(Preliminary Draft)

Gender differential in educational attainments in India
Comparing results of individual-level and household-level analysis

Zakir Husain
Associate Professor
Department of Humanities and Social Sciences,
Indian Institute of Technology, Kharagpur, India
Email: dzhusain@gmail.com

Charanjit Kaur
PhD Scholar
Department of Statistics
University of Delhi
E-mail: charan.rupal@gmail.com

Mousumi Dutta
Associate Professor
Department of Economics
Presidency University, Kolkata, India
Email: dmousumi1970@gmail.com

Gender differential in educational attainments in India Comparing results of individual-level and household-level analysis

Abstract

Studies on the gender gap in enrolment and on educational attainment in India based on analysis of individual outcomes indicate a clear bias against girls. However, another line of research that attempts to test the presence of gender discrimination in education by comparing the level of *household* expenditure on education has been less successful. In this study, we re-examine this problem using data from National Sample Survey Organization (NSSO) 64th round data (2009) on "Participation and Expenditure in Education". We first start by testing for gender differences in educational expenditure at the individual level; in the second step we shift to the household level and test for the persistence of our results on gender differentials. Our study is an improvement over earlier studies as we test for discrimination at two levels *using the same data set*.

JEL classification: I24, J16, C24

Keywords: Expenditure, Gender Discrimination, Resource allocation, India.

Highlights:

- Studies reveal the presence of gender disparities across individuals.
- Such disparities are not observed when we shift to house-level analysis.
- Recent studies have shifted to more complex econometric models.
- Another problem may be the failure of surveys to capture expenditure on education.
- We have used the same data set.
- We find gender disparity at both individual and household level.

Gender differential in educational attainments in India

Comparing results of individual-level and household-level analysis

1. Introduction

The instrumental and intrinsic value of education is well-known to researchers and policy makers. Since Schultz's pioneering work on human capital and its role in growth (Schultz 1961), the importance of acquiring education, skill and technological knowledge in a macroeconomic perspective has been widely accepted. Works of endogenous growth theorists like Mankiw et al. (1992) and Barro (1991) have established the positive association between education and economic growth. Education increases skill level and productivity of the workforce. This increases income and, hence, consumption and savings of workers. Human capital formation also facilitates households to emerge out of poverty traps. In particular, education of women is very important as it generates externalities in the form of better health, fertility and nutrition outcomes, and leads to inter-generation transfer of education and knowledge. The works of Sen (1985, 1999) has shown that education is not only important as a means for accelerating growth and development but in its own right. Sen's work focus on education as a basic right—he argues that education is an integral part of capabilities and freedom of individuals.

Given the crucial role of education in growth and development, ensuring equal access to education to all groups is important to attain inclusive growth. However, disparities between ethnic groups, socio-religious communities and – most important – gender have persisted over time in developing countries.

In India, the gender gap in education has remained a persistent feature of society. Census figures reveal that the gender gap in literacy is as high as 16.7 percent in 2011; this is only a small progress since Independence (in 1951, the gender gap was 18.3 percent). Female literacy rates are particularly low in states like Uttar Pradesh (59 percent), Jammu & Kashmir (58 percent), Jharkhand (56 percent), Bihar (53 percent) and Rajasthan (53 percent). The 2001 Census reveals that the gender gap in Primary, Secondary and Higher Secondary completion rates are 3.7 percent, 6.5 and 3.4 percent.[1] There have been several studies on the gender gap in enrolment and on educational attainment in India (Asadullah and Yalonetzky 2012, Kingdon 2005, PROBE 1999). Such studies have clearly indicate a bias against girls, which is explained by factors like social attitudes, low economic returns to educating girls, opportunity costs of educating girls who have to provide household labour, etc. (PROBE 1999). These findings are clearly in line with studies reporting various forms of gender discrimination in India (Dasgupta, 1987).

However, another line of research that attempts to test the presence of gender discrimination in education by comparing the level of household expenditure on education has been less successful. Such studies, based on Engels curve analysis of household level data, have failed to come up with strong evidence of gender discrimination. Interestingly, this failure is also a common feature of the literature on gender discrimination in intra-household allocation of resources on children in other developing countries (Deaton, 1997; Zimmerman, 2012).

In this study, we re-examine this problem using data from National Sample Survey Organization (NSSO) 64th round data (2009) on “Participation and Expenditure in Education”. Unlike other

[1] The relevant population for these calculations are 12 years and above, 16 years and above and 18 years and above.

studies, which test for gender discrimination at either the household level or the individual level, we test for such differentials at *both* levels. We first start by testing for gender differences in educational expenditure at the individual level; in the second step we test the robustness of our findings at the household level. Our study is an improvement over earlier studies as we test for discrimination at two levels *using the same data set*.

The analysis of the paper proceeds as follows: Section 2 starts with a description of the theoretical model of household decision-making. We then describe how the theoretical model is made operational and used to test for gender differences in household allocation of expenditure. This discussion is followed by a review of the empirical evidence on gender differentials in developing countries. In the next section we describe the database used in the study and the econometric methodology. Findings of the study are reported in section 4, while the last section concludes by summing up the results and indicating areas for further research.

2. Materials and method

2.1 Theoretical background[2]

Consider a household divided into two groups—adults (A) and children (C). Adults are both the income earners and decision-makers. They allocate resources to cater to the needs of adults and children. Demand a good consumed by adult members is given by:

$$q_i^A = g_i^A(x^A, p, z^A, z^C) \qquad [1]$$

when,

q_i^A: consumption of good i by adult members

x^A: Share of expenditure allocated to adult members by a sharing rule

[2] Based on Deaton (1997).

p: Price vector (assumed to be constant in our analysis)

z^A: Characteristics of adult members (number and gender composition)

z^C: Characteristics of child members (number and gender composition)

The sharing rule (x^A) may be conceptualized to be of the following functional form:

$$x^A = \theta\ (y, p, z^A, z^C) \qquad [2]$$

There is a corresponding demand function for children and the sum of these two demand functions is the observable household demand.

