Choice of Gender in High School Dropout: A Comparison of MNL and BIMNL Models

Soumik Biswas, Indian Institute of Management Lucknow–Noida Campus, India
Kaushik Bhattacharya, Indian Institute of Management Lucknow–Noida Campus, India

The Asian Conference on the Social Sciences 2023
Official Conference Proceedings

Abstract

Purpose – To estimate gender disparity in high school dropouts when households experience financial shocks, this study aims to extend the simple decision-making process of school dropouts to allow for the multiplicity of dropout choices with respect to gender in households having exactly one boy and one girl in high school. We also account for heterogeneity in the households choosing the baseline "no dropout" category and resulting bias in the estimated effects of variables by comparing Baseline Inflated Multinomial Logit (BIMNL) models with Multinomial Logit (MNL) models. Design/methodology/approach – We filter households from India Human Development Survey (IHDS) dataset having exactly one boy and one girl in high school. We estimate Multinomial Logit (MNL) model and to account for the heterogeneity of the no dropout case, we estimate BIMNL models developed by Bagozzi and Marchetti (2017) and compare average partial effects (APE) and econometric performance of MNL and BIMNL models.

Findings – We find that Households facing shocks due to marriage or crop failure are more likely to drop out the girl. BIMNL only could capture the association of crop failure with dropout of girl. Econometrically BIMNL models performed better than MNL.

Practical implications – It would be pertinent to examine whether girls are especially at higher risk of high school dropout with these more sophisticated econometric tools to draw effective policy interventions in the backdrop of financial shocks due to Covid-19 / geopolitical crisis across the world.

Keywords: Gender Disparity, High-School Dropout, Financial Shocks, Multinomial Logit, Baseline Inflated Multinomial Logit

1. Introduction

The decision-making process of school dropout is often modelled as a trade-off. This trade-off could be modelled either in a cost-benefit framework (Dreze and Kingdon, 2001) or in a utility maximization framework comparing the sum of present and future period utilities of the options of dropping out or continuing education (Kearney and Levine, 2016)[1]. However, studies on school dropouts often fail to recognize that different households in a sample may have different choice problems with respect to the number of children and gender.[2] To estimate the real impact of any disparity due to gender, households in a sample should have at least a boy and a girl. It may be noted that even for the simplest case of just one boy and one girl, households have multiple discrete choice options e.g., no dropout, dropout of the girl, dropout of the boy, and dropout of both. A simple probit or logit model would not be able to capture a choice among these four options. Further, survey data on dropouts often reports too many "no dropout" cases, leading to apparent inflation in the zero or the reference category of the multinomial choice problem. Extending the logic of Bagozzi and Marchetti (2017) in the context of the dropout problem, we argue that "zero inflation" in no dropout can come from several sources and non-recognition of this in a simple Multinomial Logit (MNL) model can induce a bias in the estimated effects of variables and this, in turn, may lead to faulty inferences. Bagozzi and Marchetti (2017) developed a Baseline Inflated Multinomial Logit (BIMNL) model that accounts for the baseline category inflation and argued that the econometric performance of this BIMNL model will be superior in terms of model fit criteria like BIC or AIC compared to standard MNL models.

In this paper, we extend the simple decision-making process of school dropouts to allow for the multiplicity of choices with respect to gender and apply this extended framework to the India Human Development Survey (IHDS) dataset. To avoid complexity, we restrict ourselves to the dropout choice problem of households with one boy and one girl only, both children being of school going age. As income shock is found to be one of the most important determinants of dropout, we examine how and to what extent dropout decisions of households are influenced by a financial shock. Several studies (Boyle et al., 2002; Heltberg et al., 2013) in some Asian and African countries (Bangladesh, Nepal, Uganda, Zambia, etc.) report that in response to adverse financial shocks, dropout rates increase substantially and it is the girl child who becomes a victim. In the Indian context, a recent study by Sowmya, Paul, and Gade (2019) on young school-going cohorts in Andhra Pradesh and Telangana, however, did not find significant effects of income shocks on enrolment in the primary level.

Theoretically, during or after a financial shock, the discounted future period utilities of education may change significantly for high school going boys and girls in the case of some households. A household in such a situation may choose to drop out a girl or a boy or both, depending upon the nature and the severity of the shock. Though, even such households can choose a "no dropout". However, following Bagozzi and Marchetti (2017), we argue that many of the households in a given sample will be effectively impervious to the effects of covariates like an adverse financial shock as these households might be unwilling to discontinue the education of any of their children, no matter what the benefits of the dropout choice of any or both of the children might provide. The inclusion of both categories of households that choose no dropout in the reference category of a multinomial choice problem may cause biased estimation of coefficients and may lead to faulty inferences. Empirically, to capture the

[1] For more detail on these approaches, see Dreze and Kingdon (2001) and Kearney and Levine (2016).

[2] For example, households that do not have a girl cannot discriminate against a girl. Many papers reported in a recent survey by Momo et al. (2019) do not recognize the difference.

multiplicity of choice, we specify and estimate the MNL model. In addition, to recognize the heterogeneity of the no dropout case, we specify and estimate the BIMNL model developed by Bagozzi and Marchetti (2017). The paper presents empirical applications of these models in India, constructing a data set from two waves of the IHDS dataset. In order to make the theoretical framework and the empirical application consistent, we estimate both these models on a filtered dataset of households that have only one boy and one girl, both of school-going age. We analyse the roles played by different types of financial shocks on dropout decisions at the high school level in India with respect to gender. We compare their econometric performance and compute the average partial effects (APE) of explanatory variables and attempt to suggest policies that may mitigate the problem.

