOTAL	100	100	84.60	13.38	2.02	100	52.08	21.26	16.66	100

Table 2: THE RESULTS FOR BELOW CLASS-10 LEVEL

VARIABLES	SECTORS			
	RURAL		URBAN	
	PROBIT(s_i>0 or =0)**	OLS(s_i)	PROBIT	OLS
log(x_i)	0.852 (0.000)*	.434 (0.265)	1.121 (0.430)	.709 (0.012)
$[log(x_i)]^2$	-.0488 (0.000)	-.026 (0.281)	-.028 (0.730)	-.040 (0.018)
n_i	-.004 (0.058)		-.066 (0.005)	
log(n_i)		-.044 (0.063)		-.028 (0.274)
HHHED'S EDU	.024 (0.206)		.577 (0.085)	
DISTANCE	-.029 (0.000)		-.208 (0.000)	
AGE	-.024 (0.002)		-.025 (0.020)	
LAND		.008 (0.094)		-.0004 (0.987)
λ		0.427 (0.000)		0.392 (0.000)
(5-9 YEARS)m	.295 (0.000)	-.119 (0.001)	1.344 (0.002)	-.027 (0.775)
(5-9 YEARS)f	.278 (0.000)	-.097 (0.242)	2.491 (0.003)	-.041 (0.722)
(10-14 YEARS)m	-.104 (0.003)	.457 (0.145)	-.817 (0.015)	.231 (0.003)
(10-14 YEARS)f	-.161 (0.001)	.378 (0.000)	-.895 (0.012)	.220 (0.009)
(15-16YEARS)m	-.344 (0.000)	.263 (0.042)	-1.527 (0.002)	.243 (0.047)
(15-16YEARS)f	-.395 (0.001)	.257 (0.052)	-2.206 (0.001)	.402 (0.005)
CONSTANT	-3.076 (0.001)	0.966 (0.000)	-2.322 (0.002)	1.607 (0.000)

*Figures in parentheses are the p-values.

** The expressions in the parentheses denote the dependent variables.

Table 3: VALUES OF F-STATISTIC (Testing $\theta_{jm} - \theta_{jf} = 0, j = P, M, S.$) FOR BELOW CLASS-10 LEVEL

EDUCATION (AGE)	SECTORS			
	RURAL		URBAN	
	PROBIT	OLS	PROBIT	OLS
PRIMARY (5-9 YEARS)	2.725 (0.009)*	13.867 (0.785)	4.004 (0.0815)	16.456 (0.8521)
MIDDLE (10-14 YEARS)	4.089 (0.004)	15.409 (0.277)	3.907 (0.001)	18.007 (0.9045)
SECONDARY (15-16 YEARS)	1.005 (0.001)	17.903 (0.9718)	3.006 (0.2849)	19.605 (0.3494)

Figures in parentheses are the p-values.

Table 4: THE MLOGIT MARGINAL EFFECTS FOR ABOVE CLASS-10 LEVEL

VARIABLES	SECTORS			
	RURAL		URBAN	
	Arts($\log(\frac{Prob(T=A)}{Prob(T=C)})$)**	Science($\log(\frac{Prob(T=S)}{Prob(T=C)})$)	Arts	Science
$\log(x_i)$	-0.00039 (0.000)*	0.00001 (0.009)	-0.0013 (0.000)	0.0002 (0.000)
$[\log(x_i)]^2$	0.00005 (0.045)	-0.00034 (0.001)	0.0002 (0.001)	-0.0001 (0.001)
AGE	0.30307 (0.285)	-0.63282 (0.046)	0.8453 (0.978)	0.0097 (0.577)
D_{1i}	0.07864 (0.876)	0.23561 (0.987)	-1.0208 (0.030)	0.9037 (0.001)
$\log(eduexp)$	-2.46390 (0.011)	0.75227 (0.030)	-1.7645 (0.001)	0.5637 (0.001)
NUMBER	-1.67452 (.856)	1.4007 (0.001)	-1.8350 (0.071)	-17.4352 (0.994)
n_i	0.04634 (0.010)	-7.9562 (0.994)	0.0734 (0.001)	-0.1917 (0.007)
HHHEAD'S EDU	-13.62658 (0.001)	12.66512 (0.004)	-11.7453 (0.005)	9.2301 (0.030)
D_{2i}	-150.5417 (0.996)	-138.0271 (0.997)	90.5324 (0.998)	101.6009 (0.996)
CONSTANT	5.30834 (0.010)	3.05223 (0.000)	3.2008 (0.002)	4.7069 (0.001)

*Figures in parentheses are the p-values.