Now, it may be seen that child characteristics affect demand in two ways. One effect is through the sharing rule (x^A). This implies that when a child is born, this is tantamount to what would have happened if the family had received an income shock as the family members have to cut back on their consumption to feed and clothe the new born child. This effect is the income effect. Simultaneously, adult consumption will have to be rearranged as a result of the birth. This is the substitution effect that operates through the demand function directly.

Rothbart (1943) argues that it may be possible to identify the allocation rule by considering exclusive goods. These are goods that are consumed exclusively by one group of members. Thus adult goods are consumed only by adults, and not by children. Examples are adult clothing, alcohol and tobacco. If we assume that these goods do not have a substitution effect, then the only effect of a birth is through the income effect. This implies that x^C drops out of the demand function [1]. Simultaneously, for simplicity, we can eliminate p from both [1] and [2] as the price vector is constant. Formally,

$$q = g[\theta\ (y, p, z^A, z^C), p, z^C, z^A]$$
$$= g[\theta\ (y, z^A, z^C), z^A]$$

as p is constant and dropping z^C in g(.) as it represent an adult good.

The effect of an extra child on expenditure of each adult good ought to be proportional to the effect of income on the expenditure on that good. It is easy to show that, after substituting [2] into [1] and then differentiating, that

$$\delta q/\delta z^C = (\delta q/\delta\theta).(\delta\theta\ /\delta z^C) \text{ and } \delta q/\delta y = (\delta q/\delta\theta).(\delta\theta\ /\delta y).$$

So that

$$\frac{\partial q/\partial z^C}{\partial q/\partial y} = \frac{\partial\theta/\partial z^C}{\partial\theta/\partial y} \qquad [3]$$

holds for all goods. This allows us to measure the effect of a change in household characteristics on the adult's share of income, measured in terms of income units.

One important limitation of this model is that we are treating children as homogeneous. However, given the gendered nature of South Asia in general and India in particular, the impact of adding a boy and the impact of having a girl on adult consumption should be different. However, as we shall see, this can be taken care of easily to enable us to test for gender discrimination in household allocation of expenditure.

2.2 Empirical verification

Given the lack of data on expenditure at the individual level, researchers are forced to rely on expenditure aggregated at the household level to test for gender discrimination. This method is based on an extension of the Working-Lester linear expenditure function:

$$w_i = \alpha + \beta\log(x_i/n_i) + \gamma\log n_i + \sum_{k=1}^{K-1} \delta_k\,(k_i/n_i) + \eta z_i + \varepsilon_i \qquad [4]$$

when,

w_i: share of i^{th} good in household expenditure

x_i: household expenditure

n_i: household members

k_i: number of members in each age-sex class

z_i: household characteristics

ε_i: error term.

Using the earlier conceptual framework and the concept of adult goods, the Ordinary Least Square estimates of the equation [4] can provide evidence of the presence of gender discrimination in household consumption. The coefficients δ_k for a particular age group denotes the effect of increasing number of children (of the relevant gender) on consumption. If there is gender bias, than the coefficients of girls for each age group will be higher than coefficients of boys in corresponding age groups. Thus, if we test for equality of coefficients for a particular age group, using the Wald test, we will be able to prove/disapprove gender bias in household consumption.

2.3 Empirical evidence

The Engels curve approach has been used to test for gender discrimination in several developing countries. Surprisingly, such studies have not always been successful in detecting significant level of gender discrimination.

For instance, Deaton's analysis of household level data in Cote d'Ivoire and Thailand to test for gender differences in intra household allocation of goods found only a small insignificant bias in favour of boys in Thailand and absence of discrimination in Cote d'Ivoire (Deaton, 1989). Analysis of NSS 1983 data for rural Maharashtra by Subramanian and Deaton (1991) found evidence of pro-male bias only in the age group 10-14 years. Analysis of 2005-06 NSS data by Lancaster et al. (2008) for Bihar, Uttar Pradesh, Kerala and Maharashtra also does not find conclusive evidence of gender discrimination. Only among the 10-16 year age group, that to only for rural Bihar and Maharashtra, is there any evidence of gender differential in intra-family allocation of expenditure in education. Kingdon's (2005) attempt to detect gender bias using individual level data from a NCAER survey also failed to reveal discrimination at the household level, although individual level equations show significant gender bias. Similarly, despite Chinese culture being characterized by a strong son preference, a study by Lee (2007) failed to find any strong evidence of gender bias in rural China using consumption expenditure from World Bank dataset (China Standards of Living Survey 1995). A study of allocation of household resources for healthcare over age and gender groups in rural Burkina Faso also failed to reveal gender differences (Sauerborn et al., 1996). However, more healthcare resources were found to be allocated to sick adults (men and women), who were considered to be productive members of households, than on children. A study under the Mexican Nutrition Collaborative Research Support Programme observed that infants and pre-schoolers did not exhibit significant gender differences in dietary quality and quantity (Backstrand et al., 1997).

On the contrary, some studies actually reveal a bias in intra-household allocation of resources in favour of girls—for instance, Himaz's study of educational expenditure in Sri Lanka (Himaz, 2010). Similarly, Masterson's study of gender differences in education expenditures in Paraguay reported share of expenditure on education to be higher for girls, vis-à-vis boys, in the age groups 5-14 and 15-19, in both rural and urban regions (Masterson, 2012).

As opposed to the lack of evidence for gender bias in intra-household allocation of resources, only a handful of studies have reported statistically significant gender bias in such allocation. Such studies either (a) point out that the population is neither homogenous nor heterogeneous, but composed of dissimilar groups formed by socio-culturally and economically similar households, whose members behave (in terms of intra-household allocation of resources) in broadly similar ways, or (b) criticise the econometric method used to estimate the Engel's curve approach on the grounds that the dependent variable is zero in many cases. Both approaches suggest a movement away from OLS estimation methods to more complex estimation methods.