Section 2 of this paper describes the analytical framework, Section 3 presents data and descriptive statistics, Section 4 carries out an empirical analysis, and Section 5 concludes the paper.

2. Analytical Framework

A dropout decision is often treated as a joint decision within a household. Both the cost-benefit framework (Dreze and Kingdon, 2001), and the utility maximization framework (Kearney and Levine, 2016) have been developed considering a single child. For example, in the utility maximization framework of Kearney and Levine (2016), a dropout of an individual from school is chosen if the following condition is met:

$$u^d_0+E(V^d)>u^e_0+E(V^e) \qquad (1)$$

It is assumed that a dropout decision is taken when the sum of the present period utility of dropout u^d_0 and the expected discounted sum of future period utilities after dropout $E(V^d)$ is greater than the sum of present period utility of continuing education u^e_0 and expected discounted sum of future period utilities after completion of education $E(V^e)$. It is also assumed that $E(V^e)>E(V^d)$.

However, with the increasing number of children in the household the choice problem becomes more complex as with n number of children, there can be 2^n number of choices. Thus for 2 high school going children (say, A and B), there are 2^2 or 4 choices: no dropout, dropout A, dropout B, or dropout both. Therefore, to estimate the impact of any disparity due to gender and also to keep the choice problem manageable we select households with just one boy and one girl. In such households, the total utilities of the household corresponding to the four dropout choices available to the household are as follows:

We can write,

$$U^0 = u^{eg}{}_0+E(V^{eg}) + u^{eb}{}_0+E(V^{eb}) \qquad (2)$$
$$U^1 = u^{dg}{}_0+E(V^{dg}) + u^{eb}{}_0+E(V^{eb}) \qquad (3)$$
$$U^2 = u^{eg}{}_0+E(V^{eg}) + u^{db}{}_0+E(V^{db}) \qquad (4)$$
$$U^3 = u^{dg}{}_0+E(V^{dg}) + u^{db}{}_0+E(V^{db}) \qquad (5)$$

Where,

U^0, U^1, U^2, *and* U^3 be the total utilities of the household corresponding to the four dropout choices available to the household, e.g. no dropout, dropout of the girl, dropout of the boy, and dropout of both respectively.

$u^{eg}{}_0$ and $u^{eb}{}_0$ are the present period utility of the household for continuing education for the girl and the boy respectively,

$u^{dg}{}_0$ and $u^{db}{}_0$ are the present period utility of the household for dropout of the girl and the boy respectively,

$E(V^{eg})$ and $E(V^{eb})$ are expected discounted sum of future period utilities of the household after completion of education for the girl and the boy respectively, and

$E(V^{dg})$ and $E(V^{db})$ are the expected discounted sum of future period utilities of the household after dropout for the girl and the boy, respectively.

Note that U^0, U^1, U^2, *and* U^3 can be different for a household. For example, the present period utility of dropout may increase due to an increase in the direct and indirect cost of schooling and also due to increased opportunity cost of a child's time[3]. Future period benefits of schooling can be both instrumental and intrinsic – which might include higher returns through employment, increased cognitive and other skills, facilitate partner selection in marriage and /or reduction of its costs (mostly for daughters), and intergenerational benefits like better education and health outcomes for offspring. However, the actual benefits of education to a child can be greater than the perceived utilities of it to the household (Dreze and Kingdon, 2001). Based on studies conducted in Madagascar and the Dominican Republic, Banerjee and Duflo (2011) conclude that simple intimation of average income gains from spending one more year in school to parents can improve the educational outcome of students.

The evaluation of utilities of dropout choices may also depend on other factors or household characteristics like parents' education level, the income of the household, the father's age, information available to parents about the benefits of education, etc. Greater income and parental education are associated with a lesser likelihood of dropping out of school (Hunt, 2008; Murnane, 2013; Momo et al., 2018). The greater age of the father is also associated with dropout (Momo et al., 2018; Siddhu, 2011). Mvroniuk et al. (2017) show that students from families with more social capital achieve higher educational milestones.

High school students in India receive lesser support compared to primary and upper primary students (e.g., in terms of no-detention policy, midday meals, school fees etc) from the government. Moreover, high school going children have a greater opportunity cost of schooling. Therefore present period utility of dropout for both girls and boys $u^{dg}{}_0$ and $u^{db}{}_0$ are likely to be greater at the high school level than the present period utility of continuing education $u^{eg}{}_0$ and $u^{eb}{}_0$ for boys and girls compared to primary / upper primary levels.

Further, the opportunity cost of a child's time or utilities of dropout can also be different for high school going girls and boys. As per prevailing social norms, a boy is generally considered more suitable to be employed in wage work outside a household and a girl in household activities. Moreover, for different households, the expected discounted sum of future period utilities of the household after completion of education or after discontinuation of education for the girl and the boy, $E(V^{eg})$, $E(V^{eb})$ and $E(V^{dg})$, $E(V^{db})$ respectively, can be different. In the Indian context, after marriage a daughter usually leaves her parents and joins her husband's

[3] Direct costs of schooling are expenditures on school fees, books, and stationeries and indirect costs include efforts on the part of parents or guardians in terms of motivating and helping the child for school (Dreze and Kingdon, 2001).

family, some parents may often see a lesser benefit in educating a daughter while others may send a daughter to school out of genuine concern for her own wellbeing.