**The expression in parentheses is the dependent variable.

Note: Commerce is the base outcome.

Table 5: THE OBSERVED AND PREDICTED PROBABILITIES FROM MLOGIT

PROBABILITIES		SUBJECTS					
		Arts		Science		Commerce	
		RURAL	URBAN	RURAL	URBAN	RURAL	URBAN
OBSERVED	P_m	.8566	.5690	.1189	.2310	.0246	.2000
	P_f	.9524	.7369	.0476	.1657	.0000	.0947
PREDICTED	$\hat{P_m}$	.7981	.6043	.1602	.2314	.0415	.1641
	$\hat{P_f}*$	.6718	.3186	.3895	.2241	.0385	.3571

These are the counterfactual probabilities.

Table 6: THE OLS REGRESSION FOR ABOVE CLASS-10 LEVEL FOR URBAN SECTOR(dependent variable log($eduexp$))

VARIABLES	SUBJECTS					
	Arts		Science		Commerce	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
$\log(x_i)$	2.48622 (0.000)*	3.31437 (0.000)	1.97030 (0.000)	7.32478 (0.000)	1.05362 (0.000)	2.74239 (0.001)
$[\log(x_i)]]^2$	-0.16944 (0.000)	-0.21972 (0.000)	-0.13385 (0.000)	-0.45143 (0.000)	-0.00621 (0.001)	-0.16774 (0.001)
DISTANCE	0.00504 (0.038)	-0.0006 (0.992)	0.00168 (0.446)	0.00280 (0.804)	0.05091 (0.001)	.05769 (0.002)
OTHEREXP	-0.00008 (0.000)	-0.00005 (0.000)	-0.00007 (0.000)	0.04028 (0.997)	-0.00046 (0.016)	-0.000025 (0.016)
n_i	0.01291 (0.000)	0.00949 (0.058)	-0.00789 (0.033)	.01242 (0.203)	-0.02453 (0.006)	.05291 (0.152)
D_{2i}	NA** (NA)	-.005043 (0.998)	NA (NA)	.4876035 (0.000)	NA (NA)	NA (NA)
λ	0.08289 (0.001)	-0.00040 (0.001)	-.02296 (0.046)	-7.45907 (0.770)	0.00408 (0.001)	.03723 (0.166)
CONSTANT	-5.68102 (0.000)	-9.68340 (0.000)	-3.91489 (0.000)	-2.45907 (0.000)	-9.00716 (0.000)	-7.80180 (0.010)

*Figures in parentheses are the p-values

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

Table 7: THE EXPENDITURE DECOMPOSITION

TOTAL $\overline{\ln(Y^m)} - \overline{\ln(Y^f)}$	WE $\sum_j P_j^f(\hat{\alpha_j^m} - \hat{\alpha_j^f}) + \sum_j P_j^f \bar{X_j^f}(\hat{\beta_j^m} - \hat{\beta_j^f})$	WU $\sum_j P_j^f(\bar{X_j^m} - \bar{X_j^f})\hat{\beta_j^m}$	BE $\sum_j \overline{\ln Y_j^m}(\hat{P_j^f} - P_j^f)$	BU $\sum_j \overline{\ln Y_j^m}(P_j^m - \hat{P_j^f})$
8.43354	2.4344	2.97177	0.461844	2.56552
% contributed in total (100)	64.06%		35.91%	
% contributed in Within (100) and Between (100)	46.91%	53.09%	16.29%	83.71%

Appendix

Table A1: CATEGORIES OF EDUCATION EXPENDITURE (NSSO 64th ROUND 2007-08)

1	TUITION FEE
2	EXAMINATION FEE
3	OTHER FEES AND PAYMENTS
4	BOOKS
5	STATIONARY
6	UNIFORM
7	TRANSPORT
8	PRIVATE COACHING
9	OTHER EDUCATIONAL EXPENDITURE

Table A2: THE SUBJECT CODES PROVIDED BY NSSO

Subjects	Codes
General courses(upto 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: THE MODIFIED SUBJECT STREAMS FOR ANALYSIS

Streams	Subjects
Science (S)	Science,Medicine,Engineering, Agriculture
Arts (A)	Arts, Law, Education
Commerce (C)	Commerce, Management, CA and alike courses