For instance, one approach is to use mixture models. This is a probabilistic model for representing the presence of subpopulations within an overall population, without requiring that an observed data set should identify the sub-population to which an individual observation belongs. Formally a mixture model corresponds to the mixture distribution that represents the probability distribution of observations in the overall population. However, while problems associated with "mixture distributions" relate to deriving the properties of the overall population from those of the sub-populations, "mixture models" are used to make statistical inferences about the properties of the sub-populations given only observations on the pooled population, without sub-population-identity information. Mixture models have been used to study gender differences in child health outcomes in Bangladesh by Morduch and Stern (1997). Interestingly, although they observed systematic gender differences in health outcomes using mixture models, such differences disappeared when regression method was used.

Alternately, the issue of zero education expenditures may be resolved by using either hurdle models (Cragg, 1971) or semi-parametric methods (Chay and Powell, 2001).

Hurdle models envisage decision-making as a two-step process. In the first step, parents decide on whether to enroll their children or not. In case of enrolled children, this is followed by determination of the allocation of household resources. The advantage of hurdle models is that the functional forms reflecting each decision-making step may be different.

Using hurdle models, Aslam and Kingdon (2008) find significant presence of pro-male bias in enrolment decisions and expenditure on education in junior and secondary school grades in Pakistan; at the primary school level, however, no gender bias in expenditure was observed. Similarly, Zimmermann (2012) used India Human Development Survey data (IHDS, 2005) to test for gender differentials in expenditure on education. Two models were used—the standard Engels Curve approach and a Hurdle model. Zimmermann found that after attainment of age 10, girls experience gender discrimination. Further, the results are more sensitive at the state level as compared to all-India level. Azam and Kingdon (2013), also using a hurdle model, observes that although noteworthy progress is accomplished in the gender equality over the period 1993-2005, pro-male gender bias still exists within household allocation of educational expenditure. Such bias is manifested through:

a) Differential enrolment at secondary level; and,

b) Differential expenditure at primary and middle level.

Further, such bias is substantially greater in rural areas.

One problem with the various techniques, ordinary least square and maximum likelihood method, applied to deal with censored data is that they often result in biased estimates and misspecification of the error distribution. One possible way to overcome existing problems in traditional statistical analysis along with the sources of misspecification in parametric estimation approaches is to use semi-parametric methods. Semi-parametric methodology specifies the functional form of the model (regression function) parametrically in one part, under some plausible assumptions, while the remaining part of the model is not parameterized (Chay and Powell, 2001).

An example of application of semi-parametric methods is Gong et al.'s study of gender differences in expenditure on adult goods (food and alcohol) and education in rural China. Using the Rural Household Income and Expenditure Survey dataset Gong et al. (2005) found the share of expenditure on education to total expenditure to be higher for boys compared to that for girls. On the other hand, gender differentials in allocation of adult goods were marginal.

3. Database and methodology

3.1 Data source

The data used in the analysis is unit-level data from the "Participation and expenditure in education" survey conducted by the National Sample Survey Organization (NSSO) between July 2007 and June 2008 in India (64^{th} round). The NSS 64th Round was designed to collect information on (a) participation of persons aged 5-29 years in the education system of the country (b) private expenditure incurred by households on education and (c) the extent of educational wastage in terms of dropout and discontinuance, and its causes. The survey is a representative national-level survey covering 290,171 individuals from 63,318 households in 7,953 villages and 155,789 households from 37,263 households in 4,682 urban blocks.[3]

In addition to NSSO 64th round data, we also used district level data from the 2001 Census and District Information System on Education (for 2005-06). District-level figures on literacy (Total and Female), Child Sex Ratio (defined as number of females per 1,000 male children in age group 0-6 years), per capita school availability, percentage of schools with female teachers and percentage of schools with separate girls' toilet were appended to the data set. In addition, state-wise daily earnings and work-force participation ratio was estimated from the 61st Round NSSO survey on "Employment and Unemployment" and used to calculate opportunity cost of education (=Earnings * Probability of getting work, when probability of getting work is assumed equal to workforce participation ratio). This was calculated for male and female workers separately, and also for rural and urban areas.

3.2 Econometric models

The advantage of using the NSSO 64th round data is that it provides information on expenditure on education for individual children. This enables us to estimate the regression:

$$EE_i = \alpha + \beta MALE_i + \gamma CV_i + \varepsilon_i \qquad [5]$$

where,

[3] Details of the survey methodology are described in Chapter 1 of NSSO Report No. 532 (NSSO, 2010).

EE_i is education expenditure for the respondent, MALE is a gender dummy (=1 if respondent is a male, 0 otherwise) and CV are control variables. Estimating the above equation using the OLS method and testing H_0 ($\beta = 0$) indicates the presence (absence) of gender discrimination.

Control variables include:

a) Age of the respondent
b) Socio-religious identity: Categories were Hindu Scheduled Castes (HSC), Hindu Scheduled Tribes (HST), Hindu Other Backward Castes (HOBC), Hindu Forward Castes (HFC), Muslims and other socio-religious groups (OSRC).
c) Household monthly expenditure (in log form)
d) Household size
e) Place of residence (rural and urban)
f) Geographical zone of residence
g) Gender of household head
h) Education of household head
i) Occupation of household head
j) Educational infrastructure: Per capita school availability, percentage of schools with female teachers and percentage of schools with separate girls' toilet
k) Proxies for culture: Total and female literacy (district-wise)
l) Proxy for son preference: child sex ratio (district-wise).[4]

Two definitions of expenditure on education are taken:

[4] Jensen (2002) argues that the omission of son preference (a factor not generally incorporated into the Engels' curve equation) can lead to biased estimates of coefficients. The logic is that families with son preference will tend to display a differential stopping behavior—not restricting their fertility till their target number of sons is attained. This implies that families with daughters will, in general, tend to be larger than families with sons. Greater competition for resources in the former type of families will lead to lower educational allotment on girls.

a) E_1 (Tuition fee + Examination fee + Other school fees + Costs of books, uniform, transport and private coaching + Other expenses), and

b) E_2 (= Tuition fee + Examination fee + Other school fees + Costs of books and uniform). For convenience we will refer to them as expenditure on education and expenditure on schooling, respectively.

