Droughts and Gender Heterogeneity in High School Dropouts in Marginal and Small Agricultural Households in India: Discrete Time Survival Analysis

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 – We examine whether droughts are differently associated with high school dropout of girls and boys from marginal and small agricultural households compared to other households. While non-agricultural households may adopt a utility maximisation framework regarding educational investments, marginal or small agricultural households have to make an additional profit maximisation decision regarding production and choice of inputs in their farm including differential deployment of female or male family labour in and outside their farms.

Methodology – With discrete-time survival analysis, we examine the association of droughts with the hazards of high school dropout by combining publicly available data from two rounds of the India Human Development Survey (2004-05 and 2011-12) with ICRISAT district-level rainfall data for India.

Findings – We find that the hazard of dropout of girls from high school significantly reduces in drought years in marginal and small agricultural households while boys in marginal agricultural households face a significantly higher hazard of high school dropout. We observe that the hazard of dropout of girls in marginal agricultural households increases significantly if there was a drought in the previous academic year.

Practical implications –Effective policy interventions are needed to provide high school education to all in the face of climate change and increasing drought frequency across India.

Originality – We uniquely contribute by combining data from different publicly available datasets and deploying discrete-time survival analysis to bring out the heterogenous relationship between drought and high school dropout of boys and girls in marginal and small agricultural households.

Keywords: Gender Heterogeneity, High School Dropout, Droughts, Discrete-Time Survival Analysis, Marginal and Small Agricultural Household

Droughts and Gender Heterogeneity in High School Dropouts in Marginal and Small Agricultural Households in India: Discrete Time Survival Analysis

Section-I: Introduction

Literature exploring the relationship of covariate shocks[1] with the educational outcome of children in the household has found that while households show a procyclical behaviour regarding covariate shocks in low-income countries by reducing investment in education and schooling, in the middle-income countries the evidence becomes more ambiguous and in high income developed countries households response to covariate shock become more countercyclical (Ferreira, Santos, Fonseca, & Haase, 2007), (Zimmermann, 2020). Unlike other covariate shocks, rainfall shocks can result in agricultural shocks, and the impacts of rainfall shocks on income, consumption, leisure, and educational investment of agricultural households can therefore be different from non-agricultural households.

We examine whether negative rainfall shocks or drought[2] are associated with high school dropout of girls and boys from marginal and small agricultural households[3] differently than other households. While non-agricultural households may adopt a utility maximisation or cost-benefit framework (Dreze & Kingdon, 2001), (Kearney & Levine, 2016), regarding educational investments of children; marginal or small agricultural households must make an additional profit maximization decision regarding production and choice of inputs in their farm including differential deployment of female or male family labour in and outside their farms.

Education costs involve direct expenditures on school fees, books, and stationery. It also includes the opportunity cost of the time spent by the household members. Perceived benefits of education, subject to the information available to the household (Fiszbein, et al., 2009), may include higher returns through employment and a few other intangible benefits. As per Agricultural Household Model (AHM) (Singh, Squire, Strauss, & [Editors], 1986), production decisions of agricultural households determine farm profits, which are a component of household income and thus influences consumption and labour supply decisions. Further, as production and consumption decisions of small-scale agricultural households are interconnected, and most of these households produce partly for sale and partly for their consumption, they also purchase some of their inputs, like fertiliser and partly provide some, such as family labour, from their resources. Further, In the case of an imperfect labour market, an agricultural household may adopt greater self-sufficiency by

[1] Income shocks due to natural disasters, changes in food prices, economic crises etc are called covariate income shocks as they affect all households in a region and Shocks like illness, injury or death of family members, job or business loss, theft or destruction of property are called idiosyncratic income shocks as these shocks affect individual households. (Sowmya, Paul, & Gade, 2019).

[2] In India, around 68% of the country is prone to drought in varying degrees and 35% which receives rainfall between 750 mm and 1125 mm is considered drought prone while 33% receiving less than 750 mm is chronically drought prone. Therefore, we have considered Meteorological drought (If annual rain fall is less than 75 % of long-term average for the district) in a district as per definition of Government of India Ministry of Jal Shakti as an event of negative rainfall shock . (Source A BRIEF ON DROUGHT, http://jalshakti-dowr.gov.in/brief-drought accessed on 17.08.2021)

[3] We categorise Farmers with below 1.00 Hectare landholding as Marginal Farmer and farmers 1.0-2.0 Hectare landholding as Small farmers as per Government of India Ministry of Agriculture & Farmers Welfare "Categorisation of Farmers" published by Press Information Bureau on 5th February 2019, https://pib.gov.in/newsite/PrintRelease.aspx?relid=188051 accessed on 21.11.2021

deploying family labour as an optimal strategy ([LaFave & Thomas, 2016], [Taylor & Adelman, 2003]). Due to resource constraint, we assume that a greater deployment of family labour in own farms for marginal and small agricultural households compared to relatively larger farmers with more than 2 hectares of landholding. Moreover, there can be gender disparity in the choice of deploying family labour on the own farm or participation in wage labour outside of own farm and investment of leisure time in education in the agricultural household for the students. At times of rainfall shocks, choices of agricultural households concerning the education/ dropout of high school going adolescents in the household would depend on the above concerns.