Evaluated utilities of the multinomial dropout choices (U^0, U^1, U^2, U^3) can change when families experience large financial shocks. For some households, the discounted future period utilities may change substantially after a financial shock resulting in a differential discounted sum of future period utilities for girls and boys. If the discounted utilities differ across gender, a household in such a situation may choose to drop out a girl or a boy or both, based on their choice utilities. However, many households in a sample can choose a "no dropout" option. These "no dropout" households could be a heterogeneous group in the sense that among them, there can also be some for whom the discounted utilities would not change after a financial shock or these households are effectively impervious to the effects of adverse financial shock as these households might be unwilling to discontinue the education of any of their children, no matter what the benefits of the dropout choice of any or both of the children might provide.

On the other hand, there may be some households that choose no dropout option based upon some observable factors like their wherewithal to continue education for both, gender preference, the age difference of the children, opportunity costs, etc at times of financial shock.

Data on dropouts often reports too many "no dropout" cases, leading to apparent inflation in the zero or the reference category of the multinomial choice problem. In our case, the choice of zero option or "no dropout" option can come from heterogeneous sources leading to "zero inflation". As the problem at hand is also a problem of choosing one among four categorical options by a household, Multinomial Logit (MNL) is one of the standard methods for empirical estimation. However, Bagozzi and Marchetti (2017) argue that if "zero inflation" comes from heterogeneous sources and non-recognition of this in a simple Multinomial Logit (MNL) model can bias the estimated effects of variables, leading to faulty inferences.

We examine the role played by financial shocks in high school dropout choices of a household with respect to gender in the Indian context after controlling for a set of explanatory variables. To capture the multiplicity of choice of the household we specify the MNL models and further to account for "zero inflation" or baseline category inflation, we specify the BIMNL model developed by Bagozzi and Marchetti (2017).

If we assume that a household chooses the alternative outcome with the highest level of utility among U^0, U^1, U^2, *and* U^3, then we can specify an MNL model to determine the probability that household i (i=1, 2,…, N) with characteristics vector $\boldsymbol{x'}_{\boldsymbol{i}}$ would choose an outcome Y_i that can take any of j+1 discrete unordered values of 0,1,…J such that:

$$Pr(Y_i = j) = \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta}_{\boldsymbol{j}})}{\sum_{j=0}^{J}\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta}_{\boldsymbol{j}})} \qquad (1)$$

where $\boldsymbol{\beta}_{\boldsymbol{j}}$ is the vector of parameters. A normalization $\boldsymbol{\beta}_{\mathbf{0}} = \mathbf{0}$ for choice category j=0 or baseline category to zero ensures that probabilities across J+1 choice categories for an observation i sum to one, under the assumption that the corresponding J+1 error terms for underlying utilities of J+1 choices are independent and identically distributed with Gumbel (Type 1 extreme value) distribution.

As specified by Bagozzi(2016), the BIMNL estimation consists of two stages: an inflation stage logistic equation for a latent binary non-inflation indicator s_i and a latent outcome stage

MNL equation for $\widetilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ with discrete unordered values of 0,1,…J given i (i=1,2,…, N). Bagozzi (2016) further elaborates that the observed outcome $Y_i = \widetilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ X s_i implies that the baseline outcome Y_i=0 can occur when s_i=0 or when s_i=1 and $\widetilde{\boldsymbol{Y}}_{\boldsymbol{i}}$ =0 and the baseline inflated MNL distribution arises as a mixture of a degenerate distribution in the baseline category and the assumed distribution of the polytomous variable $\widetilde{\boldsymbol{Y}}_{\boldsymbol{i}}$:

$$\mathbf{Pr}\ (\mathbf{Yi} = \mathbf{j}) = \begin{cases} \mathbf{Pr}(\boldsymbol{s_i} = \mathbf{0}|\boldsymbol{z_i}) + \mathbf{Pr}(\boldsymbol{s_i} = \mathbf{1}|\boldsymbol{z_i})\mathbf{Pr}\left(\widetilde{\mathbf{Y}}_{\boldsymbol{i}} = \mathbf{0}\middle|\boldsymbol{x_i}, \mathbf{s}_{\boldsymbol{i}} = \mathbf{1}\right) & \mathbf{for\ j = 0} \\ \mathbf{Pr}(\boldsymbol{s_i} = \mathbf{1}|\boldsymbol{z_i})\ \mathbf{Pr}(\widetilde{\mathbf{Y}}_{\boldsymbol{i}} = \mathbf{J}|\boldsymbol{x_i}, \mathbf{s}_{\boldsymbol{i}} = \mathbf{1}) & \mathbf{for\ j = 1, 2, .., J} \end{cases} \quad (2)$$

Where z_i and x_i are inflation and outcome stage covariates, respectively (Bagozzi, 2016).

We use two waves of the nationally representative, multi-topic, pan-India IHDS, which is a joint project between the University of Maryland (USA) and the National Council of Applied Economic Research (India). Wave 1 of IHDS or IHDS-I (Desai et al., 2005) covered over 41,000 households in 1,503 villages and 971 urban blocks throughout India, and Wave 2 or IHDS-II (Desai et al., 2012) resurveyed 83% of original households[4]. The Wave-1 fieldwork was carried out from September 2004 to August 2005, and Wave-2 from September 2011 to August 2012.

We filter resurveyed households that have only one girl and one boy—who are siblings—at school with less than 12 years of completed education in IHDS-I, but more than seven completed years of education in IHDS-II.[5] Additionally, we specify a maximum age difference of three years between them so that the external environments faced by the two during their school years are not too different and also to restrict and control for the difference in the opportunity cost of a child's time.In this way, this study becomes more like a controlled experiment . We acknowledge that there can be some unobserved factors related to motivation, the ability of the child, and similar. We assume that households invest more in children with greater ability and thus use the natural logarithm of total expenditure on the education of the girl and the boy in IHDS-I as separate proxies for their individual ability in the MNL stage of the BIMNL models.

