



Socio-economic inequalities in spending on various levels of education across Indian households: an update

Aswathy Rachel Varughese¹ · Indrajit Bairagya²

Accepted: 17 June 2023 / Published online: 4 July 2023

© The Author(s) under exclusive licence to Editorial Office, Indian Economic Review 2023

Abstract

The study examines potential channels of socio-economic inequalities in education expenditure by Indian households at three different levels of education using National Sample Survey Data (2018). Based on Heckman's two-step model estimates, the inequalities are evident in the participation choice and education expenditure by social groups, place of residence and religious minorities for education up to the secondary level. Nevertheless, the economic status of the household is less critical. Gender inequality is more evident in the expenditure incurred than the enrollment choice at the secondary level. For education at higher secondary and above levels, the crucial channel of social inequality lies in choosing a subject specialization and subsequent expenditures. As per multinomial logit estimates, the choice of streams is highly selective, favoring those with the capacity and willingness to pay. The selection corrected expenditure based on Lee correction reveals the extent of socio-economic inequalities in education expenditures at these levels of education. The apparent disparity in participation choice and expenditure can only be addressed through calibrated policy interventions, especially at higher levels of education. Failing to resolve this may exacerbate inequalities in education and labor market participation.

Keywords Household education expenditure · Socio-economic inequalities · Engel curve · Heckman selection · Multinomial logit · Lee correction

JEL Classification I20 · I21 · I22

✉ Aswathy Rachel Varughese
iamaswathyvarughese@gmail.com; aswathy@isec.ac.in

Indrajit Bairagya
indrajitise@gmail.com

¹ Department of Applied Economics, Cochin University of Science and Technology (CUSAT), Cochin, Kerala 682022, India

² Institute for Social and Economic Change (ISEC), Centre for Human Resource Development, Dr. V K R V Rao Road, Nagarbhavi, Bengaluru 560072, India

1 Introduction

Resource allocation for education by the State and the household levels aids in equalizing educational opportunities and thereby promotes economic growth (Castelló-Climent & Mukhopadhyay, 2013). In India, households incur considerable expenditure at various levels of education with an increase in real income, improved demand for education and enhanced participation in private educational institutions (Agrawal, 2014). From 2014 to 2018 NSS rounds, household education expenditure grew at a Compounded Annual Growth Rate (CAGR) of 14.78% at the secondary level, while it registered a rise of 10.19%, 9.50% and 8.24% for education at higher secondary, graduation and post-graduation levels, respectively. Given increased household spending on education across various levels, what assumes importance is the varying pattern and allocation of resources, given the diverse socio-economic backgrounds of Indian households.

The studies such as Kingdon (2005), Aslam and Kingdon (2008), Azam and Kingdon (2013) and Datta and Kingdon (2019) have established the presence of a strong gender bias and rural–urban divide, while spending for education in India. The main deviation of the present study lies in examining the social inequalities in educational resource allocation and gender differences across three levels of education. Social inequalities are household characteristics (inter-household) while gender is primarily an individual characteristic. Therefore, this study considers per capita education expenditure and per capita household consumption expenditure at the individual level data to control for the household fixed effects.

Most studies addressed gender bias in intra-household educational resource allocation for school-going children up to 16 years. However, studies capturing such differences at the higher education level in India are found scanty. Educational spending at the household level involves a two-stage decision-making process (Datta & Kingdon, 2019; Kingdon, 2005). First, the decision to get children enrolled and second, the amount incurred on those enrolled. Based on the household and individual level data, using the Double Hurdle model, Kingdon (2005) and Datta and Kingdon (2019) separated these decision-making stages controlling for various age groups. This study extends the analyses to those who are pursuing higher education.

Specifically, the study analyses the dual-stage decision-making process for education spending across three different levels of education. From the Indian context, for education up to secondary¹ (up to Class X), the preliminary decision relates to sending the child to school (decision on enrollment). In that case, what should be the level of spending (decision on the expenditure)? At higher secondary² (Class XI and Class XII) and above levels³ (Graduation and Post-Graduation), the first-stage decision-making concerns the specialization of the subject they want to study. The Gross Enrollment Ratio (GER) in higher education in India stood at 27.1 for students between 18 and 23 years of age as per the AISHE report (2020). India's

¹ Up to secondary level refers to Classes from I to X.

² Various streams are different courses that students opt for such as Humanities, Commerce and Science.

³ Higher secondary refers to Class XI and XII, whereas above Higher secondary means classes above XII, Graduation and Post-graduation.

higher education GER is lower than other educationally well-performing countries, especially its East Asian peers. Given the poor enrollment ratio at this level, the socio-economic inequalities may be evident in participation. However, here a more insightful picture of socio-economic inequalities can be drawn from addressing the issue of choice of specialization of courses/streams for higher secondary and above levels.

Moreover, at higher levels of education, the supply is limited due to its quasi-public good nature, resulting in higher costs. Hence, socio-economic inequalities could be more revealing in the choice of courses and subsequent expenditures. Therefore, a stream's choice determines the expenditure level incurred in the second stage. Each student's subject specialization and socio-economic factors may determine differences in the amount incurred for each stream.

Up to the secondary level, education is mainly considered a public good with the state actively involved in its public provisioning. On the contrary, education is considered a private or quasi-public good beyond the secondary level, wherein public provision is limited. In the Indian context, the prime focus of decades of policy initiatives has been more on education up to the secondary level than higher levels. Despite the National Education Policy (1986) (NEP, 1986) giving a considerable thrust to expanding higher education, private higher educational institutions have flourished rapidly over the years. Higher tuition fees and the lack of reservation policies have hindered the enrollment of students from the deprived social and economic categories (Bohindar, 2019). With a lesser focus on the higher education sector in terms of financing, it is vital to consider the determinants of subject specialization and household spending on education while formulating education policies. From this viewpoint, a study on the potential channels of inequalities at different levels of education regarding educational resource allocation is exceedingly essential.

The recent studies by Aslam and Kingdon (2008) and (Datta & Kingdon, 2019) have pointed out the limitation of traditional estimation of the Engel curve using a single OLS for capturing the two-stage decision-making process. Following those studies, the present study estimated the Engel curve using the Heckman Selection model for analysis up to the secondary level. Meanwhile, using a Multinomial Logit model, we estimated the probability of choosing various courses at higher secondary and above levels. Lee correction gives the selection corrected expenditure for those choosing a particular subject. Therefore, this study separated the factors determining the choice and spending at each level of education.

The analysis reveals clear evidence of social inequalities in educational resource allocation across three levels of education. Up to the secondary level the participation choice and expenditure are influenced by the social groups, place of residence and religious minorities. Notably, the household's economic status does not influence schooling and related spending much. However, the most crucial channel of social inequalities lies in choosing subjects to be pursued at education above secondary levels. At higher levels of education, females, people from rural areas and those from socially and economically deprived groups choose the least expensive over expensive professional courses. These disparities have long-term implications for educational and occupational engagement, as professional and science streams are generally more job market oriented. In light of these findings, we argue that

the dearth of policy initiatives at the higher education level may further accentuate social inequalities in the labor market. The study utilizes 75th round National Sample Survey (NSS) data (2018), covering a period of one year (June 2017–June 2018) for the empirical analysis. Hitherto, similar studies used data up to the 71st NSS round (January 2014 and June 2014). Hence, it may be worth pondering social inequalities in education expenditure allocation at various levels of education across Indian households with the latest round of data.

The rest of the paper is organized as follows: Sect. 2 outlines the related literature; Sect. 3 the conceptual framework, data and econometric specifications; followed by a descriptive analysis in Sect. 4. Results of empirical estimation and discussions are presented in Sect. 5, followed by the conclusion.

2 Related literature

The literature argues that each country's socio-economic factors influence household-level education expenditure disparities. In the Turkish context, Tansel and Bircan (2006) argue that the demography of households, parents' education and other socio-economic factors such as gender, location and economic status influence educational expenditure at the household level, which ultimately affects inter-generational mobility. Income levels of the household are a prominent factor in the intra-household allocation of educational resources in Turkey. The educational expenditure elasticities at the household levels are comparatively lower for top- and bottom-income groups, but high for middle-income groups (Acar et al., 2016). In Pakistan, the gender gap in educational attainment is mainly because of the persisting gender disparity in the intra-household allocation of educational resources (Aslam & Kingdon, 2008). Asadullah and Chaudhury (2009) observed a reverse gender disparity in respect of Bangladesh for education up to secondary schooling; this gap is even wider in urban areas of the country. They claimed that the identified reverse gender gap could be the fallout of educational policy intervention known as the Female Secondary Stipend (FSS) program in Bangladesh.