Association of women in agriculture is an age-old practice and majority of Indian rural women belong to small and marginal farmers and landless agricultural labourer's families. In general weeding and harvesting are predominantly female activities and women are overrepresented in unpaid, seasonal, and part-time work. Also, women are often paid less than men, for the same work. In view of the above it is more likely that adolescent girls going to high school would be deployed on their farms by marginal and small farming households. Experiencing reduced income and food insecurity during drought periods, marginal and small farming households may deploy adolescent high school boys to work for wages—adhering to prevailing gender roles and considering higher wage male labour. In a study on Madagascar, (Marchetta, Sahn, & Tiberti, 2019) the author also observed lagged effects of shock, indicating some persistence of the impact in the year following the shock itself.

Unlike other covariate shocks, we propose that drought or lesser rainfall shocks are more like agricultural shocks and have greater and varied impacts on the high school education of girls and boys in the marginal and small farming households among the households in a pan-India context. We hypothesise that expecting a lesser farm income and food insecurity due to drought in current year, marginal and small farming households would be compelled to take decisions to increase family income by available means. In such situations, adhering to prevailing gender roles and considering generally higher wage rates for male labour, marginal and small farming households may deploy adolescent high school boys to work for wages. This behaviour may result in higher high school dropout possibilities for boys. A drought in the academic year would reduce the requirement of farm labour and depress market wage for farm labour and may reduce the requirement of deploying labour by household girls resulting in a lesser propensity of high school dropout for girls than boys. Further, in the year following a drought year, such households may face capital constraints due to lesser farm income in the previous year and be compelled to provide for greater labour inputs in their farm to substitute capital by deploying high school going girls and thus a drought's lagged effect can reduce girls' propensity to continue high school education in small and marginal farming households. Moreover, if a drought year is followed by a normal rainfall year, facing higher labour requirements and /or higher market wages of farm labour, marginal and small farming households may increase the deployment of labour of household girls on the farm. Further, (Fulford, 2014) observed that only 14% of women and 41% of men work for wages in India. Therefore, facing a resource crunch after a drought year, the marginal and small agricultural household might see lesser benefits in educating a girl as a girl would be less likely to participate in wage work outside the farm/household and would also leave her parent's household after marriage. Thus, in the year following a drought year, the marginal and small farming households may remain invested in the high school education of boys as a diversification strategy & future insurance.

There are several international studies examining relationships between rainfall variability and educational outcome of students (some recent examples: [Colmer, 2020] for Ethiopia, [Marchetta, Sahn, & Tiberti, 2019] for Madagascar). However, such studies in the Indian context and particularly for high school level education are rare with notable exception of (Zimmermann, 2020). Zimmermann (2020) using India NSSO data found that a negative rainfall shock was associated with increased school enrollment in 2007 for 11-18 years old children, whereas the estimates were of the opposite sign in 1986 and enrollment was increasingly falling after positive rainfall shocks. Zimmermann also found that the effect was stronger for girls than for boys, more pronounced for older children. However, Zimmermann did not examine whether the outcomes are different in marginal and small agricultural households.

To test our propositions, we examine the association of droughts with the hazards of high school dropout in an academic year by combining publicly available data from two rounds of the India Human Development Survey (2004-05 and 2011-12), (Desai, Vanneman, & and National Council of Applied Economic Research, India Human Development Survey (IHDS)., 2005), (Desai, Vanneman, & and National Council of Applied Economic Research, India Human Development Survey-II (IHDS-II), 2011-12) with ICRISAT district-level rainfall data [4] for India With discrete-time survival analysis with the Gompertz link function. We check the robustness of our result with several alternative model specifications and controls.

The plan of the paper is as follows, section II discusses the analytical framework; Section III describes the data and computes some exploratory statistics. Section IV presents the results. Finally, Section V concludes the paper.

Section-II: Analytical Framework and Empirical Strategy

2.1 Analytical Framework

A non-agricultural household may adopt a utility maximisation or cost-benefit framework (Kearney and Levine, 2016, Dreze and Kingdon, 2001) regarding educational investments where a student is dropped out when 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 $E(V^e) > E(V^d)$.

As per the basic Agricultural Household Model (AHM) developed by Singh, Squire, and Strauss(1986), agricultural households must make an additional profit maximisation decision regarding production and choice of inputs in their farm like fertiliser, farm labour, deployment of family labour in own farm, etc., including differential deployment of female or male family labour in and outside their farms.

[4] http://data.icrisat.org/dld/src/biophysical.html

Following Singh, Squire, and Strauss (1986) and considering our context, the AHM can be mathematically represented as follows:

$$Max\ (U = U(X_a, X_m, X_l)) \qquad (2)$$

For any production cycle, the household is assumed to maximize the above utility function where commodities are; agricultural staple (X_a,), market purchased good (X_m), and leisure (X_l)

Subject to:
Cash Income constraint; pmXm $= p_a\,(Q - X_a) - w_f\,(L - F_f) + w_{of}\,F_{of}$

Where p_m and p_a are the prices of the market-purchased good and the staple, respectively, Q is the household's production of the staple, w_f is the market wage for farm labour and w_{of} is the market wage for off-farm labour, L is total labour input, and F_f is family labour input in own farm and F_{of} is off-farm family labour deployment.