Although the data was collected at the individual level, we also aggregated the individual level data into household level data. This enabled us to utilize the Engel's curve approach (using [4]) to verify whether the discrimination observed at the individual level carries over to the household level. In other words, we can test for gender discrimination at *two levels* using the *same* data set. The analysis is undertaken for respondents aged 5-20 years.[5]

Kernel densities of expenditure on education for both boys and girls (Figures 1a and 1b show kernel densities for E_1) indicate the presence of a substantial probability spike for value 0. Kernel densities for expenditure on only schooling and uniform (E_2) are similar to total costs of education (E_1). The presence of the spike signifies that estimation of [4] and [5] using OLS may not be methodologically correct as censoring may result in inconsistent estimates of coefficients—yielding a downwards-biased estimate of the slope coefficient and an upwards-biased estimate of the intercept.

[5] This results in only 2.85 percent of the sample with positive expenditure on education being dropped.

Figure 1a: Kernel density of expenditure on education for boys

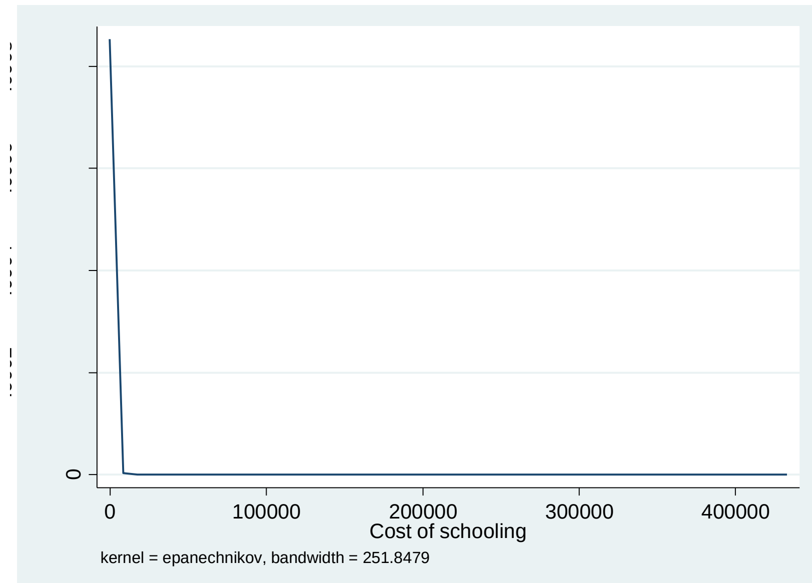
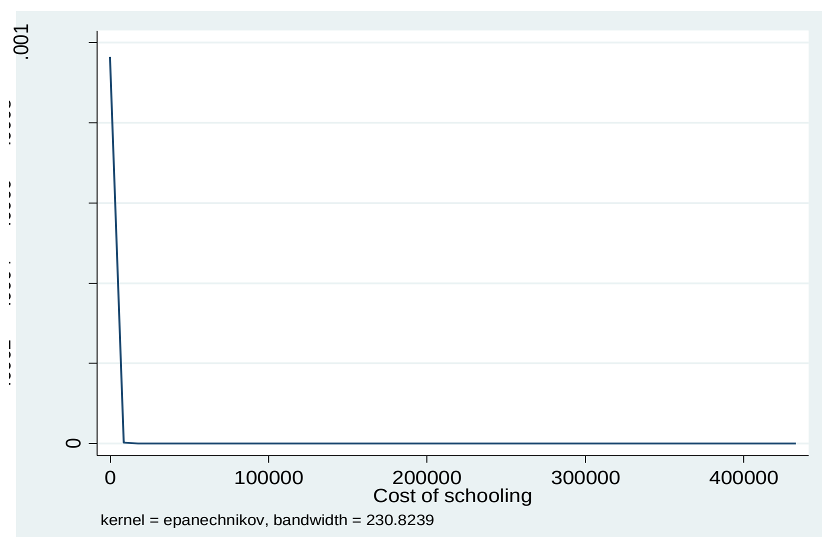


Figure 1b: Kernel density of expenditure on education for girls



To tackle the problem posed by the use of OLS in estimating censored and clustered dependent variables Tobin (1958) suggested the use of model with latent variable (w*).[6] 'Tobit' or Censored normal regression model is used as an alternative to impasse the problem provided the data follows the assumptions of normality and homoscedasticity (Deaton, 1997). The Tobit model attributes the censoring to a standard corner solution by imposing the assumption that the dependent variables censoring around zero is attributable to economic factors alone. But latter on this restrictive assumption was relaxed. Use of latent variable (w*) allows us to assumes a single mechanism that determines the choice between w =0 versus w > 0 *and* the amount of w, given w>0. In particular, the probability of w>0 given x (where x is independent variable) $\partial P\left(w > 0 \mid x\right) / \partial x_j$ and expectation of w>0 given x conditioned for w>0 $\left(\partial E\left(w > 0 \mid x, w > 0\right) / \partial x_j\right)$ are constrained to take same sign.

One problem with the Tobit model is that it can be too restrictive because a single mechanism governs the "enrolment decision" (y = 0 versus y > 0) and the "amount decision" (how much y is if it is positive). In a Tobit model, for a continuous variable xj, the partial effects on P(y> 0|x) and E(y|x, y > 0) have the same signs (different multiples of βj). So, it is impossible for xj to have a positive effect on P(y > 0|x) and a negative effect on E(y|x, y > 0). This may also occur for discrete covariates.

Further, for continuous variables x_j and x_h:

---

6 $w_i = \begin{cases} w_i^* & \text{if } w_i^* > 0 \\ 0 & \text{if } w_i^* \le 0 \end{cases}$

$$\frac{\partial P(y > 0|\mathbf{x})/\partial x_j}{\partial P(y > 0|\mathbf{x})/\partial x_h} = \frac{\beta_j}{\beta_h} = \frac{\partial E(y|\mathbf{x}, y > 0)/\partial x_j}{\partial E(y|\mathbf{x}, y > 0)/\partial x_h} \quad [6]$$

So, if x_j has twice the effect as x_h on the enrolment decision, xj must have twice the effect on the expenditure decision, too. Two-part models allow different mechanisms for the enrolment and expenditure decisions.