From the Indian milieu, Kingdon (2005) identified strong gender bias in India at sub-national levels regarding household allocation of education expenditure. The deep-rooted social norms are often blamed for the persisting gender gaps. The practices such as early marriage and continuing dowry systems typically hinder the educational aspirations of girl children, especially in rural areas. In such areas, a small proportion of girls are allowed to attain higher education compared to their male counterparts, given that the socially acceptable age of marriage is much earlier for girls. Hence, the perceived returns to education differ drastically for boys and girls (Maertens, 2013). Girl children from socially and economically deprived communities are more vulnerable to these practices, resulting in poor educational outcomes. Moreover, the coerced spending on their marriages and continuing dowry system influence spending decisions for their education, leading to gender discrimination in intra-household resource allocation (Chiplunkar & Weaver, 2021). State-specific educational policy interventions such as Conditional Cash Transfer (CCT) schemes and the provision of bicycles for those girls attending secondary schools improved

their access to schools. The persisting socio-economic inequalities also accentuate disparities in intra-household educational resource allocation. The CCTs effectively channeled the limited resources to economically backward families (Sekher & Ram, 2015). Public financing compensates for household spending through such policies and such programs aid in reducing gender and social gaps in school participation. Nevertheless, the dearth of such initiatives at the higher levels of education may again discriminate against the deprived and the marginalized for selecting job-oriented courses as well as their related expenditure. Moreover, it will enhance participation, but the families' discriminatory spending patterns and treatment may impact their learning outcomes (Das & Sarkhel, 2020).

Studies such as Chaudhuri and Roy (2006); Lancaster et al. (2008); Saha (2013) confirmed the presence of gender bias using various rounds of NSS data at national and sub-national levels. They further observed that gender discrimination in allocating educational resources is prominent in backward states and developed or high-income states. Individual characteristics like gender and household characteristics such as place of residence, caste, religion and certain household factors influence the decision to spend more on education, particularly in a diverse society like India (Majumder & Mitra, 2017). Conventionally, most of the existing studies estimated various functional forms of Engel curves such as Working-Lesser, linear, semi-log, double-log, double-semi-log, hyperbolic, log inverse and log-log inverse using simple ordinary least square (OLS) or Tobit models. Studies such as Aslam and Kingdon (2008) and Datta and Kingdon (2019) argued that the conventional Engel curve approach would not suffice while detecting gender bias in the intra-household allocation of education resources. According to them, the traditional approach does not address the bias involved in the two-stage decision-making process of enrollment choice and expenditure decision. Thus, they estimated the Working-Lesser functional form of the Engel curve using hurdle models, controlling for various age groups. Datta and Kingdon (2019) found that gender discrimination in education spending changed drastically over the years, comparing the NSS rounds of 1995 and 2014. They identified potential channels of gender discrimination such as the bias regarding enrollment decisions (first stage) and subsequent spending decisions (second stage). The second stage is conditional education expenditure for those who are enrolled. Specifically, they underscored individual child-level analysis scores over household-level analysis, deploying household fixed effect. Majumder & Mitra (2017) found that West Bengal has a pro-male bias regarding the enrollment decision but not in the spending decisions for those enrolled for below class X. For above class X, female students are discriminated against while choosing the courses to be pursued in urban areas during 2007–08, 64th NSS round.

Hitherto, the available literature primarily focused on the gender gap in education spending decisions at an aggregate level while controlling for ages relevant for school-going children. There exists a dearth of studies focusing on socio-economic inequalities in the specialization of subjects at higher levels of education and related education expenditure. Therefore, the present study aims to fill the existing research gap by considering different levels of education and deploying appropriate empirical techniques for each level. As the inequalities differ for education above the secondary level, households' decision includes choosing courses or specialization. Given

the expensive nature of higher education, particularly graduation and post-graduation, it is essential to understand the socio-economic inequalities regarding education spending decisions at this level. The choice of a subject by the student claims importance while making the spending decision for education, implying the amount to be spent across various disciplines.

3 Conceptual framework, data and methodology

3.1 Educational expenditure up to secondary level

As referred to above, the choice of education spending constitutes the enrollment choice, a binary decision at the first stage, which may be non-linear. The second stage is a conditional expenditure, for positive education expenditure that may follow a log-normal distribution. It is specifically applicable to education up to the secondary level. Hence, as per Aslam & Kingdon (2008), conventionally estimating an Engel curve using a single OLS equation may be wrong to model two separate decision-making processes. Following Datta & Kingdon (2019), the Working-Lesser form in which the functional form of the Engel curve is corrected and estimated.

The Heckman selection model addresses observations with zero values in the dataset. In such a case, this method addresses the selection bias as a form of omitted variable bias that can be corrected by adding a control to the model that reflects the probability of selection into the sample. Hurdle models are also suitable for such observations, wherein two separate hurdles are crossed for incurring education expenditure. The participation (first hurdle) and quantity (second hurdle) equations give the socio-economic determinants of household education expenditure. With respect to education expenditure and schooling participation, some studies use either Heckman selection or the double-hurdle model to analyze such decision-making processes. The study presents results based on the Heckman selection model as the statistical structure of both models is similar (Ahmadzai, 2018; Humphreys, 2013). Moreover, the existing studies comparing the results in empirical analysis using both models testified that the results are on similar lines (Madden, 2008). Further, in the next level of this analysis, modeling the decision-making process with respect to the choice of courses or streams at the higher secondary and above levels, the study uses multinomial logit. Multinomial logit analyses the probability of enrolling for a particular subject and the related selection corrected expenditure using Lee correction. These analyses come as an extension of the sample selection rather than the hurdle model. Hence, the study analyses the dynamics of household spending on education up to the secondary level using the Heckman selection model and reports the results accordingly.

In the extant studies, to model the dual-stage decision-making process, the estimation of the Working-Lesser functional form of the Engel curve is deemed superior (Chaudhuri & Roy, 2006; Deaton, 1989). In the first stage, the Heckman selection model (1979) estimates the probability of a child getting selected for attending school, given the socio-economic background of the household. If the child is attending school, the amount to be incurred, based on the socio-economic

profile of the household, comprises the second stage (Deaton, 1989). The distribution of household educational expenditure encompasses a large set of data hovering at zero. The exclusion of samples with zero educational expenditure for those students not currently attending school leads to a sample selection bias (Hjortsberg, 2003). The usual OLS fails to reflect such cases with the consequence of being an inconsistent estimator (Wooldridge, 2010). The Heckman selection model (1977) considers the potential sample selection bias arising out of zero educational expenditure (You & Kobayashi, 2011). In the Heckman selection model, through a Probit model, the probability of incurring some expenditure or nil is analyzed using a participation equation. Then, given the positive likelihood, factors determining the spending are quantified using OLS and its outcome equation (Humphreys, 2013).

There are two latent variables here, Z_{1i}^* and Z_{2i}^*

The selection equation is expressed in terms of a latent variable Z_{1i}^* , wherein it depends on factors that influence the spending decision of the households.

The latent variable Z_{1i}^* cannot be directly observed here, It implies that the spending decision is represented by a binary variable Z_{1i} , wherein $Z_{1i} = 1$ implies that Z_{1i}^* has a strictly positive value, meaning the household is willing to send their children for attending educational institutions.

$$Z_{1i} = 1, \text{ if } Z_{1i}^* > 0,$$

$$Z_{1i} = 0, \text{ otherwise.}$$

The selection equation is as follows:

$$Z_{1i}^* = \alpha_0 + \alpha_1 \ln(\text{PerCapita Total Household Expenditure}_i) + \alpha_2 \ln(\text{PerCapita Total Household Expenditure}_i)^2 + \alpha_3 X_i + \alpha_4 Y_i + \varepsilon_i \quad (1)$$

where X_i captures household- and individual-specific characteristics such as social group, religious minorities and gender, while Y_i is the vector of control variables influencing education participation and ε_i is the error term. Out of the total sample, those values are only observed when the household educational expenditure is greater than zero. This is the first stage which gives the probability of children going to school or not, determined by the socio-economic backgrounds of the households. The exclusion restriction in the Heckman selection model here is reflected through the variable, households having a computer with internet connections. According to Johnson et al. (2016), the tangible asset base's ownership pattern influences families' schooling participation. However, due to the unavailability of direct data from the database used here, the households having a computer with an internet connection are used for capturing the tangible asset base that may directly impact the choice of education but not the expenditure. The weak correlation of the asset base variable (households having a computer with an internet connection) with education expenditure (0.04) without statistical significance ascertains the feasibility of this variable as an exclusion criterion. The strong statistically significant correlation

of the instrument in the selection equation empirically supports the identification strategy.

The second latent variable, Z_{2i}^* (per student education budget share of household i) which can only be observed when the household is spending on education and cannot be observed otherwise (if, $Z_{1i} = 0$). The outcome equation represents the observed value of Z_{2i}^* which is represented as (*Lnedushare*) in Eq. (2) below.

The variable $\ln(\text{householdsize}_i)$ allows for individual scale effect.