And, in the RHS of the income constraint the term, $p_a\,(Q - X_a) - w_f\,(L - F_f)$ can be denoted as π or farm profits.

Household time constraint; $T = X_l + F_f + F_{of}$

Where T is the total stock of household time

Household agricultural production $Q = Q\,(A, K(Z_{t-1}), L(Z_t, Z_{t-1}), Z_t)$

Where A is the household fixed quantity of landholding, K is aggregated capital deployment for inputs like fertilizer, pesticide, herbicides, farm implements etc.; Drought or negative rainfall shock in the last year Z_{t-1}, Drought, or negative rainfall shock in the production year Z_t.

It can be assumed that any directional change in leisure due to negative rainfall shocks is likely to similarly impact time devoted to education and a positive effect on leisure time would result in lesser dropout propensity while a negative effect would result in greater high school dropout propensity.

To cope with a negative weather shock from lower rainfall in the current year, the marginal and small agricultural households would attempt to maximize farm profits and utility. In such situations, adhering to prevailing gender roles and considering generally lesser wage rates for female labour, marginal and small farming households may deploy adolescent high school-going boys to work for wages to increase family income. This behaviour may result in lesser leisure time for boys including time devoted to education and result in greater high school dropout propensity for boys. On the other hand, a drought in the academic year would reduce the requirement of farm labour and depress market wage for farm labour and may reduce the requirement of deploying labour by household girls resulting in greater leisure time and a lesser propensity of high school dropout for girls.

i.e., as per our proposition, it is expected that for boys in the marginal and small agricultural households $\boldsymbol{(\delta X_l / \delta Z_t)_b < 0}$ and for girls in marginal and small agricultural households the $\boldsymbol{(\delta X_l / \delta Z_t)_g > (\delta X_l / \delta Z_t)_b}$ of boys in such households.

After a drought year, marginal and small agricultural households may be compelled to provide greater labour inputs in their farm and may increase the deployment of adolescent girls in their farms.

For girls in marginal and small agricultural households, we posit that $(\delta X_l / \delta Z_{t-1})_g < 0$

2.2 Empirical Strategy:

Though many of the studies examining dropout use variants of probit or logit analysis – due to the presence of left and right censored data and incomplete information –survival analysis methods are increasingly being used to study dropout of education. (Momo, Cabus, Witte, & Groot, 2019) in a literature review of school dropout studies observe that the longitudinal studies used logistic regression and survival analysis methods such as the discrete-time logit model and the Cox regression model. Table 1 we present some of the recent studies in school dropout where survival analysis methods have been used.

Table 1:Some recent Survival Analysis studies in various countries on school dropout

Study	Country, Sample	Method
(Goel & Zakir, 2018)	India, Cross section Data NSSO,2011-12, 456,976 Individuals, from 101,718 households with ~233 thousand males and ~223 thousand females in Rural & urban areas	Survival analysis with Discrete-time hazard Model: complementary log-log model. Cohort Analysis was deployed to identify temporal trends in gender differences in the educational survival rate
(Liu & Hannum, 2017)	China, Sample size ~ 60 Longitudinal China Health and Nutrition Survey (CHNS)	Cox proportional hazard model of dropping out of school in young adulthood
(Zhang, 2017)	China, publicly available large-scale survey data (RUMIC)	Proportional hazards regressions (or survival analysis) with three different stratifications are conducted to find out the school outcome of migrant children, as compared with their urban and rural cohorts. The dependent variable for the proportional hazard regression is the hazard ratio of schooling duration.
(Valdivieso, 2015)	Peru, Panel data on Peruvian children from Young Lives program. 2,052 children who were born in 2001 and 2002 and 714 children who were born in 1994 and 1995, with 3 interventions, in September 2002, March 2007 and August 2009.	Kaplan-Meier survival curves, Cox Proportional Hazards (PH) model, Adjusted survival curve using Cox PH model and Testing of PH assumption
(Thomas, Singh, & Klopfenstein, November 2015)	USA, Longitudinal Data from Texas USA tracking ~175,000 students for 5 years	Cox proportional hazards model with time-varying effect and shared frailty
(Gelli, et al., 2014)	Mali, Sample size ~ 7500 primary education	Cox regression model with time-invariant covariates
(Tamusuza, December 2011)	Uganda, National Longitudinal surveys 2000 to, 2002/3,2005/6,2009/10 ,~11,600 school-age children and ~4200 households	A discrete-time Cox regression model to identify factors related to children's survival through to the final grade

(Marshall, 2011)	Guatemala, up to four years of information for~ 850 children who were in first grade in 1999.	Uses grade level transitions framework for analysing stratification. The multinomial dependent variable is year-specific and includes passing, failing, and dropping out as opposed to dichotomous indicators for completing each level of schooling.
(Ferreira, Santos, Fonseca, & Haase, 2007)	Portugal, 10-year longitudinal study of 445 participants from age 7 to 17	HierarchicalProportional hazards regression analysis is used to model the probability of the event of dropping out of school.

Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs (Kleinbaum & Klein, 2008). Survival analysis focuses primarily on two central pieces of information, whether a participant suffers the event of interest like death, failure etc. during the study period and the follow-up time for everyone being followed.

Let the random variable T be the survival time in years in high school education and t be the observed value of T. Following (Kleinbaum & Klein, 2008), the survivor function S(t) is the probability that a student does not drop out of high school education less than some specified time t. If we denote F(t) as the cumulative probability of dropping out of a student till year t and the underlying probability density function of T is f(t) then:

$$F(t) = P(T \leq t) = \int_0^t f(u)du \qquad (3)$$

And then the Survivor function:

$$S(t) = P(T > t) = 1 - F(t) \qquad (4)$$

In survival analysis, instantaneous failure rate or risk/hazard of an event like dropout at some time t conditional on whether the student has not dropped out of high school until that time is defined as hazard.

The hazard function h(t) can be defined as:

$$h(t) = \lim_{\delta t \to 0}\{\frac{P(t < T \leq t + \delta t | T > t)}{\delta t}\} \qquad (5)$$

Sir David Cox in 1972 developed a continuous proportional hazard model to estimate the hazard function (Cox, 1972). Discrete survival time data may arise because either the time scale is intrinsically discrete, or survival occurs in continuous time, but spell lengths are observed only in intervals (Jenkins, 2005). In the case of high school dropouts, the spell lengths are observed only in intervals, and discrete-time survival analysis methods can be deployed to analyse the probability or hazard of a student dropout each year.

One way to estimate discrete time survival analysis is by estimating a logistic regression model. The equation for the logistic regression model is as follows:

$$\ln [p/(1-p)] = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta kXk \qquad (6)$$

Where p is the probability of the event occurring, X1, X2, .., Xk are the predictor variables, β0, β1, β2, ..., βk are the regression coefficients, and ln is the natural logarithm.

The logistic regression model can be used to estimate the odds ratio (OR) of the event occurring for each predictor variable. The OR represents the increase or decrease in the odds of the event occurring associated with a one-unit increase in the predictor variable, holding all other variables constant.

However, another way to model this probability is to use the Gompertz distribution as the link function.

$$\ln[-\ln(1-p)] = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta kXk \qquad (7)$$

Where p is the probability or hazard of dropout, X1, X2, ..., Xk are the predictor variables, β0, β1, β2, ..., βk are the regression coefficients, and ln is the natural logarithm.

Fang & van de Schoot (2019) suggests that though both Logistic and Gompertz link functions are suitable for most cases and usually lead to similar parameter estimates where underlying hazards are small; the two links result in different interpretations of the estimated parameters. Fang & van de Schoot (2019) further explains that in a logistic model, the exponential term of a parameter estimate quantifies the difference in the value of the odds per unit difference in the predictor, while in the Gompertz model, it is the value of the hazard (i.e., which is a probability). Therefore, the Gompertz link results in a more intuitive model interpretation.

With discrete-time survival analysis with the Gompertz link function, we examine the association of droughts with the hazards of high school dropout in an academic year by combining publicly available data from two rounds of the India Human Development Survey (2004-05 and 2011-12) with ICRISAT district-level rainfall data for India. The two rounds of the India Human Development Survey (IHDS), IHDS-I conducted in 2004-05 consist of 41,554 households and 215,754 individuals and IHDS-II, conducted in 2011-12 resurveyed 83% of original households and 150,988 individuals.

We construct an analytical sample from these individuals who were a student in IHDS-I but did not enter high school and who were surveyed in IHD-II and completed at least 8 years of education. In our case the “Event” of interest is Dropout from High school, and the “Time” is in years from passing out of 8^{th} grade or from the start of the study period till the student drops out or passes out of high school or the study period ends. In this case, an observation is right censored if a student passes out of high school i.e., complete 12 years of education or the student is still studying in high school while the study period has ended. On the other hand, an observation is left censored if at the time of the start of the study period the student had already enrolled in high school i.e., had completed education for more than 8 (eight) years and less than 12 years. We avoid left censored data by limiting the completed education of students in round 1 to less than 9 years, i.e., we followed all students in our study from entry into high school i.e., from 8 years of completed education. Though, we have incomplete observations for the right-censored data where the students do not complete high school education during the study period.

We convert the above sample in a person-year format drawing out one record for each year of high school attendance of an individual and combining the person-year data with ICRISAT district-level rainfall data. Our person-year data contains the information that whether a student has dropped out in a particular year and the information on whether there was a drought in that year or in the previous year along with household-specific and individual-specific information. An individual is considered a high-school dropout in a person-year if she/he was not a student as per IHDS-II and whose completed years of education as per IHDS-II is one grade below that grade. We define negative rainfall shocks or drought if the rainfall deficiency in a corresponding academic year is more than 25% of long-term average of the district by broadly following the Meteorological drought definition.