It is therefore useful to have a general way to think about two-part models without specific distributions. Let w be a binary variable that determines whether y is zero or strictly positive. Let y* be a nonnegative, continuous random variable. Assume y is generated as y = w X y*. Other than w being binary and y* being continuous, there is another important difference between w and y*—we effectively observe w because w is observationally equivalent to the indicator 1[y > 0] (P(y* = 0)). But y* is only observed when w = 1, in which case y* = y. Generally, we might want to allow w and y* to be dependent, but that is not as easy as it seems. A useful assumption is that w and y* are independent conditional on explanatory variables x, which we can write as:

$$D(y^*|w, x) = D(y^*|x) \quad [7]$$

This assumption typically underlies two-part or *hurdle models*. One implication is that the expected value of y conditional on x and w is easy to obtain:

$$E(y|x,w) = w \,.\, E(y^*|x,w) = w \,.\, E(y^*|x) \quad [8]$$

When w = 1, we can write E(y|x, y > 0) = E(y*|x), so that the so-called "conditional" expectation of y (where we condition on y > 0) is just the expected value of y* (conditional on x). The so-called "unconditional" expectation is:

$$E(y|x) = E(w|x)E\,(y^*|x) = P(w = 1|x)E(y^*|x) \quad [9]$$

Cragg (1971) proposed a natural two-part extension of the type I Tobit model. The conditional independence assumption is assumed to hold, and the binary variable w is assumed to follow a probit model:

$$P(w = 1|x) = \Phi(x\gamma) \quad [10]$$

Further, y* is assumed to have a truncated normal distribution with parameters that vary freely from those in the probit. This implies that we can write $y^* = x\beta + u$, where u given x has a truncated normal distribution with lower truncation point $-x\beta$.

Because $y = y^*$ when $y > 0$, we can write the truncated normal assumption in terms of the density of y given $y > 0$ (and x):

$$f(y|x, y > 0) = [\Phi(x\beta/\sigma)]^{-1}\phi[(y - x\beta)/ \sigma]/ \sigma, y > 0, \quad [11]$$

where the term $[\Phi(x\beta/\sigma)]^{-1}$ ensures that the density integrates to unity over $y > 0$. The density of y given x can be written succinctly as

$$f(y|x) = [1 - \Phi(x\gamma)]^{1[y=0]}\{ \Phi(x\gamma)[\Phi(x\beta/\sigma)]^{-1}\phi[(y - x\beta)/ \sigma]/ \sigma\}^{1[y>0]}, \quad [12]$$

where we must multiply $f(y|x, y > 0)$ by $P(y > 0|x) = \Phi(x\gamma)$. This is called the truncated normal hurdle (THN) model. A nice feature of the TNH model is it reduces to the type I Tobit model when $\gamma = \beta/\sigma$. The conditional expectation has the same form as the Type I Tobit because D(y|x, y > 0) is identical in the two models:

$$E(y|x, y > 0) = x\beta + \sigma\lambda (x\beta/ \sigma) \quad [13]$$

In particular, the effect of x_j has the same sign as β_j (for both continuous or discrete changes). The partial effects of this mode are given by:

$$\hat{\beta} * \exp (\hat{\beta}\ \bar{x} + \sigma^2/2) \quad [14]$$

In addition to OLS and Tobit models, we have therefore also used the hurdle model. In section 2.3 we had seen such models have been used to test for gender differentials and usually produce better results than OLS models.

3.3 Sample profile

A summary of the data used in this study is given in Tables 1 and 2. Summary results for age group 5-20 years show that the mean age of respondents is respectively 12.54, 12.39 and 12.84 years for total, rural and urban areas. Predictably, the mean monthly expenditure (in log) for rural area is less than the mean expenditure in urban areas. Urban households spend, on an average, three times more on education than rural household. This also holds for education costs (comprising of school fees, uniform, examination fees and stationary). Child sex ratio is higher in rural areas, compared to urban areas. This may reflect easier access to technology for pre-natal sex determination and selective abortion in urban areas.

Table 1: Descriptive statistics for 5-20 age groups — Total, Rural and Urban

Variable	**Mean**			**Median**		
	Total	**Rural**	**Urban**	**Total**	**Rural**	**Urban**
Individual and household characteristics:						
Age of respondents	12.54	12.39	12.84	12	12	13
Household monthly expenditure (log)	6.49	6.32	6.83	6.44	6.29	6.80
Household size	5.88	5.99	5.66	5	6	5
Expenditure on education:						
Cost of Education (E_1)	2009.38	1102.56	3881.48	410	264	1300

Cost of Schooling (E_2)	1287.85	667.55	2568.44	195	130	655
Educational infrastructure:						
Per capita school availability	13.80	14.39	12.57	10.78	11.42	9.77
Percentage of schools with female teachers	24.33	25.91	21.06	23.85	25.08	20.87
Percentage of schools with separate girls' toilet	50.84	48.72	55.22	50.16	47.88	56.1
Proxies for culture:						
Total Literacy	63.80	61.81	67.90	64.5	62.1	69.2
Female Literacy	52.75	50.34	57.72	53	50.1	58
Child Sex Ratio	929	932	923	943	946	939

Source: Based on the authors' calculation using NSS 64th round on "Participation and expenditure in education".

Table 2 portrays household characteristics. Analysis reveals that 63 per cent households reside in rural area. Further, Hindu Other Backward Castes (HOBC) is numerically the dominant socio-religious class in both rural and urban areas (combined), followed by Hindu Scheduled Castes (HSC) in rural areas and Hindu Forward Castes (HFC) in urban areas. Analysis of household characteristics indicates that few households are headed by females (about one out every eight families). This holds for both rural and urban areas. In rural areas most of the household heads are either illiterate or have below primary level education (55 percent). In contrast, almost 70 percent of urban households are headed by persons with at least primary education levels. Predictably, the majority (72 percent) of rural households have household heads working in the primary sector. Urban households, on the other hand, are scattered among three occupations—‘Craft, Trade, etc.’, ‘Primary Producers’ and ‘Managerial & Professional’.