Therefore, the second stage of conditional OLS regression for all positive educational expenditures can be written as:

$$\begin{aligned} \text{Lnedushare} = & a_1 + a_2 \ln(\text{Per capita Total Household Expenditure}_i) \\ & + a_3 \ln(\text{Per capita Total Household Expenditure}_i)^2 \quad (2) \\ & + a_4 X_i + a_5 Y_i + \beta \lambda_i + \epsilon_i, \end{aligned}$$

where X_i captures the household and individual characteristics, whereas Y_i are the control variables. In Eq. (2), certain control variables have been incorporated that influence the expenditure like distance capturing the proximity of educational institutions and the type of institutions capturing whether the student enrolled in a government or private institution. These variables influence the educational expenditure for those enrolled at respective education levels. Since the data on non-participating individuals do not give desired information about those variables, they are not included in the selection equation.

λ_i , the Inverse Mills Ratio, considers the possible selection bias from a censored dependent variable, which causes a concentration of observations at zero value.

If β is statistically zero, it implies no sample selection bias.

Considering the size and heterogeneity of the country, the results are reported with state dummy variables in the result discussions. Details of explanatory variables are presented in Table 1 with summary statistics in Appendix Table 8.

The study uses 75th NSS round data for enrollment and education expenditure from June 2017 to June 2018. It is the latest nationally representing NSS round available for education expenditure drawn through stratified multi-level sampling covering 1,13,757 households and 5,13,366 individuals. This round includes 2,86,456 persons falling in the age group of 3–35 years. Data on household education expenditure at a unit level are obtained for individual level analysis. The study uses information from blocks 2–8, schedule 25.2 on household consumption expenditure, education expenditure, household characteristics such as social groups, religious minorities etc. and individual characteristics like gender. It also uses other variables like gender and educational status of the household head, salary-earning status, etc. The second stratum in this round covers the age group between 3 and 35 years, implying particulars of education-related information for those attending and currently not participating. The education expenditure data include households with zero and positive expenditures. In the data set, zero expenditures encompass two cases—those with nil expenditures due to non-participation and the non-reported cases. In total, 44% of households have positive education expenditure. While

Table 1 Details of the variables under study

Explanatory variables	Description
Log per capita total expenditure	Natural log of Per Capita Total Household Expenditure
Log hhszise	Natural log of the size of household, meaning the number of family members
Female	The variable takes value 1, if the gender is female, otherwise 0; base category = male
Rural	The variable takes value 1, if the place of residence is rural, otherwise 0; base category = urban
Marginalized communities	The variable takes value 1, if the community is SC or ST, otherwise 0; base category = Non-SC-ST
Religious minorities	The variable takes value 1, if it is a non-Hindu, otherwise 0; base category = Hindu
Secondary educated hh head	The variable takes value 1, if it is a secondary educated household head, otherwise 0; base category = below secondary educated
Femhh	The variable takes value 1, if it is a female-headed household, otherwise 0; base category = male-headed household
Salaried	The variable takes value 1, if it is a salaried household, otherwise 0; base category = non-salaried
Distance	The variable takes value 1, if the distance to the nearest school/college is > 5 kms, otherwise 0; base category = distance less than 5 kms
Asset base	This is proxied for a household with computer and internet connections, the variable takes value 1 if it is true, otherwise 0; base category is households without computer and internet connections
Instype	The variable takes value 1, if it is a government educational institute, otherwise 0; base category = private institute

considering those households with secondary school-going children, 55% reported positive education expenditure. Similarly, for those households with children studying at a higher secondary level, 42% reported positive expenditure. Finally, for households with children attending higher education levels, 41% reported positive expenditure. With respect to the distribution of students attending educational institutions across various levels, 91,379 attend education up to the secondary level. In contrast, 28,875 students attend higher secondary schools and 21,140 participate in higher education. At the higher secondary level, 69% choose humanities, while 16 and 15% attend science and commerce, respectively. Meanwhile, at the higher education level 53% attend humanities; whereas, 15, 24 and 8% choose science, commerce and professional courses, respectively.

3.2 Educational spending decision at higher secondary and above levels

For higher secondary and above levels, the choice of specialization of streams matters as much as the spending decision. Our discussions are on the potential students facing various choices of streams at these levels. More specifically at the higher secondary level (classes XI and XII), the options are humanities, science

and commerce. In contrast, the options for higher education level (graduation and post-graduation) are humanities, science, commerce and professional courses. The category, of professional courses encompasses medicine, engineering, agriculture, law, management and chartered account as well as allied courses. IT and computer courses and courses from Industrial Training Institutes are merged with the engineering category while others are clubbed with the management category.

The potential students may choose an outcome that maximizes their utilities and availability of educational institutions for pursuing those courses. For instance, choosing commerce gives a student maximum utility, and the student will opt for that stream, given educational institutions are available in that locality. However, in reality, students may not choose this according to the utilities, on the contrary they may end up choosing a course depending upon an array of social and economic factors. Therefore, the present paper analyses the influence of socio-economic backgrounds like gender, place of residence, social groups and other household characteristics on the choice of streams and subsequent expenditures. Heckman selection model or double-hurdle model as discussed for classes up to secondary level may not be suited for the following analysis as education above secondary levels has polychotomous choices. Hence, in this section, we use the multinomial logit to estimate coefficients for each outcome and the Polychotomous choice model developed by Lee (1983) for selection corrected expenditure. The multinomial logit model captures how these variables influence the probability of an individual student i choosing a particular course j , considering the choice as endogenously determined. This probability may be as follows:

$$P_{ij} = \frac{e^{X_{ij}\gamma_j}}{\sum_j e^{X_{ij}\gamma_j}}, \quad (3)$$

where X_j represents a vector of variables influencing the selection decision with respect to a course at the respective levels of education. γ_j is the vector of coefficients for these variables. Since the choice is determined by various factors, the samples are selected only from the students studying the particular course. Therefore, adjusting the information obtained from Eq. (3) with the expenditure equation is necessary to avoid potential sample selection bias.

At the higher secondary level and above levels, the probability of choosing a course has been estimated using multinomial logit, constitutes the first stage. In the second stage, the estimated expenditure is analyzed for the extent of influence of the socio-economic backgrounds of the households on education spending. The student's expenditure incurred for a certain stream can be represented with a conditional expenditure equation. To obtain the selection corrected expenditure using the Lee correction technique (1983), the transformed expenditure equation can be written as follows:

Table 2 Average household education expenditure at various levels (1987–2018) (in Indian rupees). Source: Authors’ calculation from the respective NSS rounds (1987–2018)

	1987	1996	2008	2014	2018
Secondary	2055	3586 (6.38)	4249 (1.42)	8490 (12.23)	14,738 (14.78)
Higher secondary	3152	4341 (3.62)	5886 (2.57)	13,107 (14.27)	19,322 (10.19)
Graduation	4913	10,854 (9.21)	17,244 (3.93)	23,480 (5.28)	33,758 (9.50)
Post-graduation	11,697	19,288 (5.71)	25,427 (2.33)	32,048 (3.93)	43,985 (8.24)

Expenditure at constant prices is calculated using GDP deflators (2011–12)

1. Figures in parentheses represent the compounded annual growth rate (CAGR in percentage)
2. Expenditures on Diploma/Certificate courses up to secondary are included in secondary level, higher secondary is included in higher secondary and graduation and above are included for the graduation and post-graduate levels of education, instead of treating them separately

$$\begin{aligned}
 lneduexp_{ji} = & \beta_0 + \beta_1 \lambda_i + \beta_2 \ln(\text{Per capita Total Household Expenditure}_i) \\
 & + \beta_3 \ln(\text{Per capita Total Household Expenditure}_i)^2 \\
 & + \beta_4 X_{ji} + \beta_j Y_{ji} + u_{ji},
 \end{aligned} \tag{4}$$

where *lneduexp* is the per capita education expenditure on courses or streams selected *j* for an individual *i* and *u_{ji}* is the error term. *X_{ji}* are the variables capturing household specific and individual characteristics such as social group, religious minorities and gender. *Y_{ji}* is the vector of independent variables influencing the expenditure. This will help to capture the socio-economic inequalities in inter-stream expenses for education. *λ_i* is the Inverse Mills Ratio incorporated as an independent variable in the second stage conditional OLS. If *β* is statistically zero, it implies no sample selection bias. Considering the size and heterogeneity of the country, the results are reported with state dummy variables for the respective results and discussions.

Of the three possible outcomes at the higher secondary level, commerce is the base category. Therefore, at this level, our analysis provides the coefficient of the variables for the other two outcomes—humanities and science and their interpretation will be in comparison with the base category. The dependent variables for Eq. (3) are the log of odds ratio of being in humanities vs commerce;

$$\left[\log \frac{\text{Probability of choosing Humanities}}{\text{Probability of choosing Commerce}} \right] \quad \text{and} \quad \text{science} \quad \text{vs} \quad \text{commerce};$$

$$\left[\log \frac{\text{Probability of choosing Science}}{\text{Probability of choosing Commerce}} \right] \text{ for } i^{\text{th}} \text{ individual.}$$

At the higher education level, we have incorporated a fourth category—professional courses. Similarly, the dependent variables are the log of odds ratio of being in humanities vs science, commerce vs science and professional courses vs science.