We consider households with less than 1 hectare of agricultural land as marginal agricultural households and households having 1 to 2 hectares of agricultural land as small agricultural households as per the definition of the Government of India and the categorisation is placed in Table 2.

Table 2:Categorisation of farmers

Sl. No.	Category	Size-Class
1.	Marginal	Below 1.00 hectare
2.	Small	1.00-2.00 hectare
3.	Semi- Medium	2.00-4.00 hectare
4.	Medium	4.00-10.00 hectare
5.	Large	10.00 hectare and above
Source: Press Information Bureau, Government of India, Ministry of Agriculture & Farmers Welfare		

Our sample with both urban and rural residences is having 36,388 person-year observations for 13,704 students and our rural subsample consists of 24,215 person-year observations for 9,383 students.

From the person year dataset, we estimate discrete-time hazard models with the Gompertz link function, and our variables of interest are whether the student belongs to a small or marginal agricultural household, the gender of the student and whether the household faced drought in the academic year or its previous year and interactions of the above terms.

For robustness, we estimate two analytical samples with rural and both rural and urban residents and we control for year of high school, some student characteristics as per IHDS-I such as student age and its square, log of the log of total investment in the education of the child (sum of school fees, books, and uniform and private tuition). We also control for some household characteristics like highest adult education in years as per IHDS-I, the log of per capita consumption as per IHDS-I, whether the household is an agricultural labour household as per IHDS-I and its interaction with gender, whether a household belongs to any of the disadvantaged groups (Scheduled Caste [SC], Scheduled Tribe [ST], Other Backward Classes [OBC], and Muslim), dependency ratio of the household in IHDS-I and Change in dependency ratio in IHDS-II and interaction of the previous dependency ratios with gender, whether the household has experienced single or multiple financial shocks or losses between IHDS-I and IHDS-II and also the interaction of the shock variable with gender. We also control for all drought-prone districts with rainfall less than 1125 mm per annum[5] and its interaction with gender. In Gompertz models, the exponential term of parameter estimates quantifies the hazard's value (i.e., probability) and results in an intuitive model interpretation. Further, in robustness check, we estimate the above models with additional variables of excess rainfall in the academic year or in the previous year. Where we consider a year with excess rainfall if yearly rainfall is more than 25% of its long-term average.

Section-III: Data and Descriptive Statistics

Our sample with both urban and rural residences is having 36,388 person-year observations for 13.704 students and our rural subsample consists of 24,215 person-year observations for 9,383 students. Table 3 reports the overall number and proportions of discrete variables in our sample across gender and dropout status.

Table 3:Descriptive Table for Discrete Variables across Gender and Dropout Status

		Overall		Boys		Girls	
		No Dropout	Dropout	No Dropout	Dropout	No Dropout	Dropout
Observations (n)	13704	11327	2377	5954	1372	5373	1005
Variables				Number (%)			
Agricultural Labour Household=1	1569 (11.4)	1135 (10.0)	434 (18.3)	646 (10.8)	219 (16.0)	489 (9.1)	215 (21.4)
Marginal agricultural Household=1	1907 (13.9)	1503 (13.3)	404 (17.0)	840 (14.1)	252 (18.4)	663 (12.3)	153 (15.2)
Small Agricultural Household=1	910 (6.6)	732 (6.5)	178 (7.5)	388 (6.5)	118 (8.6)	344 (6.4)	60 (6.0)
Drought prone districts	7870 (57.4)	6358 (56.1)	1512 (63.6)	3343 (56.1)	860 (62.7)	3015 (56.1)	652 (64.9)
Large expenditure Shock /loss between IHDS-I and II : MULLOSSCC							
Household experienced No shock=0	4619 (33.7)	4045 (35.7)	575 (24.2)	2145 (36.0)	366 (26.7)	1900 (35.4)	209 (20.7)
Household experienced one Shock =1	4425 (32.3)	3603 (31.8)	822 (34.6)	1923 (32.3)	447 (32.6)	1680 (31.3)	375 (37.3)
Household experienced more than one Shock =2	4660 (34.0)	3679 (32.5)	981 (41.3)	1887 (31.7)	559 (40.8)	1793 (33.4)	422 (41.9)
Other Backward Caste Household =1	4862 (35.5)	4024 (35.5)	838 (35.3)	2151 (36.1)	492 (35.9)	1873 (34.9)	346 (34.4)
Scheduled Caste Household =1	3189 (23.3)	2527 (22.3)	662 (27.9)	1286 (21.6)	363 (26.5)	1240 (23.1)	299 (29.8)
Scheduled Tribe Household =1	727 (5.3)	532 (4.7)	195 (8.2)	314 (5.3)	110 (8.0)	218 (4.1)	85 (8.5)
Muslim Household =1	1556 (11.4)	1225 (10.8)	331 (13.9)	598 (10.0)	187 (13.6)	627 (11.7)	144 (14.3)

[5] Source A BRIEF ON DROUGHT, http://jalshakti-dowr.gov.in/brief-drought accessed on 17.08.2021.