IHDS-II survey enquires about whether a household has had to incur large financial loss or expenditure due to one or more of the following seven reasons: Major illness, accidents, drought, flood, fire, loss of job, marriage, crop failure, death, and other losses. We examine the association of high school dropout choices of a household with a dummy variable, which takes a value of 1 if a household has experienced any one of the seven categories of losses.

In an alternate specification, we also examine the association of high school dropout choices with the following four types of losses: Major illness or accidents, marriage, crop failure, and death, which are more frequent in our sample.

IHDS does not provide information on the timing of the shock. We assume that the shocks are randomly distributed in the 6–7 years and if we find a significant relationship with this data, this indicates that with more information, the effect of shocks on gender choice in dropout can be more pronounced.

[4]See https://www.ihds.umd.edu/

[5] These conditions ensure that both the girl and the boy had the opportunity to enroll in high school.

At the inflation stage, we control for covariates such as the logarithm of per-capita consumption in IHDS-I, education of mother and father (in completed years) separately, father's age (in years), a social capital index of the household[6], and whether the household resides in a rural area. For the multinomial stage, we consider factors like whether the household has suffered a large financial loss, the difference between the ages of the girl and the boy, the log of expenditure on education of the girl and the boy, the father's age, log of per-capita consumption and the mother's and the father's education in the multinomial stage.

We compare the performance of MNL and BIMNL models by computing standard measures like AIC or BIC. We also compute APE for both MNL and BIMNL models using and extending algorithms and R codes[7], as given by Bagozzi and Marchetti (2017)[8] and Bagozzi (2016)[9]. While Bagozzi and Marchetti (2017) and Bagozzi (2016) compute partial effect at means, we develop R code for computing APE[10]. We draw a random sample of 1,000 from the multivariate normal population with respective maximum likelihood estimates of MNL and BIMNL model coefficients as means with Cholesky decomposition of variance-covariance matrices as per Krinsky and Robb (1986)—also suggested by Greene (2018). The sample average of the partial effects yields APE for each draw. The empirical variance of these 1,000 observations is used to compute the statistical significance of APE estimates for each model.

We report APEs of both MNL and BIMNL methods in Tables 2 and 3 for two alternate loss specifications. Table 2 shows the aggregate loss dummy indicating whether the household experienced any one of the seven categories of losses as an independent variable. In Table 3 models, we use dummy variables for the following four types of losses: Major illness or accidents, marriage, crop failure, and death. The first four columns in Tables 2 and 3 show the APEs of the MNL model and the next four show the APEs of the BIMNL model respectively.

3. Data and Descriptive Statistics

In Table 1 we present a summary of the descriptive statistics of the analytical sample used in this study based on the IHDS-I and IHDS-II datasets. As in this study, we are comparing two estimation methods, we are reporting descriptive statistics without survey weights. The sample contains 1,018 households that made one among four choices in the interim period of the two surveys: dropout of neither (759; 74.10%), drop out of the girl (101; 9.92%), drop out of the boy (95; 9.33%), and dropout of both (63; 6.19%). We observe that 57.1% of the 1018 households have experienced at least one large financial loss between IHDS-I and IHDS-II survey rounds. Among households choosing to drop out a girl, the proportion of experiencing at least one loss is highest (72.3%), followed by households choosing a boy dropout option (62.1%), and among households choosing the no dropout option the proportion is lowest (54%). Table 1 also presents information on specific losses experienced by households. We observe that proportion of households experiencing loss due to major illness and accidents (30.5%) is

---

[6] Following (Mvroniuk, Vanneman and Desai 2017) we also compute a simple index of social capital based on whether the household has contacts in the formal sectors like, education and healthcare sectors and also in government as per 2005 IHDS-I survey data . The index takes a value of 0 to 3.

[8] Bagozzi & Marchetti(2017) provided code and replication files at url https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/3JR2YL last accessed, 28 May 2020.

[9] We use standard formulas for computing APE of a covariate as per the MNL model. While for computing APE of a covariate for BIMNL model, we use formulas provided by Bagozzi (2016), given in Appendix A.7 to A.10.

[10] We write a "margins" method for MNL and BIMNL model objects, which works when we attach the library of the "margins" package (Thomas J. Leeper, 2018, Margins: Marginal Effects for Model Objects. R Package version 0.3.23)

highest among the households choosing to drop out a boy, while the sample average is 24.2%. On the other hand, 44.6% of households that have chosen to drop out a girl have experienced large financial expenditures due to marriage. This proportion is the highest among the four groups and the sample proportion is 28.7%. The sample average of experiencing a crop failure is 14.7% while in households choosing girl dropout and boy dropout options the proportions are 21.8% and 20.0% respectively. The proportion of households experiencing large financial loss due to death is the highest (14.8%) among the group choosing the no dropout option with a sample average of 14.0%. We observe that among households choosing the girl dropout option, girls are almost 1-year senior to the boy and almost the reverse is true for households choosing a boy dropout option. We also observe that the average value of the following variables; Log of total educational expense for a girl and a boy, Log of per capita consumption is household, mother's and father's education in years; are highest in the households choosing the no dropout option compared to other three options and the average value of all these variables are lowest in the group choosing dropping out both children. The average father's age is lowest in the group choosing no dropout option and highest in the group choosing drooping out both children. The social capital index of the household is also highest among the households choosing no dropout options. The proportion of households from rural areas is highest among households choosing to drop out the girl.