Table 2: Household characteristics (in percentage)—Total, Rural and Urban

Variable	**Percentage**		
	Total	**Rural**	**Urban**
Place of residence:	-	62.95	37.05
Socio-religious Categories:			
Hindu Forward Castes (HFC)	22.36	16.18	32.86
Hindu Scheduled Castes (HSC)	16.32	18.66	12.35
Hindu Scheduled Tribes (HST)	7.31	9.92	2.88
Hindu Other Backward Castes (HOBC)	31.72	34.29	27.35
Muslims	11.74	10.19	14.37
Other socio-religious groups (OSRC)	10.55	10.77	10.19
Gender of household head:			
Male	88.46	88.72	88.02
Female	11.54	11.28	11.98
Education of household head:			
Illiterate/Just Literate	34.13	43.20	18.71
Below primary	9.89	11.47	7.20
Below secondary	30.06	30.35	29.56
Below Higher Secondary	10.80	8.08	15.43

Completed Higher Secondary	5.61	3.34	9.48
Above Higher Secondary	9.51	3.56	19.61
Occupation of household head:			
Primary Producers	34.79	41.98	22.05
Primary Workers	20.94	30.61	3.80
Managerial & Professional	10.75	4.71	21.43
Craft, Trade, etc.	15.21	10.33	23.87
Technical Workers & Clerk	7.03	3.26	13.71
Elementary Occupations	11.28	9.11	15.14

Source: Based on the authors' calculation using NSS 64th round.

4. Results and discussion

4.1 Gender differential at individual level

This section reports results for econometric analysis based on equation [5]. This equation tests for gender discrimination in education expenditure taking individual children as the unit of analysis. Equation [5] is estimated for the sample aged 5-20 years (Total, Rural and Urban) and for the age groups 5-10 years, 11-15 years and 16-20 years. Results are reported in Table 3a for total expenditure on education (E_1) and in Table 3b for expenditure on schooling (E_2).

Table 3a: Summary results of test for gender differential in cost of education at individual level

	All India			
	5-20 years	**5-10 years**	**11-15 years**	**16-20 years**
Coefficient of Male	0.1691	0.1462	0.1327	0.1345

t	21.30***	12.96***	11.85***	8.85***
N	91258	42595	33146	15517
F	3120.05	1570.82	1044.90	379.09
R^2	0.4293	0.4480	0.4097	0.3499
		All India: Rural		
	5-20 years	**5-10 years**	**11-15 years**	**16-20 years**
Coefficient of Male	0.2064	0.1599	0.1441	0.1647
t	20.43***	11.85***	10.42***	8.19***
N	58961	29076	21636	8249
F	1129.72	546.72	365.51	124.44
R^2	0.2870	0.2832	0.2621	0.2411
		All India: Urban		
	5-20 years	**5-10 years**	**11-15 years**	**16-20 years**
Coefficient of Male	0.1140	0.1179	0.1209	0.1031
t	9.00***	5.82***	6.42***	4.48***
N	32297	13519	11510	7268
F	1110.06	465.95	406.52	177.85
R^2	0.4194	0.4203	0.4263	0.3401

*Note: *** denotes 1% level of significance.*

Table 3b: Summary results of test for gender differential in cost of schooling at individual level

		All India		
	5-20 years	**5-10 years**	**11-15 years**	**16-20 years**

Coefficient of Male	0.1908	0.1611	0.1552	0.1547
t	22.45***	13.48***	13.15***	9.41***
N	90802	42281	33030	15491
F	2872.19	1527.47	934.53	327.12
R^2	0.4104	0.4430	0.3838	0.3175

All India: Rural

	5-20 years	**5-10 years**	**11-15 years**	**16-20 years**
Coefficient of Male	0.2223	0.1640	0.1650	0.1736
t	20.93***	11.82***	11.60***	8.11***
N	58639	28842	21559	8238
F	983.04	494.08	309.30	107.01
R^2	0.2605	0.2647	0.2317	0.2148

All India: Urban

	5-20 years	**5-10 years**	**11-15 years**	**16-20 years**
Coefficient of Male	0.1442	0.1527	0.1436	0.1343
t	10.26***	6.72***	6.94***	5.32***
N	32163	13439	11471	7253
F	990.03	437.06	351.87	150.85
R^2	0.3928	0.4062	0.3922	0.3046

*Note: *** denotes 1% level of significance.*

Results for equation [5], based on both definitions of expenditure on schooling (E_1 and E_2), reveal similar results (Table 3a and 3b). The coefficient of the Male dummy is positive and significant at 1% level in all cases. At the individual level, therefore, our data set proves conclusive evidence that Indian parents spend more on the education of boys vis-à-vis girls. Let us now turn to the household level and see whether the evidence for discrimination is equally conclusive.

4.2 Gender differential at household level

To test for discrimination at the household level we have to estimate equation [4] using household level data and test for equality of coefficients. This has been done for both definitions of education—when w_1=log(E_1/Total household expenditure) and w_2=log(E_2/Total household expenditure) using the OLS method.

Contrary to standard results, we find a statistical difference in coefficients across gender, with the coefficient for boys being greater than that of girls. Only in one case—the 16-20 year age group residing in urban areas—does the Wald test fail to show gender differential (for w_2, estimated using OLS).

When estimating equation [5] using OLS regression both zero and positive values of w_i are considered on the grounds that the decision to enroll a child in school also affects the decision on spending. Another justification for including all households is based on the assumption that the dependent variable (expenditure on education) is normally distributed instead of following a log-normal distribution. However, it is often found that in the case of analysis of households expenditure share on education there is clustering of household at zero reported value so that the share of education in household expenditure too is censored at zero. Figure 1 reveals that there is a probability spike at zero. Use of OLS as the estimation method is not appropriate in such cases. OLS regression results tend to display a downward bias and may also be inconsistent.

To test the robustness of the results of the OLS model we have also estimated the Tobit variant. Results are even more conclusive for the Tobit model as significant gender differential is observed in all age groups (Table 4).