Rural Household = 1	10293 (75.1)	8415 (74.3)	1877 (79.0)	4458 (74.9)	1057 (77.1)	3957 (73.6)	820 (81.6)

Note: Author's computation from IHDS I and IHDS II data using IHDS-I individual weights. Weighted observations across categorical variables rounded to the nearest integer. Percentage of observations depicted inside Parenthesis

Means and standard deviations of continuous variables are presented in Table 4 across gender and dropout categories.

Table 4:Mean and Standard deviation of Continuous variables across Gender and Dropout

		Overall		Boys		Girls	
		No Dropout	Dropout	No Dropout	Dropout	No Dropout	Dropout
Observations	13704	11327	2377	5954	1372	5373	1005
Variables	Mean(sd)						
Age of student in years IHDS-I	10.70 (2.58)	10.43 (2.55)	12.00 (2.31)	10.55 (2.54)	12.44 (2.30)	10.29 (2.56)	11.40 (2.19)
Square of age of student IHDS-I	121.16 (56.01)	115.25 (54.31)	149.31 (55.47)	117.72 (54.50)	159.97 (56.77)	112.51 (53.97)	134.76 (50.12)
Highest adult education (completed years) in household-IHDS-I	7.95 (4.58)	8.40 (4.54)	5.82 (4.15)	8.13 (4.60)	5.78 (4.14)	8.70 (4.45)	5.87 (4.17)
Log of percapita consumption expenditure in household -IHDS-I	6.36 (0.61)	6.40 (0.60)	6.15 (0.58)	6.40 (0.61)	6.20 (0.58)	6.40 (0.60)	6.09 (0.58)
Log of total annual education expense IHDS-I	6.47 (1.42)	6.55 (1.42)	6.07 (1.33)	6.62 (1.41)	6.23 (1.28)	6.47 (1.42)	5.84 (1.36)
Increment in dependency ratio in IHDS-II over IHDS-I in the Household	-0.08 (0.16)	-0.07 (0.16)	-0.13 (0.17)	-0.08 (0.16)	-0.14 (0.18)	-0.06 (0.16)	-0.10 (0.16)
Dependency ratio in the Household in IHDS-I	0.55 (0.13)	0.55 (0.12)	0.58 (0.12)	0.54 (0.12)	0.56 (0.12)	0.56 (0.13)	0.60 (0.12)

Note: Author's computation from IHDS I and IHDS II data using IHDS-I individual weights

Section-IV: Empirical Results

Through discrete-time survival analysis, we examine the association of rainfall shocks with the hazards of high school dropout in an academic year in two analytical samples drawn from the students from rural residences (model 1) and as well as students from both urban and rural residences (model 2) respectively. We report the results in Table 5 where the estimated coefficients for our selected variables of interest for the discrete-time hazard models with the Gompertz link function along with their standard errors for the variables of interest is presented. In Table 5, model 1 is estimated on both urban and rural residences with 36388 person-year observations for 13704 students while model 2 is estimated on students from the rural subsample with 24215 person-year observations for 9383 students.

Table 5: Results

Prob(Y=1)Dependent Variable Y: Whether Individuals who were students in the IHDS-I survey round but did not enter high school, dropped out of education in the high school year 9,10,11 or 12 as per the IHDS-II survey? (1: "Yes", 0: "No")				
Variable Description	**Model 1: Urban & Rural**		**Model 2: Rural**	
	1a	1b	2a	2b
Girl in Marginal Agricultural Household and Drought in Current Academic year	-1.217** (0.490)	0.30**	-1.140** (0.507)	0.32**
Student in Marginal Agricultural Household and Drought in Current Academic year	0.623* (0.322)	1.86*	0.602* (0.328)	1.83*
Girl in Small Agricultural Household and Drought in Current Academic year	-1.114**(0.516)	0.33**	-0.928* (0.532)	0.40*
Student in Small Agricultural Household and Drought in Current Academic year	0.453 (0.342)	1.57	0.416 (0.350)	1.52
Girl in Marginal Agricultural Household and Drought in Previous Academic year	0.951**(0.444)	2.59**	0.849*(0.463)	2.34*
Student in Marginal Agricultural Household and Drought in Previous Academic year	-0.480* (0.292)	0.62*	-0.492 (0.305)	0.61
Girl in Small Agricultural Household and Drought in Previous Academic year	0.093 (0.519)	1.10	-0.076 (0.540)	0.93
Student in Small Agricultural Household and Drought in Previous Academic year	0.422 (0.352)	1.53	0.421 (0.363)	1.52
Girl and Drought in Current Academic year	0.284*(0.172)	1.33*	0.167 (0.210)	1.18
Girl and Drought in Previous Academic year	-0.113 (0.181)	0.89	-0.016 (0.217)	0.98
Drought in Current Academic year	-0.114 (0.097)	0.89	-0.084 (0.120)	0.92
Drought in Previous Academic year	-0.003 (0.097)	1.00	-0.010 (0.122)	0.99
GIRL in Agricultural Labour Household	0.679***(0.182)	1.97***	0.627***(0.202)	1.87***
Girl in Marginal Agricultural Household	0.085 (0.257)	1.09	0.052 (0.275)	1.05
Girl in Small Agricultural Household	0.182 (0.303)	1.20	0.076 (0.318)	1.08
Student in Agricultural Labour Household	-0.001 (0.107)	1.00	-0.006 (0.118)	0.99
Student in Marginal Agricultural Household	0.205 (0.167)	1.23	0.233 (0.176)	1.26
Student in Small Agricultural Household	-0.002 (0.192)	1.00	0.033 (0.196)	1.03
Girl in Drought prone districts	0.078(0.133)	1.08	0.231(0.158)	1.26
Student in Drought prone districts	0.189***(0.085)	1.21***	0.147(0.102)	1.16