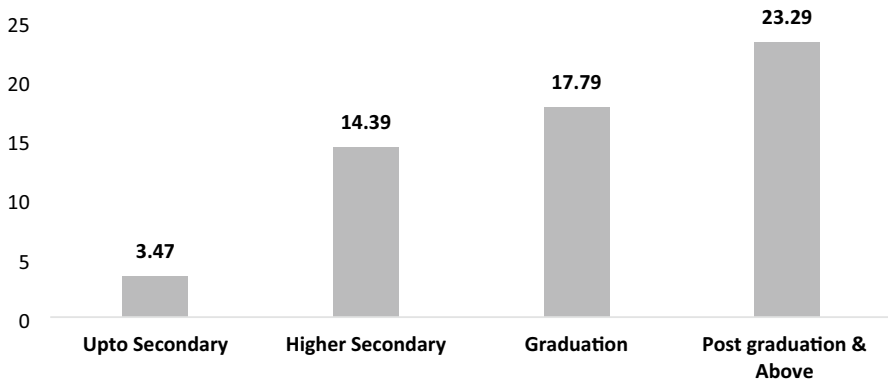


Fig. 1 Educational budget share of the households across various levels of education (in percentage). Source: Authors' calculation from NSS (2018), 75th round

4 Descriptive analysis

The average educational expenditure at constant prices shows an upsurge during 1987–2018 across different levels of education. From Table 2, it is quite evident that households are spending more at graduation and post-graduation levels. Notably, the CAGR of educational expenditure across all levels has shown a marked rise, except for higher secondary levels in 2018. All National Education Policies and programs like District Primary Education Programme (DPEP) (1994), Sarva Shiksha Abhiyan (SSA) (2001), Right to Education Act (RTE) (2010) have ensured improved access to education at the elementary level. RTE emphasizes free and compulsory education for children up to 14 years of age. In line with the above trends with respect to increasing the average expenditure, Tilak (2002) observed that despite government policy interventions at the elementary level, households were still spending a considerable amount on educating their children at this level. He further opined that owing to quality deterioration in state-run schools, parents preferred to send their children to private schools at this level despite high tuition fees. Therefore, according to Tilak (2002), there is no “free education” in India at the elementary level.

Enhanced participation across levels of education reflects the increased demand for education, which is another plausible reason for a hike in the average household expenditure. As per NITI Aayog Statistics, the Net Enrollment Ratio (NER)⁴ at the secondary level improved from 41.9% in 2012 to 53.81% in 2018. Similarly, NER at the higher secondary level improved from 23.73% to 33.58% during the same period. The same argument applies to higher educational levels, which is evident from a rise in the Gross Enrollment Ratio (GER) from 19.4% to 25.01%, with increased enrollments in private universities. After analyzing the movement of

⁴ NER is the number of students of the age of a particular level of education that are enrolled in that level of education, expressed as a percentage of the total population in that age group (Eurostat/Glossary: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Net_enrolment_rate).

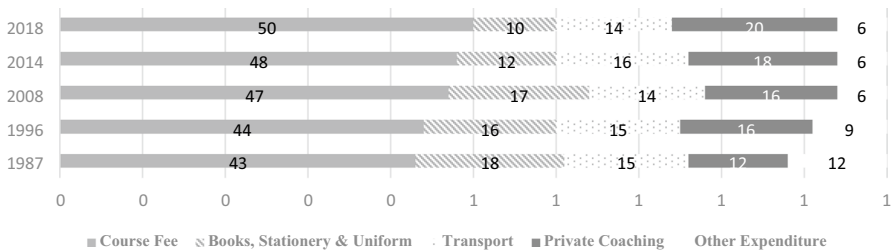


Fig. 2 Component-wise share of household education expenditure (in percentage). Source: Authors' calculation from NSS (2018), 75th round

total educational expenditure, it is necessary to examine the proportion of education spending to the total expenditure of Indian households. It throws some light on the education budget share of the households.

It is more intriguing to examine the proportion of educational expenditure to the total household expenditure than the movement of educational expenditure in absolute terms. Here, the total annual expenditure of households is calculated using the monthly consumption expenditure data of the 75th round, NSS.⁵ Over the years, the educational budget share has increased for the households, given the increased demand for education, returns from education and more extensive availability of private educational institutions. From Fig. 1, it is evident that households keep aside a comparatively minimum share of expenditure for education up to the secondary level. On the contrary, the share of household resources kept aside for education at higher secondary and above levels is considerably high.

In India, the share of public educational expenditure at the elementary and secondary level in the total educational expenditure accounts for 46.72% in 2017–18. The lion's share of public budgetary allocation goes into this level, followed by higher secondary, which might have resulted in this lower proportion of educational spending at the household level. However, the government's provision of education is limited at the higher educational level. The share of government expenditure at this level in the total educational expenditure constitutes a mere 17.12% for 2017–18, as per Ministry of Human Resource and Development (MHRD) statistics, pointing to a lesser allocation at the higher education level. On the contrary, households spend a considerable proportion at this level. Due to the quasi-public nature of higher education and limited public provision at this level, household financing of higher education has been adopted worldwide. Cost-sharing in higher education is widely practised through the imposing of tuition fees. Higher education has the characteristics of rivalry and excludability wherein people tend to internalize the economic surplus (Weiszäcker & Wigger, 1999). Hence, it is crucial to examine the component-wise share of education expenditure.

Tuition fee claims the highest share in educational expenditure by households about 50% in 2018 (Fig. 2). The share of private tutoring also shows a considerable rise over the years whereas the proportion spent on books and stationery has

⁵ $Total\ household\ expenditure = (Monthly\ consumption\ expenditure * 12)$.

Table 3 Proportion of household education expenditure by gender, place of residence (in percentage). Source: Authors' estimation based on NSS (2018), 75th Round

	Rural	Urban	Mean dif- ferences (t-value)	Male	Female	Mean dif- ferences (t-value)	Rural male	Urban male	Mean dif- ferences (t-value)	Rural female	Urban female	Mean dif- ferences (t-value)
Up to secondary	6.17	10.23	1.62	7.09	7.92	2.89	5.74	8.46	11.24*	2.85	9.55	21.47**
Higher secondary	10.12	15.6	2.34*	16.03	11.56	0.68	9.82	12.41	21.4*	6.62	14.71	2.51*
graduation	18.63	20.16	0.86**	18.45	18.34	1.14**	15.18	19.87	4.63**	10.78	18.61	1.82**
Post-graduation & above	19.41	23.61	1.85**	19.67	21.47	2.57**	16.73	20.56	0.064**	12.11	21.93	0.96**

*, ** shows the statistical significance at 1 and 5%, respectively

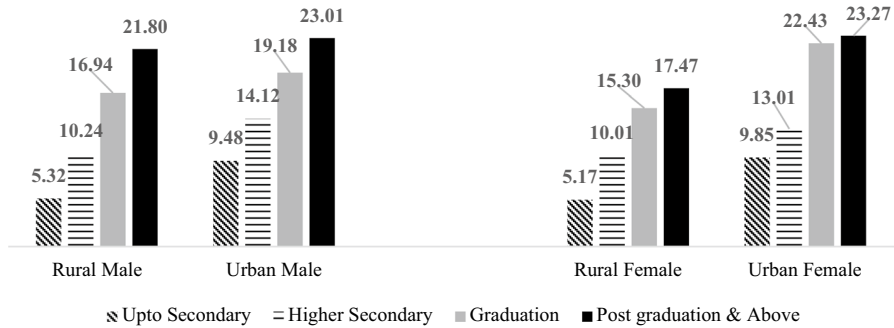


Fig. 3 Proportion of per-student expenditure by gender and place of residence (in percentage). Source: Authors' calculation from NSS (2018), 75th round

reduced. The public policy intervention in terms of providing books and stationery items at the secondary level might have benefited the households.

The presence and availability of private educational institutions are other means of cost-sharing at this level. There is a more prominent presence of private institutions in higher education, accounting for 78% of the total higher educational institutions as of 2018, as per MHRD statistics. Besides, fee structures of private institutions far exceed those of government institutions. The limited government intervention has made households incur a sizable amount on higher education. Therefore, it is necessary to understand the pattern of education budget share by gender, place of residence and social group to get better clarity on the socio-economic inequalities in education spending.

From Table 3, it is evident that the proportion of educational expenditure is more for urban than rural areas across levels of education. It is certainly intriguing to note that a higher proportion is spent on females at all levels of education, excepting a marginal decrease at the higher secondary level. This calls for a further disaggregated level analysis and investigation of per-student expenditure.