Table 1: Descriptive Statistics

	Overall				Stratified on Dropout Choice of Households			
					No Dropout	Girl Dropout	Boy Dropout	All Dropout
Observations	1018				759	101	95	63
Variables*	count(%)/ mean(SD)	Pctl(25)	Pctl(50)	Pctl(75)	count(%)/ mean(SD)	count(%)/ mean(SD)	count(%)/ mean(SD)	count(%)/ mean(SD)
Whether household experienced large financial loss/expenditure (Dummy) =1*	581 (57.1)	--	--	--	410 (54.0)	73 (72.3)	59 (62.1)	39 (61.9)
Large financial loss/expenditure: Major illness / Accidents (Dummy) =1*	246 (24.2)	--	--	--	181 (23.8)	23 (22.8)	29 (30.5)	13 (20.6)
Large financial loss/expenditure: Marriage- (Dummy)=1*	292 (28.7)	--	--	--	181 (23.8)	45 (44.6)	35 (36.8)	31 (49.2)
Large financial loss/expenditure: Crop Failure- (Dummy)=1*	150 (14.7)	--	--	--	97 (12.8)	22 (21.8)	19 (20.0)	12 (19.0)
Large financial loss/expenditure: Death - (Dummy)=1*	143 (14.0)	--	--	--	112 (14.8)	10 (9.9)	13 (13.7)	8 (12.7)
Girl's Age -Boy's Age in Year	0.16 (2.29)	-2	1	2	0.22 (2.28)	1.03 (2.21)	-1.00 (1.98)	-0.19 (2.19)
Log total education expense on Girl IHDS-I	6.79 (1.44)	5.97	6.89	7.82	6.93 (1.50)	6.47 (0.96)	6.48 (1.29)	6.06 (1.26)
Log total education expense on Boy IHDS-I	6.84 (1.52)	5.99	6.95	7.94	6.97 (1.57)	6.39 (1.50)	6.77 (1.06)	6.19 (1.24)
Log of per capita consumption in IHDS-I	6.57 (0.67)	6.14	6.55	7	6.66 (0.69)	6.33 (0.56)	6.46 (0.56)	6.17 (0.49)
Mother's education in years	5.10 (4.53)	0	5	9	5.97 (4.52)	2.30 (3.34)	3.14 (3.84)	2.13 (3.05)
Father's education in years	7.40 (4.50)	4	8	10	8.21 (4.39)	5.05 (3.95)	5.42 (4.11)	4.40 (3.72)

Father's age in years	40.29 (6.75)	35	40	45	39.59 (6.45)	41.32 (7.34)	42.12 (6.82)	44.27 (7.25)
Social Capital Index (numbers 0 to 3)	1.20 (1.15)	0	1	2	1.32 (1.18)	0.94 (1.08)	0.78 (0.95)	0.84 (0.95)
Rural (Dummy) =1*	644 (63.3)	--	--	--	455 (59.9)	79 (78.2)	64 (67.4)	46 (73.0)

Note: Computed by authors from IHDS-I and IHDS-II data set, survey weights have not been used. * For these variables the observations represent sample count for variable value =1 and parenthesis has sample proportion

4. Empirical Results

Households experiencing large financial losses are more likely to opt for a dropout of the girl. Estimated APEs are 4.3 percentage points as per both the MNL and BIMNL models in Table 2. However, the likelihood of choosing the dropout of the boy or the dropout of both is not significant even at the 10 % level. In Table 3, we observe that households experiencing large financial loss/expenditure due to marriage are more likely to choose the girl dropout option by 4.1 and 4.0 percentage points as per MNL and BIMNL models respectively, while no significant association with boy dropout or all dropout options is observed. Moreover, as per the BIMNL model in Table 3, the probability of choosing the girl dropout option increases by 5.3 percentage points when a household has experienced a crop failure. There is no significant estimate of choosing the dropout of a boy or both. The MNL model fails to discover this association. Considering sample estimates of dropout for girls and boys in the 10% range, the above estimated effects are quite substantial. No significant APE estimates are observed for the other two loss variables: Loss/expenditure due to major illness or accidents and large financial loss/ expenditure due to death.

Households with an elder girl are more likely to choose the dropout of the girl as per models in both Tables 2 and 3. Both models in Tables 2 and 3 indicate that households at higher consumption levels are less likely to choose dropout of both. Households with a higher education level of the mother are less likely to choose to drop out the girl; 6.7/6.5 and 5.8/5.8 percentage points less for MNL and BIMNL models respectively in Table 2/ Table 3 respectively. Households with higher education of the father are less likely to choose to drop out of the boy, though the APE estimate is almost half of that of the mother's education increment for girls. Households with a better-educated mother and father are also more likely to choose the no dropout option, while the estimated APEs of choosing the no dropout option associated with the mother's education are 7.7 to 11.3 percentage points in Tables 2 and 3. The estimates are 1.2 to 2,3 times more than that of increment due to a higher father's education.

With seniority in the father's age, households are more likely to choose dropout of both students and dropout of the boy as per both MNL and BIMNL models in Tables 2 and 3 and the estimates of BIMNL models are higher than that of MNL models for both the options in Tables 2 and 3. Therefore, with the increased age of the father, the choice of dropout shifts towards the boy and dropping out both, and the BIMNL model better extracts the association. As per MNL models in both Tables 2 and 3, with one more acquaintance in a social network, the likelihood of choosing no dropout increases, and the likelihood of choosing boy dropout decreases. BIMNL estimates of choosing no dropout options are close to that of MNL models in Tables 2 and 3 while BIMNL estimates that social network is more evenly associated with a lesser likelihood of choosing boy and girl dropout options. The rural dummy was not significant in either model in both tables. The BIMNL model outperforms the MNL model as per AIC and BIC criteria in both Tables 2 and 3, while in Table 2 BIMNL model is the best among the four models as per AIC and BIC criteria.