Prob(Y=1)Dependent Variable Y: Whether Individuals who were students in the IHDS-I survey round but did not enter high school, dropped out of education in the high school year 9,10,11 or 12 as per the IHDS-II survey? (1: "Yes", 0: "No")				
Variable Description	**Model 1: Urban & Rural**		**Model 2: Rural**	
GIRL	-1.411***(0.406)	0.24***	-1.582***(0.491)	0.21***
Rural	-0.208*** (0.071)	0.81	---	
AIC	**14838.39**		**10179.22**	
Deviance	**15130.56**		**10634.83**	
Dispersion	**0.918**		**0.927**	
Pseudo-R^2 (McFadden)	**0.12**		**0.11**	
Num. obs.	**36388**		**24215**	
Number of Students	**13704**		**9383**	
Notes: a). Estimated coefficient values of selected variables of interest have been reported for discrete-time hazard models with Gompertz link function using IHDS-I survey weights for individuals, Standard errors are inside the parenthesis, b) column 1a and 2a depicts estimated coefficients with standard error in parenthesis and column 2a and 2b depicts exponentiated value of the coefficients as per column 1a and 2a respectively b) ***p < 0.01, **p < 0.05, *p < 0.1				

In Gompertz models, the exponential term of parameter estimates quantifies the hazard's value (i.e., probability) and results in a more intuitive model interpretation. For example, in Table 5, the exponentiated value of 1.86 of the significant coefficient estimate for students in marginal agricultural households who had experienced drought in the current academic year and exponentiated value of 0.3 for the significant coefficient estimate for girls in marginal agricultural households who had experienced drought in the current academic year in model 1, estimated on students from both urban and rural residence, denotes that holding other things constant, in such situations and households, while the hazard of dropout from high school for boys significantly increases 86% (1.86-1) in that academic year, girls face 70% (1-0.3=0.7) lesser hazard or probability of dropout compared to boys.

Interpreting in the same manner, as per Table 5, we observe that in model 2, estimated on the rural sample, boys from marginal agricultural households face an 83% greater hazard or probability of dropout in case of drought in the current academic year and girls in such situations and households face 68% (1-0.32) lesser hazard or probability of dropout compared to boys. Also, the hazard of dropout of girls during drought in the current academic year reduces significantly by 67% (1-0.33) and 60% (1-0.4) in small agricultural households in model 1 and model 2, respectively. However, the hazard of dropout of students (boys) from high school during drought in the current academic year in the small agricultural household does not become significant in both models.

Further, in Table 5, we observe that the dropout hazard of girls in marginal agricultural households increases by 159% and 134% if there was a drought in the previous academic year in model 1 and model 2, respectively. We also observe that if there was a drought in the previous academic year, hazard of dropout of boys reduces significantly by 38% in marginal agricultural household as per model 1 though in model 2 it is not significant.

In general, we observe no significant parameter estimates for drought in the current academic year or previous academic year for students. However, in the overall model (model1) we observe that at times of drought in current academic year girls face a 33% higher chance of dropout.

In addition, as per Table 5, we find that girls from agricultural labour households have an 97% and 87% greater hazard of dropout from high school compared to boys in such households in model 1 and model 2, respectively[6]. Further, we also observe a 21% higher dropout hazard for students from drought-prone districts in model 1, containing students from both urban and rural locations.

Robustness check: In Table 6 we present the results of robustness check estimates of the models in Table 5 where we estimate the same models with additional variables of excess rainfall in the academic year or in the previous year. We observe that our robustness results are also in the same line with our main results.

These results support our propositions that, unlike other covariate shocks, drought or lesser rainfall shocks are more like agricultural shocks and have greater and varied impacts on the high school education of girls and boys in the marginal and small farming households among the households in a pan-India context. Our findings are somewhat different from Zimmerman's (2020) in a way that we observe that gender choices in high school dropouts during and after a drought are different in small and marginal agricultural households compared to other households. Therefore, targeted policy support would be required to these households to reduce high school dropouts of students from these households.