It is observed that in rural areas, the average size of households is larger as compared to urban areas. Owing to the bigger family size, the proportion of household educational expenditure to the total household expenditure may be misleading. Thus, the proportion of per-student educational expenditure to the average household expenditure gives an accurate picture of the burden on families. In short, rural families need to provide for a greater number of children, while urban nuclear families spend exclusively on one or two children. Here, the mean household size in rural areas is nine, while it is only four in urban areas. Hence, the per-student expenditure and its proportion are even higher for urban households, and further analysis, thus, proceeded with per-student expenditure.

The share of per-student expenditure⁶ is considerably higher for urban females, and excepting the secondary level, it is even higher than for urban males (Fig. 3).

⁶ Per-student expenditure is obtained from the unit level data and calculated as $\frac{\text{Expenditure on education for respective levels}}{\text{Total number of individuals in the household who attending that particular level of education}}$

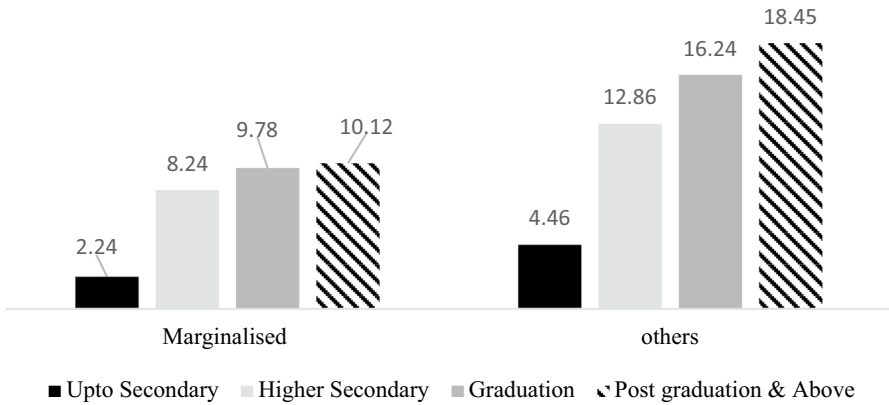


Fig. 4 Proportion of per-student expenditure by social groups (in percentage). Source: Authors’ calculation from NSS (2018), 75th round

Table 4 Stream-wise average annual per capita education expenditure at various levels (in Indian rupee). Source: Authors’ calculation from NSS (2018), 75th round

Streams/Courses	Higher secondary	Above higher secondary/Higher education
Humanities	3864	2721
Commerce	6763	5831
Science	8695	11,661
Professional courses	–	18,658

Rural females find themselves discriminated against when it comes to educational spending at different levels. A wider gap in educational expenditure is observed for rural females at graduation and post-graduation levels compared to secondary and higher secondary levels. The poor economic status of rural households, coupled with a limited provision of higher education by the government, prevents them from enrolling their children at these levels.

An analysis of the share of per-student expenditure by social groups unveils an insightful picture (Fig. 4). In India, economic status is strongly associated with the social status of households, and it also becomes evident from the below-given figure. The mainstream communities in India keep aside a more significant proportion of their expenditure on their children’s education. This calls for required policy initiatives to uplift marginalized communities and religious minorities, especially at higher levels of education. All major National Education Policies have given a greater thrust to the upliftment and inclusion of marginalized social communities in the country. The poor economic condition of households further prevents them from accessing education, significantly higher education. Therefore, sustained efforts are required on the part of the government to uplift the socially and economically backward marginal groups.

A detailed descriptive analysis on the dynamics of household spending for education by gender, place of residence and social groups unveiled that the average household educational expenditure has observed a leap. At the same time, socio-economic status plays a vital role in determining its level of spending, especially at higher levels.

From Table 4, the average yearly per capita spending is higher for science than for humanities and commerce at the higher secondary level. Education above higher secondary professional courses is expensive as households incur more than others. At both levels of education, humanities courses are less expensive than others. Hence, it may be interesting to connect the socio-economic background of the families with the choice of subject specialization. From the above descriptive analysis, the presence of socio-economic inequalities while allocating educational resources is evident. The following empirical analysis aids to get better clarity on the channels of inequalities across three different levels of education.

5 Results of econometric estimation and discussions

5.1 Decision-making for education spending up to secondary level

The following analysis compares the traditional Engel curve approach using OLS with the Heckman selection model in education spending decision-making up to the secondary level.

Table 5 presents the estimated coefficients of the Heckman two-step selection model for analyzing the education expenditure up to the secondary level. Column 3 shows the estimation of the traditional Engel curve, using simple OLS, while columns 1 and 2 report the selection and outcome equations of the Heckman selection model, respectively. Given, the limitation of single OLS in capturing the two-stage decision-making, the marginal effects calculated at mean values of the independent variables from the selection equation of the Heckman selection model are deemed more reliable to understanding the socio-economic factors determining school participation.

From the selection equation, the per capita expenditure and its square have a positive sign despite being statistically insignificant. The household expenditure is a proxy for household income, underscoring the economic status. The result implies that household income is not a vital factor influencing the enrollment decision at this level of education. With educational policy interventions like Sarva Shiksha Abhiyan (SSA) (2001) and the Right to Education Act (RTE) (2009) providing free and compulsory education, the economic status of the households might have become irrelevant for the school participation. However, the size of the families plays a vital role in the participation decision, pinpointing that the larger families are less likely to send their children to school at the secondary level. Aslam and Kingdon (2008) also observed a similar trend in their study. They justified it based on parents' preference for the male child increases the family size, resulting in higher family size. Hence, the girl students may have more siblings, impacting the household enrollment decision.

Table 5 Determinants of household educational expenditure up to secondary level. Source: Author's estimation

Variables	Heckman selection model		OLS (3)
	Selection equation (probit- first stage) (1)	Outcome equation (conditional OLS) (2)	
Natural log of Per capita Total Expenditure	0.091 (0.067)	-0.950*** (0.010)	-0.314** (0.167)
Natural log of Per capita Total Expenditure (Square)	0.008 (0.005)	0.041 *** (0.044)	0.04*** (0.008)
Natural log of household size	-0.006*** (0.002)	0.623* (0.041)	0.544 (0.014)
Female	-0.018 (0.001)	-0.121*** (0.001)	-0.845* (0.015)
Rural	-0.008*** (0.002)	-0.487*** (0.014)	-0.44*** (0.01)
Marginalized	-0.022*** (0.002)	-0.284*** (0.019)	-0.305*** (0.01)
Religious minorities	-0.010*** (0.004)	-0.637*** (0.005)	-0.151*** (0.011)
Secondary educated household head	0.049*** (0.004)	0.317 (0.015)	0.398*** (0.011)
Female-headed households	-0.017*** (0.007)	-0.013*** (0.022)	-0.07** (0.015)
Salaried households	0.023*** (0.003)	-0.028 (0.018)	-0.06*** (0.013)
Distance		0.078*** (0.020)	-0.1*** (0.016)
Government institutions		-0.386** (0.604)	-0.184*** (0.182)
Households having computers with internet connections	0.028*** (0.042)		
State dummy	Yes	Yes	Yes
Constant		7.581 *** (1.560)	1.112*** (0.082)
Adjusted R square			0.284
Mills Lambda	-1.191 ***		
Rho	-0.811		
Sigma	1.469		
Wald Chi Square	11,564.3		

Table 5 (continued)

Variables	Heckman selection model		OLS (3)
	Selection equation (probit- first stage) (1)	Outcome equation (conditional OLS) (2)	
Total number of observations	166,145		
Censored	74,765		
Prob> Chi Square	0.00		

Standard errors are given in parenthesis

*, **, *** show the statistical significance at 10, 5 and 1% levels, respectively

Dependent variable for the selection equation is a binary variable capturing the school participation, while natural log of share of education expenditure of a household for outcome equation

The coefficients of Selection and Participation Equations are marginal effects

Social inequalities, captured by the rural–urban divide, marginalized communities, religious minorities are more revealing here. Notably, children from rural pockets are deprived of schooling choices. Similar is the case with children from marginalized communities and religious minorities, underscoring the gap in existing policy interventions. The marginal effect of females is negative and non-significant; thus, gender is not a crucial factor for school enrollment. Specific education policies like the National Programme for Education of Girls at Elementary Level NPEGL (2003) and Kasturba Gandhi Balika Vidhyalaya KGBV (2004) for enhancing girl students' participation as well as providing residential facilities for girl students might have augured well in increasing their enrollment.

The household head's education level is an important factor determining the school enrollments; however, a female-headed household has a lesser chance of sending children to schooling. Meanwhile, the sign of the coefficient for a regular salary-earning household is positive and statistically significant, implying that the pattern of occupation matters in the school enrollment decision. The asset base of the household, captured through the households having computers with internet connections, turned out to be positive and significant, implying that it strongly influences enrollment decisions at the secondary level.