Table 4: Summary results for testing gender differential at household level among age groups

Method	Sample	m510r=f510r	m1115r=f1115r	m1620r=f1620r
OLS	**All**	124.37***	140.42***	164.42***
(W_1)		(0.0000)	(0.0000)	(0.0000)
	Rural	73.98***	124.78***	343.69***
		(0.0000)	(0.0000)	(0.0000)
	Urban	52.54***	36.23***	0.81
		(0.0021)	(0.0000)	(0.3693)
OLS	**All**	92.65***	96.70***	153.71***
(W_2)		(0.0000)	(0.0000)	(0.0000)
	Rural	61.71***	107.01***	251.49***
		(0.0000)	(0.0000)	(0.0000)
	Urban	36.04***	19.81***	10.69
		(0.0000)	(0.0000)	(0.0011)
Tobit	**All**	324.23***	353.78***	353.15***
(W_1)		(0.0000)	(0.0000)	(0.0000)

	Rural	225.30***	317.62***	565.86***
		(0.0000)	(0.0000)	(0.0000)
	Urban	95.64***	64.58***	7.41
		(0.0000)	(0.0000)	(0.0065)
Tobit	**All**	273.39***	295.13***	352.41***
(W_2)		(0.0000)	(0.0000)	(0.0000)
	Rural	208.36***	298.35***	479.32***
		(0.0000)	(0.0000)	(0.0000)
	Urban	69.47***	43.64***	23.25***
		(0.0000)	(0.0000)	(0.0000)

Note: '***' denotes 5% level of significance.

4.3 Hurdle model

Finally, we present results of the two-stage hurdle models (Table 5). An important advantage of hurdle models is that they allow us to identify whether the gender differential in expenditure on education is a result of non-enrolment in schools (at the entry point, or at different stages), or is a result of lower allotments due to enrolment in 'cheaper' and/or 'nearer' schools and economizing on private coaching.

Table 5: Results of Hurdle model for household-level data

Model		**Total**			**Rural**			**Urban**		
		Male	**Female**	**Chi-square**	**Male**	**Female**	**Chi-square**	**Male**	**Female**	**Chi-square**
Probit	**5-10**	1.1289	.9966	59.80***	1.0216	.8806	51.28***	1.400	1.3178	5.61***
				(0.0000)			(0.0000)			(0.0178)

	11-15	.7809	.6241	57.36*** (0.0000)	.7021	.5437	42.63*** (0.0000)	.9753	.8174	15.47*** (0.0001)
	16-20	.0944	.0784	0.82 (0.3665)	-.01189	-.0507	3.20 (0.0736)	.3007	.3462	2.10 (0.1469)
		Male	**Female**	**F stat**	**Male**	**Female**	**F stat**	**Male**	**Female**	**F stat**
W_1	**5-10**	0.5477	0.4879	29.67*** (0.0000)	0.5895	0.5107	27.88*** (0.0000)	0.4223	0.3992	2.10 (0.1473)
	11-15	0.1517	0.0813	56.10*** (0.0000)	0.1243	0.0449	38.08*** (0.0000)	0.1623	0.1164	13.28 (0.0003)
	16-20	-0.0030	-0.0396	16.33*** (0.0001)	-0.0620	-0.1021	9.92 (0.0016)	0.0575	0.0342	4.02 (0.0450)
W_2	**5-10**	0.8896	0.7841	33.45*** (0.0000)	0.9715	0.8211	38.87*** (0.0000)	0.6711	0.6506	0.23 (0.6286)
	11-15	0.3105	0.1839	74.10*** (0.0000)	0.2613	0.1231	49.21*** (0.0000)	0.3323	0.2400	19.66*** (0.0000)
	16-20	-0.0424	-0.0945	14.79*** (0.0001)	-0.1427	-0.1994	9.27 (0.0023)	0.0639	0.0339	2.66 (0.1032)

Note: '***' denotes 5% level of significance.

Probit model reports coefficients in top panel, while the hurdle model reports marginal effects for both definitions of expenses on education (bottom two panels).

Table 5 reveals that, in both rural and urban areas, boys have a significantly lower probability of enrolment than girls in primary schools and in the post-primary level (age groups 5-10 and 11-15 years, respectively). In urban areas, the gender differential actually reverses after the secondary level (age group 16-20 years), but the difference is insignificant even at the 10 percent level. In rural areas, the coefficients of both boys and girls are negative. This is somewhat unusual. The negative value implies that as the number of children in this age group increases, households reduce expenditure on education. Note, however, that the reduction in expenditure is significantly more for an increase in number of girls—indicating the presence of gender differential. All India results are similar to those for the rural sample.

When we turn to the expenditure decision, we find that gender differences persist for the all-India and rural samples. In urban areas, however, gender differential is significant only for the age group 11-15 year age group. This is true for both W_1 and W_2. There are several reasons why results vary between rural and urban areas. A common explanation for gender differential is that sons are considered as future investment for the family so more household resources are allocated on their education. This effect is stronger in rural areas due to shortage of schools providing higher education. Moreover, in rural areas, female workers are generally engaged in primary occupations, where education is not a pre-requisite of securing work. Other possible reasons for gender differential in rural areas are household chores, early marriage, social norms, low mobility.

5. Conclusion

This study had attempted to test for the presence of gender differential in expenditure on education in India. While other studies use household level data or individual level data our analysis is undertaken at both household and individual levels, using the same data set, generated using a questionnaire framed to collect detailed data on expenditure and other indicators of educational attainments. This allows us to overcome a possible shortcoming of household level surveys of expenditure, viz. their failure to capture expenditure on education accurately (Zimmerman, 2012).

Results confirm the presence of gender differentials at both the individual and household level. This discrimination can be attributed to either one, or both, of two processes—enrolment decision, and decision to spend. The hurdle model allows us to model expenditure decisions as a two-part process, capturing this aspect. Results of the hurdle model indicates that, in rural areas, the decision to enrol and the decision to spend are both characterized by discriminatory attitudes towards the girl child. In contrast, in urban areas, discrimination marks only the enrolment process—but not the decision to spend. In other words, once a girl is enrolled, she gets a share of household resources equal to her male siblings.