Table 2: Estimated Average Partial Effects (APE) of MNL and BIMNL Models of High School Dropout With Aggregate Loss Dummy

Dependent Variable: Dropout Choice of Households -- None, Girl, Boy, Both

	MNL Model APE of Dropout of				BIMNL Model Global APE of Dropout of			
	None	Girl	Boy	Both	None	Girl	Boy	Both
Whether Household experienced large financial Loss /Expenditure- (Dummy), 0-> 1[a]	-0.031 (0.025)	0.043 (0.018)**	0 (0.019)	-0.012 (0.016)	-0.032 (0.022)	0.043 (0.017)**	-0.002 (0.016)	-0.009 (0.014)
Girl's Age -Boy's Age in Year, 1->2[b]	-0.004 (0.006)	0.022 (0.006)***	-0.014 (0.002)***	-0.004 (0.003)	0.002 (0.005)	0.018 (0.005)***	-0.015 (0.003)***	-0.005 (0.003)**
Log total education expence on Girl IHDS-I, 5.97->6.89[c]	0.02 (0.015)	0.002 (0.011)	-0.012 (0.011)	-0.01 (0.008)	0.017 (0.014)	0.002 (0.01)	-0.01 (0.011)	-0.009 (0.008)
Log total education expence on Boy IHDS-I, 5.99->6.95[d]	-0.01 (0.014)	-0.001 (0.01)	0.012 (0.011)	-0.001 (0.008)	-0.013 (0.01)	0.002 (0.008)	0.011 (0.01)	0 (0.007)
Log of per capita consumption in IHDS-I, 6.14->6.55[e]	0.015 (0.01)	-0.006 (0.007)	0.003 (0.008)	-0.013 (0.005)**	0.013 (0.009)	-0.004 (0.008)	0.005 (0.008)	-0.014 (0.006)**
Mother's education, 0->5[f]	0.113 (0.025)***	-0.067 (0.02)***	-0.019 (0.016)	-0.027 (0.015)*	0.085 (0.029)***	-0.058 (0.019)***	-0.009 (0.017)	-0.018 (0.014)
Father's education, 4->8[g]	0.049 (0.016)***	-0.015 (0.01)	-0.021 (0.011)**	-0.013 (0.009)	0.06 (0.021)***	-0.02 (0.012)	-0.028 (0.014)**	-0.012 (0.009)
Father's Age, 40->45[h]	-0.053 (0.01)***	0.008 (0.007)	0.019 (0.007)***	0.026 (0.006)***	-0.059 (0.011)***	0.008 (0.007)	0.023 (0.007)***	0.027 (0.006)***
Social Network Index, 0->1[i]	0.036 (0.014)**	-0.006 (0.009)	-0.027 (0.011)**	-0.003 (0.008)	0.035 (0.013)***	-0.014 (0.005)***	-0.013 (0.005)***	-0.008 (0.003)***
Rural (Dummy) , 0->1[j]	-0.019 (0.028)	0.022 (0.021)	0 (0.021)	-0.003 (0.018)	-0.031 (0.027)	0.012 (0.011)	0.012 (0.01)	0.007 (0.006)
Number of Observations	1018				1018			
AIC	1522.88				1506.13			
BIC	1685.42				1668.68			

Notes: Standard Errors in (parentheses); Significance Codes: '***' 0.01 '**' 0.05 '*' 0.1; We use following levels of the independent variables to compute Average Partial effect (APE) a: Whether household experienced large financial Loss/Expenditure (Dummy) at 1 yes and 0 No; b: Girl's Age -Boy's Age in Year at 75th percentile value of 2 and median value 1,c:Log total education expense on Girl IHDS-I, at the median value 6.89 and 25th percentile value 5.97 , d:Log total education expense on Boy IHDS-I, at the median value 6.95 and 25th percentile value 5.99, e: Log of per-capita consumption in IHDS-I at the median value 6.55 and 25th percentile value 6.14, f: Mother's education at median value of 5 years of education and 25th percentile value of 0 year of education, g: Father's education at median value of 8 years of education and 25th percentile value of 4 years of education, h: Father's Age at 75 percentile value of 45 years and median value of 40 years, i: Social Network Index at median value of 1 and 25th percentile value of 0, j: Rural (Dummy) at 1 Yes and 0 No

Table 3: Estimated Average Partial Effects (APE) Of MNL And BIMNL Models of High School Dropout With Four Specific Losses