From the coefficients of conditional OLS, the coefficient of log of per capita household expenditure is negative and highly significant. At the same time, the squared term is positive and significant at 1% level. It implies that, as per capita income increases, the education budget share decreases after reaching a threshold level. As income increases, households do not set apart as much as a higher proportion for education expenditure at this level, contradicting Engel's law. The educational policies of free and compulsory education or subsidized provision might have aided in shrinking their education budget shares at this level. Moreover, rural and economically backward households may not have substantially higher education budget shares even with increased income. With bigger families, the conditional budget share may be high for those enrolled in schooling. Children from rural areas, marginalized communities and religious minorities are deprived of education spending, which is evident from the magnitude and direction of their coefficients. The social inequalities are pronounced in the choice of schooling and spending on education. Despite decades of policy interventions aimed at uplifting the socially deprived in educational attainments, they are yet to catch up with economically advantageous groups owing to the disparities in resource allocation. Although enrollment decision is gender-neutral at this level, the estimated amount incurred on female students is less, showing gender inequality at the second-stage decision-making. Therefore, at this level, the main channel of gender inequality is observed not in the enrollment decision, but rather in the expenditure decision, corroborating the results of Datta and Kingdon (2021). While the targeted policy interventions aided in reducing the gender gap in enrollments, the wide variations in household spending on education limit the scope of such policy interventions.

Although the education level of the household heads influences the enrollment decision, it does not matter for the level of spending. Female-headed households show a particular bias in enrollment as well as spending decisions. The estimated expenditure of salaried households turns out to be less although statistically

insignificant. As the distance increases education expenditure of the households increases considering the transportation cost families need to bear. Government institutions are low cost compared to private given their lower tuition fees. At the school level, enrollment in private institutions is higher in India. Nevertheless, those who are enrolled in government-run institutions spend a lesser amount on education. Despite higher tuition fees, for those who attend private institutions, households spend considerably high and several shortcomings of the government-run institutions might be the reason behind their preference for private institutions (Sengupta, 2020). The significant coefficient of Mills Lambda implies that the selection equation is relevant.

5.2 Decision-making for education at higher secondary level

As discussed above, the potential channels of socio-economic inequalities are different for education above secondary levels. The choice of various streams or courses becomes relevant at the above secondary level. Thus, the following section deals with the socio-economic determinants of choosing different courses and their related expenditure. Here, choosing expensive or superior streams that are job-oriented is often the potential source of socio-economic inequalities and subsequent resource allocation.

The results of the multinomial logit model estimating the probabilities of choosing humanities (column 1) and science (column 2), keeping commerce as the base outcome, are presented in Table 6. In some instances, the sign of marginal effects is distinct from the coefficients, as coefficients of this estimation represent the changes in the probability of one outcome vis-à-vis the probability of the base outcome. On the other hand, marginal effects indicate the probability of one particular outcome (Bairagya, 2018). Hence, there can be different signs for coefficients and marginal effects. The Inverse Mills Ratio (IMR) has been calculated from the predicted probabilities. IMR is included as an independent variable as the selection correction term in the functional form of the Engel curve, thus arriving at the selection corrected expenditure for the respective courses.

From the multinomial logit estimates, with an increase in household expenditure, a proxy of household income, there is a more likelihood of joining science and less probability of joining humanities than commerce. The estimates of selection corrected expenditure (columns 3 and 4) imply that as per capita household expenditure increases, the education expenditure increases at a decreasing rate for science. In contrast, for humanities, the education expenditure reduces with increased per capita household expenditure. As family size increases, humanities are opted for, with science less preferred over commerce. Humanities are less expensive in tuition fees than commerce or science could be the plausible reason behind it. Commerce and Science streams are generally the most sought after as there is a greater demand for them in the job market. With bigger families, the expenditure on science is higher than for other courses. Choice of streams is a clear source of inequality for those

from rural areas, marginalized communities and religious minorities as the probability of joining humanities scores over commerce and science. There exists a clear disparity in the spending decision for such students. There is a rural–urban divide in education spending at this level. Private tuition and coaching are higher in urban areas, especially for science students, marking a wide disparity between rural and urban expenditure, reflecting learning outcomes (Agrawal, 2014).

The statistically non-significant coefficients for female enrollments (columns 1 and 2) highlight that, at the higher secondary level, gender inequality is muted, implying gender does not matter in the choice of the streams. However, female science students are discriminated against for the amount incurred for education at the higher secondary level. Interestingly, the educational expenses for those females enrolled in humanities are more than for males. It may be because of the enhanced participation of urban female students as observed from the initial descriptive analysis. The education level of the household heads does not influence the choice of courses at the higher secondary level. Nevertheless, education expenditure increases with educated household heads. The probability of choosing humanities is slightly high for female-headed households, while they spend less for those who get enrolled. The wage-earning status of the family does not matter for the choice of courses, while the education spending of the salaried households is more for science courses. Notably, education expenditure is more for those enrolled in the science stream with an increase in distance than others. It may imply the less availability of government institutions that may increase the families' transportation costs. The estimated expenditure is less for government institutions than private. The dearer tuition component makes households spend considerably higher if the child is in a private institute pursuing a higher secondary-level course.

5.3 Decision-making for education spending at higher education level

In the following section, the socio-economic factors influencing the choice of streams and the resultant expenditure on the education at higher education levels are discussed. The courses considered for the analysis are humanities, commerce and professional, while keeping science as the base category.

In Table 7, coefficients are at marginal effects based on multinomial logit estimation at higher education level (columns 1, 2 and 3). Columns 4, 5 and 6 show the selection corrected expenditure for the respective courses, keeping science as the base category.

As the income of the households increases, professional courses are preferred over other courses, which is evident from the sign of estimated coefficients of per capita household expenditure and its squared term using the multinomial logit model. According to the selection corrected expenditure estimates, education expenditure for commerce and professional courses increases with an increase in household income until it reaches a threshold limit, whereas it is vice versa for humanities.

Table 6 Socio-economic determinants of household education expenditure at the higher secondary level. Source: Authors' Estimation

Variables	Humanities (Mlogit) (1)	Science (Mlogit) (2)	Selection Corrected Expenditure—Humanities (3)	Selection Corrected Expenditure—Science (4)
Natural log of Per capita Total Expenditure	-0.911*** (0.201)	0.997*** (0.181)	-0.951** (0.681)	0.840** (0.357)
Natural log of Per capita Total Expenditure (Square)	-0.062*** (0.01)	0.048*** (0.012)	0.061*** (0.028)	-0.060*** (0.021)
Natural log of household size	0.032* (0.010)	-0.013 (0.020)	-0.381*** (0.03)	0.424*** (0.038)
Female	0.008 (0.019)	-0.024 (0.018)	0.090*** (0.041)	-0.081*** (0.004)
Rural	0.077** (0.010)	-0.002*** (0.01)	-0.142*** (0.022)	-0.156*** (0.037)
Marginalized	0.047*** (0.011)	-0.021*** (0.010)	-0.138*** (0.043)	-0.119*** (0.014)
Religious minorities	0.040*** (0.018)	-0.034*** (0.018)	-0.045*** (0.022)	-0.022 (0.028)
Secondary Educated Household Head	0.089 (0.010)	-0.100 (0.001)	0.142*** (0.019)	0.132*** (0.021)
Female-headed Households	0.019* (0.001)	-0.029* (0.033)	-0.081*** (0.002)	-0.072*** (0.051)
Salariated Households	0.008 (0.019)	0.010 (0.011)	-0.020** (0.030)	0.006** (0.042)
Distance			-0.001*** (0.042)	0.097*** (0.051)
Government Institutions			-0.339*** (0.002)	-0.304*** (0.028)
Households having computers with internet connections	0.171*** (0.055)	0.012 (0.042)		
Constant	Yes	Yes	2.233*** (2.674)	0.936*** (2.297)
State Dummy Variables			Yes	Yes
Loglikelihood	-7758.83			
LR Chi square	4081.28		2197.03	2018.52
Wald Chi square				
Pseudo R square	0.211			
P > Chi square	0.00		0.00	0.00
Adjusted R square			0.179	0.176

Table 6 (continued)

Variables	Humanities (Mlogit) (1)	Science (Mlogit) (2)	Selection Corrected Expenditure—Humanities (3)	Selection Corrected Expenditure—Science (4)
lambda1 (Selection correction term)			0.079***	
lambda2 (Selection correction term)				0.051***
Number of observations	28,875			

Commerce is considered as the base outcome; Standard errors are presented in parenthesis; Chi2 is the test statistic with Ho: coefficients are jointly equal to zero. The coefficients are marginal effects
*, **, *** show the statistical significance at 10, 5 and 1% levels

In the multinomial logit model, dependent variables are the log of odds ratio of being in Humanities vs. Commerce and Science vs Commerce whereas, natural log of per capita education expenditure for selection corrected expenditures. For Columns 1 and 2, coefficients are marginal effects

It is quite evident that with an increase in household income, more expensive and job-oriented courses are preferred by households. However, the education expenditure does not increase as much as the income hike, confirming Engel's law. At the higher education level, bigger families prefer humanities and commerce over science, while professional courses are not desired. Regarding tuition fees, professional courses are more expensive than other streams. Moreover, given the limited public provision at the higher education level, especially professional courses, reliance on private educational institutions becomes necessary. Hence, with bigger families, affordability becomes a concern and thus, opting for less expensive courses.