These findings enable fine tuning of policies to reduce discrimination in education. In rural areas, policy makers should attempt to increase both enrolment and provide subsidies for educating girls. The policy of giving cycles to school-going children in Bihar, for instance, has been successful. Distribution of free textbooks and stationary and subsidised school uniforms may be other means of reducing gender disparities in expenditure on education. Parental motivation for educating girls may also be strengthened by ensuring that female teachers are posted; availability of separate girls' toilets is another useful step. In urban areas, on the other, the focus must be on increasing enrolment and retention. The focus must be on increasing awareness among parents. Such intervention strategies may be made even more target-oriented and specific by extending this study in other directions.

For instance, our analysis has been undertaken at the all-India level. Although we have incorporated dummies for geographical zones to account for possible variations in culture across regions, a possible extension is to undertake the analysis for the major states. Secondly, we have considered aggregate expenditure on education. An interesting exercise would be to look at component-wise expenditure on education—or, at least for school fees, private tuition, and similar major components of expenditure on education—and test exactly for which items there is discrimination.

Another limitation of this study is that we are assuming that the educational standard and expenditure is a measure of educational attainments. That this is a strong assumption has been shown by the All India Survey of Education (ASER) reports. However, institutions differ with respect to the quality of education they provide. Apart from expenditure or standards, it is also necessary to undertake a study to assess the quality of education received by boys and girls and examine whether it differs markedly.

References

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal, *Econometrica*, 41, 997–1016.

Asadullah, M. N., & Yalonetzky, G. (2012). Inequality of Educational Opportunity in India: Changes over Time and Across States. *World Development*, 40(6), 1151–1163.

Aslam, M. and Kingdon, G. G. (2008). Gender and household education expenditure in Pakistan. *Applied Economics*, 40(20), 2573–2591.

Azam, M., & Kingdon, G. G. (2013). Are Girls the Fairer Sex in India? Revisiting Intra–household Allocation of Education Expenditure. *World Development*, 42(1), 143–164.

Backstrand, J. R., Allen, L. H., Pelto, G. H., & Chávez, A. (1997). Examining the gender gap in nutrition: an example from rural Mexico. *Social Science & Medicine*, 44(11), 1751–1759.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics*, 106(425), 407–443.

Chay, K. Y., & Powell, J. L. (2001). Semiparametric Censored Regression Models. *Journal of Economic Perspectives*, 15(4), 29–42.

Cragg, J. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39, 829–844.

Dasgupta, M. (1987). Selective Discrimination against Female Children in Rural Punjab, India. *Population & Development Review*, 13(1), 77-100.

Deaton, A. (1989). Looking for Boy-Girl Discrimination in Household Expenditure Data. *The World Bank Economic Review*, 3(1), 1–15.

Deaton, A. (1997). The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. The Johns Hopkins University Press, Baltimore, Maryland, USA.

Duraisamy, P. (2002). Changes in returns to education in India, 1983–94: By gender, age-cohort and location. *Economics of Education Review*, 21(6), 609–622.

Gong, X., van Soest, A. and Zhang, P. (2005). The effects of the gender of children on expenditure patterns in rural China: a semi parametric analysis. *Journal of Applied Econometrics*, 20(4), 509–527.

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

Jensen, R. (2002). Fertility preferences and female disadvantage: equal treatment, unequal outcomes? Mimeo, John F. Kennedy School of Government, Harvard University.

Kingdon, G. (2005) .Where has all the Bias Gone? Detecting Gender Bias in the Intra-household Allocation of Educational Expenditure in Rural India. *Economic Development and Cultural Change*, 53(2), 409–452.

Lancaster, G., Maitra, P. and Ray, R. (2008). Household expenditure patterns and gender bias: evidence from selected Indian states. *Oxford Development Studies*, 36(2), 133–157.

Larme, A. C. (1997). Health Care Allocation and Selective Neglect in Rural Peru. *Social Science & Medicine*, 44(11), 1711–1723.

Lee, Yiu-fai D. (2008). Do families spend more on boys than on girls? Empirical evidence from rural China. *China Economic Review*, 19, 80–100.

Mankiw, N.G., Romer, D., & Weil, D.N. (1992). A contribution to the empirics of economic Growth. *Quarterly Journal of Economics*, 107, 407–437.

Masterson, T. (2012). An Empirical Analysis of Gender Bias in Education Spending in Paraguay. *World Development*, 40 (3), 583–593.

Miller, B. D. (1997). Social class, Gender and Intra-household Food Allocations to Children in South Asia. *Social Science & Medicine*, 44(11), 1685–1695.

Morduch, J.J., & Stern, H. S. (1997). Using mixture models to detect sex bias in health outcomes in Bangladesh. *Journal of Econometrics*, 77, 259–276.

NSSO (2010), Education in India, 2007–08: Participation and Expenditure, Report No. **532** NSS 64th Round (July 2007–June 2008)**.**

Probe Team. (1999). Public Report on Basic Education in India. New Delhi, Oxford University Press.

Rothbart, E. (1943). Note on a Method of Determining Equivalent Income for Families of Different Composition in: *Wartime Patterns of Saving and Spending,* Charles Madge, Cambridge University Press, appendix 4.

Sauerborn, R., Berman, P. & Nougtara, A. (1996). Age bias, but no gender bias, in the intra–household resource allocation for health care in rural Burkina Faso. *Health Transition Review*, 6(2), 131–145.

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, 51, 1–17.

Sen, A.K. (1985). Commodities and Capabilities. Oxford, Oxford University Press.

Sen, A.K. (1999). Development as Freedom. Oxford, Oxford University Press.

Subramanian, S. & Deaton, A. (1991). Gender effects in Indian consumption patterns. *Sarvekshana*, 14(4), 1–12.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26, 24–36.

Xiaodong, G., A.VanSoest & Zhangc, P. (2005). The effects of the Gender of Children on Expenditure Patterns in Rural China: A Semiparametric Analysis. *Journal of Applied Econometrics*, 20, 509–527.

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