Dependent Variable: Dropout Choice of Households -- None, Girl, Boy, Both								
	MNL Model APE of Dropout of				BIMNL Model Global APE of Dropout of			
	None	Girl	Boy	Both	None	Girl	Boy	Both
large financial Loss/Expenditure-Major illness / Accidents MI1 (Dummy), 0-> 1[a]	0.012 (0.029)	-0.018 (0.021)	0.025 (0.021)	-0.019 (0.015)	-0.005 (0.021)	-0.012 (0.017)	0.029 (0.018)	-0.012 (0.014)
large financial Loss/Expenditure-Marriage- MI4 (Dummy), 0-> 1[a]	-0.07 (0.029)**	0.041 (0.022)*	0.005 (0.021)	0.024 (0.017)	-0.064 (0.029)**	0.04 (0.02)**	0.002 (0.019)	0.022 (0.015)
large financial Loss/Expenditure-Crop Failure- MI5 (Dummy), 0-> 1[a]	-0.041 (0.037)	0.021 (0.026)	0.018 (0.027)	0.002 (0.021)	-0.066 (0.037)*	0.053 (0.027)*	0.016 (0.025)	-0.003 (0.018)
large financial Loss/Expenditure-Death - MI6 (Dummy), 0-> 1[a]	0.032 (0.036)	-0.03 (0.024)	-0.001 (0.026)	-0.001 (0.021)	0.013 (0.026)	-0.028 (0.02)	0.007 (0.024)	0.008 (0.02)
Girl's Age -Boy's Age in Year, 1->2[b]	-0.003 (0.006)	0.021 (0.006)***	-0.014 (0.002)***	-0.004 (0.003)	0.003 (0.005)	0.018 (0.005)***	-0.015 (0.003)***	-0.005 (0.003)*
Log total education expense on Girl IHDS-I, 5.97->6.89[c]	0.018 (0.015)	0.002 (0.011)	-0.011 (0.011)	-0.009 (0.008)	0.015 (0.012)	0.006 (0.01)	-0.011 (0.011)	-0.009 (0.008)
Log total education expense on Boy IHDS-I, 5.99->6.95[d]	-0.009 (0.014)	-0.002 (0.01)	0.012 (0.011)	-0.001 (0.008)	-0.012 (0.008)	-0.001 (0.008)	0.012 (0.009)	0.001 (0.007)
Log of per capita consumption in IHDS-I,6.14->6.55[e]	0.015 (0.01)	-0.005 (0.008)	0.002 (0.007)	-0.013 (0.005)**	0.011 (0.008)	-0.003 (0.007)	0.006 (0.008)	-0.014 (0.006)**
Mother's education,0->5[f]	0.104 (0.026)***	-0.065 (0.02)***	-0.017 (0.017)	-0.022 (0.015)	0.077 (0.028)***	-0.058 (0.02)***	-0.005 (0.016)	-0.014 (0.013)
Father's education,4->8[g]	0.05 (0.016)***	-0.015 (0.01)	-0.022 (0.011)**	-0.013 (0.008)	0.062 (0.02)***	-0.019 (0.012)	-0.031 (0.013)**	-0.013 (0.008)
Father's Age,40->45[h]	-0.048 (0.011)***	0.006 (0.007)	0.019 (0.007)**	0.023 (0.006)***	-0.055 (0.011)***	0.006 (0.007)	0.025 (0.008)***	0.025 (0.006)***
Social Network Index,0->1[i]	0.037 (0.013)***	-0.007 (0.01)	-0.027 (0.011)**	-0.004 (0.008)	0.038 (0.012)***	-0.015 (0.005)***	-0.015 (0.005)***	-0.009 (0.003)***
Rural (Dummy) ,0->1[j]	-0.016 (0.031)	0.025 (0.022)	-0.004 (0.021)	-0.005 (0.019)	-0.024 (0.027)	0.01 (0.01)	0.009 (0.01)	0.006 (0.006)
Number of Observations	1018				1018			
AIC	1533.15				1507.53			
BIC	1740.03				1714.41			

Notes: Standard Errors in (parentheses); Significance Codes: '***' 0.01 '**' 0.05 '*' 0.1; We use following levels of the independent variables to compute Average Partial effect (APE) a: Whether household experienced large financial Loss/Expenditure- due to reason specified (Dummy) at 1 Yes and 0 No; b: Girl's Age -Boy's Age in Year at 75th percentile value of 2 and median value 1,c:Log total education expense on Girl IHDS-I, at the median value 6.89 and 25th percentile value 5.97 , d:Log total education expense on Boy IHDS-I, at the median value 6.95 and 25th percentile value 5.99, e: Log of per-capita consumption in IHDS-I at the median value 6.55 and 25th percentile value 6.14, f: Mother's education at median value of 5 years of education and 25th percentile value of 0 year of education, g: Father's education at median value of 8 years of education and 25th percentile value of 4 years of education, h: Father's Age at 75 percentile value of 45 years and median value of 40 years, i: Social Network Index at median value of 1 and 25th percentile value of 0, j: Rural (Dummy) at 1 Yes and 0 No

5. Conclusion

As far as gender discrimination in dropouts is concerned, the choice sets of different households in a sample could be different. We use a filtered dataset, where all households have exactly one boy and one girl. To capture the multiplicity of choices of such households, we specify models of multinomial discrete choice family. We also observe that households opting for "no dropout" for both the boy and the girl could be heterogeneous due to the possible presence of households that will not discriminate and will also not drop out any child irrespective of gender even after an adverse financial shock due to their perception of the value of education and some other households that will choose the same option after evaluating a tradeoff. To capture this effect, we specify the BIMNL model.

Our results indicate that in households that experience large financial loss, a girl faces a significantly higher chance of dropout compared to a boy. More specifically, households that have experienced large financial loss/expenditure due to marriage or crop failure are more likely to choose to drop out of the girl from high school education. BIMNL model is able to capture the association of crop failure with the dropout of the girl while the MNL cannot. We found that elder girls face a significantly higher chance of dropout. Interestingly, the more the father's age, the more is the likelihood of dropout for a boy or dropout of both, and BIMNL model estimates are higher than MNL estimates. Mother's education significantly reduces the probability of dropout of girls and increases the likelihood of choosing no dropout. Further, networked households experience lower dropout and BIMNL estimates that social network is more evenly associated with a lesser likelihood of choosing boy and girl dropout options.