There is clear evidence of socio-economic inequalities in the choice of specialization and related expenditure for students from rural areas, marginalized communities and religious minorities at the higher education level. The dearth of policy initiatives on higher education considering the social inequalities may also aggravate the existing educational gap. Unlike at secondary and higher secondary levels, for higher education, gender inequality is more pronounced both for the choice of courses and expenditure incurred. From the multinomial logit estimates, females prefer humanities and commerce over science and professional courses. The education level of the household head matters only for the enrollment in professional education and increases the education expenditure as well. The salary-earning status of the household does not matter for enrollment in traditional courses. However, it increases the probability of getting enrolled in professional courses. Nevertheless, they spend considerably high compared to others. Enrollment in traditional courses disregards the family head's gender while it matters to choosing professional courses. However, female-headed households spend a lesser amount, especially on professional courses. The proximity of educational institutions at this level influences the amount spent. The spending turns out to be at the higher end of the institutions that are far away. Given the less expensive nature of such institutions, government institutions are preferred for education at this level. Especially for professional education, the tuition fee is subsidized at government institutions and hefty at private institutions. The enrollment rate at higher education is already low for the country as against other well-performing countries. The persisting socio-economic discrimination is the primary reason for falling behind the curve.

6 Conclusion

Using the 75th NSS round (2018), the study extensively focused on the potential channels of social inequalities in allocating educational resources by households. From the descriptive analysis, during 1987–2018, the educational budget share has increased for the households, given the increased demand for education, returns from education and more extensive availability of private educational institutions. The share of household resources kept aside for education at higher secondary and

above levels is considerably high. Tuition fee claims the highest share of educational expenditure by households. The share of private tutoring also shows a considerable rise over the years, whereas the proportion spent on books and stationery has reduced. Although the timeframe of the data for the current study does not account for the Covid-19 pandemic period, it is evident that the onslaught of the unprecedented pandemic might have altered the structure of household spending on education. The paradigm changes and associated educational inequalities that surfaced during the pandemic era is yet to be accounted for with greater clarity (Al-Samarrai et al., 2020).

Under the Engel curve framework, Heckman's two-step model's coefficients reveal the extent of social inequalities in resource allocation up to the secondary level of education. The social inequalities in participation choice and spending decision are evident at this level. The study highlights the group-based inequalities in enrollment decisions and subsequent education expenditure by rural–urban divide, social groups and religious minorities. The persisting social gap in resource allocation may further widen the education attainment gap. There is a need to revisit the exiting policy interventions to fill the gap in enrollment choice and subsequent spending. Notably, the economic status of the households claims no importance for the participation decision and expenditure at this level backed by the subsidized provision of education. As per capita income increases, the education budget share decreases after reaching a threshold level. At this level, as income increases, households do not set apart as much as the higher proportion for education expenditure, contradicting Engel's law. Policies such as SSA (2001) and RTE (2009) providing free and compulsory education may have enhanced the enrollments barring the income level of the families.

At this level, enrollment choice is not affected by the gender considerations of the households instead it comes into the picture through the amount spent for the girl child. There was a preference for private schools at the secondary level, especially in urban areas with a pro-male bias. The elevated household education expenditure for male students at this level may be due to increased private school enrollment. The inclination towards private schools is contested in light of debates on superior quality and infrastructural amenities (Singh & Sridhar, 2002). (Refer Appendix Table 9). The proximity and type of educational institution influence educational spending. Although the education level of the household heads influences the enrollment decision, it does not matter for the level of spending. Female-headed households show a particular bias in enrollment as well as spending decisions. The estimated expenditure of salaried households turns out to be less although statistically insignificant for education up to the secondary level.

The main channel of social inequalities is the choice of specialization of subjects at the higher secondary and above levels. Students from rural areas, marginalized and deprived, choose less expensive courses than others. The subsequent expenses are also less than socially and economically better off students. Affordability and the hassle of clearing competitive entrance exams trigger less expensive preferences among those students. With an increase in household expenditure there is a more likelihood of joining science and less probability of joining humanities than commerce. As per capita household expenditure increases, the education expenditure increases at a decreasing

Table 7 Determinants of household education expenditure at higher education level. Source: Authors' estimation

Variables	Humanities (1)	Commerce (2)	Professional Course (3)	Selection Corrected Expenditure—Humanities (4)	Selection Corrected Expenditure—Commerce (5)	Selection Corrected Expenditure—Professional Courses (6)
Natural log of Per capita Total Expenditure		-0.001* (0.005)	-0.004* (0.003)	0.004* (0.003)	-3.676*** (0.609)	3.403* (0.049)
Natural log of Per capita Total Expenditure (Square)		-0.006* (0.044)	-0.033*** (0.002)	0.003* (0.002)	0.184*** (0.033)	-0.322* (0.005)
Natural log of household size		0.05*** (0.036)	0.036** (0.031)	-0.046*** (0.036)	-0.515*** (0.023)	-0.338*** (0.017)
Female		0.003*** (0.027)	0.031** (0.252)	-0.058** (0.01)	-5.202*** (0.632)	-6.221** (0.018)
Rural		0.005*** (0.015)	0.021*** (0.002)	-0.096** (0.007)	-0.294*** (0.02)	-0.176** (0.005)
Marginalized		0.003*** (0.025)	0.015*** (0.001)	-0.026*** (0.025)	-0.024 (0.03)	-0.322*** (0.006)
Religious minorities		0.069*** (0.046)	0.092*** (0.004)	-0.006** (0.021)	-0.178*** (0.029)	-0.112** (0.078)
Secondary educated household head		-0.018 (0.017)	-0.095 (0.007)	0.012*** (0.013)	0.009 (0.017)	0.413*** (0.033)
Female-headed households		0.002 (0.027)	0.001 (0.252)	-0.005*** (0.016)	-5.231 (0.629)	-0.516*** (0.001)
Salaried households		0.059 (0.021)	0.002 (0.006)	0.072*** (0.004)	0.008* (0.03)	0.414*** (0.028)
Distance					0.102*** (0.037)	0.168*** (0.004)
Government institutions					-0.110*** (0.083)	-0.511** (0.001)
Households having computers with internet connections		0.002*** (0.012)	0.014*** (0.002)	0.061*** (0.020)		
Constant				-2.856*(2.959)	-1.867*** (0.004)	-3.152*** (0.041)
State Dummy variables	Yes	Yes	Yes	Yes	Yes	Yes
Loglikelihood						
LR Chi square						

Table 7 (continued)

Variables	Humanities (1)	Commerce (2)	Profes- sional Course (3)	Selection Corrected Expenditure— Humanities (4)	Selection Corrected Expenditure—Com- merce (5)	Selection Corrected Expendi- ture—Professional Courses (6)
Wald Chi square			2847.06		1948.26	1582.58
Pseudo R square	0.714					
P > Chi square	0.00		0.00	0.00	0.00	0.00
Adjusted R square			0.311	0.114		0.125
lambda1 (Selection correction term)			-1.142***			
lambda2 (Selection correction term)				-1.606***		
lambda4 (Selection correction term)						-1.952***
Number of observa- tions	21,140					

Science is considered as the base outcome; Standard errors are presented in parenthesis; Chi2 is the test statistic with Ho: coefficients are jointly equal to zero. Coefficients are marginal effects

*, **, *** show statistical significance at 10, 5 and 1% levels, respectively

The dependent variables for multinomial logit model is the log of odds ratio of being in Humanities vs Science, Commerce vs Science and Professional courses vs Science whereas, natural log of per capita education expenditure for a particular course. For columns 1, 2 and 3 coefficients are marginal effects

rate for science. In contrast, for humanities, the education expenditure reduces with increased per capita household expenditure. It may be because, mainly for the economically deprived families humanities is a feasible option.

As family size increases, humanities are opted for, with science less preferred over commerce. Humanities are less expensive in terms of tuition fees than commerce or science could be the plausible reason behind it. Commerce and Science streams are generally the most sought after as there is a greater demand for them in the job market. With bigger families, the expenditure on science is higher than for other courses. At the higher secondary level, gender does not matter for the enrollment in any of the streams. However, families spend less on female science students at the higher secondary level. Interestingly, the educational expenses for those females enrolled in humanities are more than for males. It may be because of the enhanced participation of urban female students.

Females in higher education prefer humanities and commerce over science and professional courses. Professional courses are preferred to the science stream when there is an increase in household income. At the higher education level, bigger families prefer humanities and commerce over science, while professional courses are not desired. Regarding tuition fees, professional courses are more expensive than other streams. Moreover, given the limited public provision at the higher education level, especially professional courses, the reliance on private educational institutions becomes necessary. Hence, with bigger families, affordability becomes a concern, thus opting for less expensive courses. At higher secondary and above levels, an important reason for an inflated household educational expenditure could be the availability of private tuition, as pointed out by some studies (Kim & Park, 2010; Tansel & Bircan, 2006).

Therefore, the channels of socio-economic inequalities in educational spending at each level are different. The social gaps in educational participation choice and spending decisions are evident up to the secondary level. At the same time, it is visible in the choice of specialization of the courses and subsequent expenditure at higher secondary and above levels. The socio-economic disparities in resource allocation underscore the necessity of policy initiatives specifically removing spending differences. It gives a clarion call for re-orienting existing policy initiatives to improve participation and make it egalitarian. At the higher education level, the choice of streams is highly selective and favors those with the capacity and willingness to pay. Given its quasi-public nature, higher education continues to get less importance in educational policy interventions and associated spending. The growth spill-over of higher education and its direct link to the job market have been debated constantly. Given the persisting social inequality in educational resource allocation at the higher education level, any lacuna in policy interventions minimizing the gap may aggravate the educational attainment gap as well.

Appendix

See Tables 8 and 9.

Table 8 Summary Statistics—Mean

Variables	Up to secondary	Higher secondary	Above higher secondary
Per capita Education Expenditure as whole	8490	13,107	28,264
Per capita Education Expenditure—Humanities		15,607	32,961
Per capita Education Expenditure—Commerce		18,742	38,928
Per capita Education Expenditure—Science		22,450	49,580
Per capita Education Expenditure—Professional Courses			69,274
Proportion of students enrolled for Humanities		64	68
Proportion of students enrolled for Commerce		21	12
Proportion of students enrolled for Science		15	11
Proportion of students enrolled for Professional Courses			9
Per student education budget share for school-going children	8.4		
HH size	5.46	5.21	3.06
Proportion of female students	56	45	38
Proportion of rural students	52	47	29
Proportion of Marginalized communities	34	26	18
Proportion of Religious minorities	52	43	35
Proportion of secondary educated household head	42	39	28
Distance from educational institutions more than 5 km	0.78	2.51	3.16
Share of female-headed household	25	18	13
Proportion of households with computer and internet connection (Asset base)	12	17	21
Share of gov't institutions chosen by those who are attending educational institutions	86	78	53
Proportion of salaried household	35	41	48

Table 9 Proportion of enrolled students attending private educational institutions (in percentage). Source: Authors' calculation

	1987		1996		2008		2014		2018	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Up to secondary	10	8	18	13	21	17	29	25	35	30
Higher secondary	8	4	11	10	20	14	27	21	31	25
Above higher secondary	11	9	23	14	42	34	51	44	58	48

Acknowledgements We thank the anonymous reviewers for the fruitful comments and remarks to improve the initial version of the manuscript

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by ARV and IB. The first draft of the manuscript was written by ARV and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding We declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest We combinedly declare that no conflicting interest (financial or non- financial) are directly or indirectly related to the work submitted for publication. We have no relevant financial or non-financial interests to disclose.

Ethical approval We combinedly declare that all ethical standards of the journal are met while preparing the manuscript.

Consent to participate and publish As per the requirement of the journal, we combinedly give the consent for the same.

References

- Acar, E. Ö., Günalp, B., & Cilasun, S. M. (2016). An empirical analysis of household education expenditures in Turkey. *International Journal of Educational Development*, *51*, 23–35. <https://doi.org/10.1016/j.ijedudev.2016.03.007>
- Agrawal, T. (2014). Educational inequality in rural and urban India. *International Journal of Educational Development*, *34*, 11–19.
- Ahmadzai, H. (2018). *Tests for separability and factor market participation in Afghanistan*. CREDIT Research Paper.
- Al-Samarrai, S., Gangwar, M., & Gala, P. (2020). The Impact of the COVID-19 Pandemic on Education Financing. World Bank Other Operational Studies 33739, the World Bank.
- Asadullah, M. N., & Chaudhury, N. (2009). Reverse gender gap in schooling in Bangladesh: insights from urban and rural households. *Journal of Development Studies*, *45*(8), 1360–1380.
- Aslam, M., & Kingdon, G. G. (2008). Gender and household education expenditure in Pakistan. *Applied Economics*, *40*(20), 2573–2591. <https://doi.org/10.1080/00036840600970252>
- Azam, M., & Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra—household allocation of education expenditure. *World Development*, *42*, 143–164.
- Bairagya, I. (2018). Why is unemployment higher among the educated. *Economic & Political Weekly*, *53*(7), 43–51.
- Bohindar, S. (2019). Corporate route of higher education. *Frontline*, *36*(14), 8–11.
- Castelló-Climent, A., & Mukhopadhyay, A. (2013). Mass education or a minority well educated elite in the process of growth: The case of India. *Journal of Development Economics*, *105*, 303–320. <https://doi.org/10.1016/j.jdeveco.2013.03.012>
- Chaudhuri, K., & Roy, S. (2006). Do parents spread educational expenditure evenly across the two genders? Evidence from two North Indian states. *Economic and Political Weekly*, 5276–5282.
- Chiplunkar, G., & Weaver, J. (2021). Marriage markets and the rise of dowry in India. Available at SSRN 3590730.

- Das, U., & Sarkhel, P. (2020). Does more schooling imply improved learning? Evidence from a conditional cash transfer programme in India (SSRN Scholarly Paper ID 3703566). *Social Science Research Network*. <https://doi.org/10.2139/ssrn.3703566>
- Datta, S., & Kingdon, G. G. (2019b). *Gender Bias in Intra - Household Allocation of Education in India: Has It Fallen over Time?* IZA Discussion Papers.
- Datta, S., & Kingdon, G. G. (2021). Gender Bias in Intra - Household Allocation of Education in India: Has it fallen over time? In *DoQSS Working Papers* (No. 21–06; DoQSS Working Papers). Quantitative Social Science - UCL Social Research Institute, University College London. <https://ideas.repec.org/p/qss/dqsswp/2106.html>
- Deaton, A. (1989). Looking for boy-girl discrimination in household expenditure data. *The World Bank Economic Review*, 3(1), 1–15.
- Hjortsberg, C. (2003). Why do the sick not utilise health care? *The Case of Zambia. Health Economics*, 12(9), 755–770.
- Humphreys, B. R. (2013). Dealing with zeros in economic data. *University of Alberta Department of Economics*, 1, 1–27.
- Johnson, N. L., Kovarik, C., Meinzen-Dick, R., Njuki, J., & Quisumbing, A. (2016). Gender, assets, and agricultural development: Lessons from eight projects. *World Development*, 83, 295–311.
- Kim, J.-H., & Park, D. (2010). The determinants of demand for private tutoring in South Korea. *Asia Pacific Education Review*, 11(3), 411–421.
- Kingdon, G. G. (2005). Where has all the bias gone? Detecting gender bias in the intra—household allocation of educational expenditure. *Economic Development and Cultural Change*, 53(2), 409–451.
- Lancaster, G., Maitra, P., & Ray, R. (2008). Household expenditure patterns and gender bias: evidence from selected Indian states. *Oxford Development Studies*, 36(2), 133–157.
- Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica: Journal of the Econometric Society*, 51, 507–512.
- Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *Journal of Health Economics*, 27(2), 300–307.
- Maertens, A. (2013). Social norms and aspirations: age of marriage and education in rural India. *World Development*, 47, 1–15.
- Majumder, A., & Mitra, C. (2017). Gender bias in household education expenditure: the case of West Bengal. *Indian Growth and Development Review*.
- Sachar Commission, R. of. (2015, August 8). *Sachar Committee Report | Ministry of Minority Affairs | Government of India*. <https://minorityaffairs.gov.in/reports/sachar-committee-report>
- Saha, A. (2013). An assessment of gender discrimination in household expenditure on education in India. *Oxford Development Studies*, 41(2), 220–238.
- Sekher, T., & Ram, F. (2015). Conditional cash transfers for girls in India: Assessment of a girl child promotion scheme from beneficiary perspective. *International Institute for Population Sciences*.
- Sengupta, A. (2020). Rapid growth of private universities. *Economic and Political Weekly*, 55(22), 7–8.
- Singh, S., & Sridhar, K. S. (2002). Government and private schools: Trends in enrolment and retention. *Economic and Political Weekly*, 4229–4238.
- Tansel, A., & Bircan, F. (2006). Demand for education in Turkey: a tobit analysis of private tutoring expenditures. *Economics of Education Review*, 25(3), 303–313. <https://doi.org/10.1016/j.econedurev.2005.02.003>
- Tilak, J. B. G. (2002). Elasticity of household expenditure on education in rural India. *South Asia Economic Journal*, 3(2), 217–226. <https://doi.org/10.1177/139156140200300206>
- von Weizsäcker, R. K., & Wigger, B. U. (1999). Risk, resources, and education: public versus private financing of higher education. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.880829>
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT press.
- You, X., & Kobayashi, Y. (2011). Determinants of out-of-pocket health expenditure in China. *Applied Health Economics and Health Policy*, 9(1), 39–49.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.