Large financial loss due to marriage in the household reduces wherewithal of households to invest in education and there is more propensity of households choosing dropout of a girl in such a situation. This is similar to an observation made by Boyle et al. (2002) that an economic slowdown that gradually constrains cash income may result in girls being withdrawn from school occasionally or permanently. In this context, our finding that sudden shocks (e.g. crop failures) lead to a higher probability of dropout of girls is contrary to the earlier findings. For example, Boyle et al. (2002) find that a sudden shock such as the death of a key income earner may result in the permanent withdrawal of a boy to take up an income-earning role. However, we find that our approach of the recognition of the multiplicity of options available to households is important here. If we recognize that households have options of dropping either or both a girl and a boy, then these differences in our findings with earlier studies can be reconciled. A limitation in our study is that the results may not be interpreted in the causal sense. Our results are however associative in nature. Future studies may focus on the casual aspect in more detail.

Econometrically, we find that BIMNL models perform better than MNL models and BIMNL estimates are different for some critical variables in our study. As BIMNL models recognize the multiplicity of choice and can capture heterogeneity in the no dropout choice effectively, their applications on appropriately filtered datasets may be a good approach to study gender discrimination in dropout. In the current context, as many households are currently suffering from a financial shock due to Covid-19 in different parts of the world, it would be interesting to examine whether girls are especially at higher risk of dropout with these more sophisticated econometric tools.

Acknowledgement

We thank the IHDS team for making their dataset publicly available, and we also thank Benjamin E. Bagozzi and Kathleen Marchetti for making publicly available the R code for

Baseline Inflationary Multinomial Logit (BIMNL) model, which we have used in this study. Authors bear full responsibility for any remaining errors.

References

Bagozzi, Benjamin E. 2016. "The baseline-inflated multinomial logit model for international relations research." Conflict Management and Peace Science 33 (2): 174-197.

Bagozzi, Benjamin E., and Kathleen Marchetti. 2017. "Distinguishing Occasional Abstention from Routine Indifference in Models of Vote Choice." Political Science Research and Methods 5 (2): 277-249.

Banerjee, Abhijit V., and Esther Duflo. 2011. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. New York: Perseus Books.

Boyle, S., A. Brock, J. Mace, and M. Sibbons. 2002. Reaching the Poor: The 'Costs' of Sending Children to School. Synthesis Report, London: DFID.

Desai, Sonalde, Reeve Vanneman, and New Delhi. and National Council of Applied Economic Research. 2005. India Human Development Survey (IHDS). Vols. ICPSR22626-v8. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-06-29. . doi:http://doi.org/10.3886/ICPSR22626.v8

Desai, Sonalde, Reeve Vanneman, and New Delhi. and National Council of Applied Economic Research. 2011-12. India Human Development Survey-II (IHDS-II). Vols. ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-07-31. doi:http://doi.org/10.3886/ICPSR36151.v2

Dreze, Jean, and Geeta Gandhi Kingdon. 2001. "School Participation in Rural India." Review of Development Economics vol. 5,issue 1,1-24.

Gandjour, Afschin. 2007. "Mutual dependency between capabilities and functionings in Amartya Sen's capability approach." Social Choice and Welfare 31 (2): 345-350. doi:10.1007/s00355-007-0283-7

Greene, William H. 2018. Econometric Analysis, 8th Edition, Indian Subcontinent Reprint. Noida, Uttar Pradesh: Pearson India Education Services Pvt. Ltd.

Heltberg, Rasmus, Naomi Hossain, Anna Reva, and Carolyn Turk. 2013. "Coping and Resilience during the Food, Fuel, and Financial Crises." The Journal of Development Studies 49 (5): 705-718. doi:10.1080/00220388.2012.746668

Hunt, Frances. 2008. Dropping Out from School: A Cross Country Review of the Literature. CREATE PATHWAYS TO ACCESS Research Monograph No 16, University of Sussex. https://files.eric.ed.gov/fulltext/ED504047.pdf

Kearney, Melissa S., and Phillip B. Levine. 2016. "Income Inequality, Social Mobility, and the Decision to Drop Out of High School." Brookings Papers on Economic Activity. 333-396.

Krinsky, I., and L. Robb. 1986. "On Approximating the Statistical Properties of Elasticities." Review of Economics and Statistics 68(4), pp. 715-719.

Momo, Michelle S. M., Sofie J. Cabus, Kirstof De Witte, and Wim Groot. 2019. "A systematic review of literature on the causes of early school living in Africa and Asia." Review of Education 7 (3): 496-522. doi:https://doi.org/10.1002/rev3.3134

Murnane, Richard J. 2013. "U.S. High School Graduation Rates: Patterns and Explanations." Journal of Economic Literature 51(2),370-422,http://dx.doi.org/10.1257/jel.51.2.370

Mvroniuk, Tyler W., Reeve Vanneman, and Sonalde Desai. 2017. "Getting a Child through Secondary School and to College in India: The Role of Household Social Capital." Sociology of Development Vol. 3 No.1, Spring 2017;(pp. 24-46).

Siddhu, G. 2011. "Who makes it to the secondary school? Determinants of transition to secondary schools in rural India." International Journal of Educational Development 31 (4): 394-401. https://doi.org/10.1016/j.jedudev.2011.01.008

Sowmya, Dhanaraj, Christy Mariya Paul, and Smit Gade. 2019. "Household income dynamics and investment in children: Evidence from India." Education Economics 27 (5): 507-520. doi:10.1080/09645292.2019.1599325

Contact emails: efpm05011@iiml.ac.in
kbhattacharya@iiml.ac.in