Table 6:Results with robustness check

Prob(Y=1)Dependent Variable Y: Whether Individuals who were students in the IHDS-I survey round but did not enter high school, dropped out of education in the high school year 9,10,11 or 12 as per the IHDS-II survey? (1: "Yes", 0: "No")				
Variable Description	**Model 1: Urban & Rural**		**Model 2: Rural**	
	1a	1b	2a	2b
Girl in Marginal Agricultural Household and Drought in Current Academic year	-1.215** (0.484)	0.30**	-1.136** (0.499)	0.33**
Student in Marginal Agricultural Household and Drought in Current Academic year	0.623** (0.310)	1.87**	0.600* (0.315)	1.82*
Girl in Small Agricultural Household and Drought in Current Academic year	-1.117** (0.515)	0.33**	-0.930* (0.530)	0.40*
Student in Small Agricultural Household and Drought in Current Academic year	0.452 (0.342)	1.57	0.417 (0.349)	1.52
Girl in Marginal Agricultural Household and Drought in Previous Academic year	0.947** (0.442)	2.58**	0.846* (0.458)	2.33*
Student in Marginal Agricultural Household and Drought in Previous Academic year	-0.480* (0.286)	0.62*	-0.490* (0.296)	0.61*

[6] As we did not find any significant coefficient estimates of drought in current and previous year for both boys and girls in households with agricultural labour as principal occupation, for the sake of parsimony, we do not estimate and report such interaction effects in our results.

Girl in Small Agricultural Household and Drought in Previous Academic year	0.104 (0.516)	1.11	-0.076 (0.535)	0.93
Student in Small Agricultural Household and Drought in Previous Academic year	0.421 (0.352)	1.52	0.420 (0.363)	1.52
Girl and Drought in Current Academic year	0.292* (0.165)	1.34*	0.163 (0.200)	1.18
Girl and Drought in Previous Academic year	-0.124 (0.181)	0.88	-0.031 (0.215)	0.97
Drought in Current Academic year	-0.105 (0.098)	0.90	-0.081 (0.119)	0.92
Drought in Previous Academic year	0.002 (0.102)	1.00	-0.002 (0.127)	1.00
GIRL in Agricultural Labour Household	0.677*** (0.182)	1.97***	0.627*** (0.202)	1.87***
Girl in Marginal Agricultural Household	0.086 (0.257)	1.09	0.054 (0.276)	1.06
Girl in Small Agricultural Household	0.176 (0.304)	1.19	0.077 (0.318)	1.08
Student in Agricultural Labour Household	-0.001 (0.107)	1.00	-0.006 (0.118)	0.99
Student in Marginal Agricultural Household	0.204 (0.167)	1.23	0.231 (0.176)	1.26
Student in Small Agricultural Household	0.000 (0.192)	1.00	0.034 (0.196)	1.04
Girl in Drought prone districts	0.078 (0.136)	1.08	0.238 (0.161)	1.27
Student in Drought prone districts	0.186** (0.083)	1.20**	0.144 (0.100)	1.16
Excess Rain in Current Academic year	0.045 (0.121)	1.05	0.016 (0.155)	1.02
Excess Rain in Previous Academic year	0.013 (0.145)	1.01	0.033 (0.184)	1.03
GIRL	-1.404*** (0.415)	0.25***	-1.564*** (0.502)	0.21***
Rural	-0.206*** (0.072)		---	
AIC	**14845.15**		**10179.22**	
Deviance	**15129.70**		**10634.57**	
Dispersion	**0.918**		**0.928**	
Pseudo-R² (McFadden)	**0.12**		**0.11**	
Num. obs.	**36388**		**24215**	
Number of Students	**13704**		**9383**	

Notes: a). Estimated coefficient values of selected variables of interest have been reported for discrete-time hazard models with Gompertz link function using IHDS-I survey weights for individuals, Standard errors are inside the parenthesis, b) column 1a and 2a depicts estimated coefficients with standard error in parenthesis and column 2a and 2b depicts exponentiated value of the coefficients as per column 1a and 2a respectively b) ***p < 0.01, **p < 0.05, *p < 0.1

Section-V: Conclusion and Policy Implications

We examine whether droughts are differently associated with high school dropout of girls and boys from marginal and small agricultural households compared to other households. With discrete-time survival analysis, we examine the association of droughts with the hazards of high school dropout by combining publicly available data from two rounds of the India

Human Development Survey (2004-05 and 2011-12) with ICRISAT district-level rainfall data for India.

We find that while boys in marginal agricultural households face a significantly higher hazard of high school dropout, the hazard of dropout of girls from high school significantly reduces in drought years in marginal and small agricultural households. We observe that the hazard of dropout of girls in marginal agricultural households increases significantly if there was a drought in the previous academic year. Additionally, we also observe that girls in agricultural labourer households are more prone to dropping out of high school. Thus, our study shows heterogenous procyclical and countercyclical behavior of marginal and small agricultural households regarding educational investment in the children according to gender at times of drought.

Our study has limitations that the student's participation in certain high school grade will be approximated based upon completed years in IHDS-I and the month/year of survey and we have assumed no grade repeat in between the two survey rounds, incorrect reporting etc.

However, we uniquely contribute by combining data from different publicly available datasets and deploying discrete-time survival analysis to bring out the heterogenous relationship between drought and high school dropout of boys and girls in marginal and small agricultural households.

Effective policy interventions are needed to provide high school education to all in the face of climate change and drought. Particularly, policy makers should consider implementing policies aimed at reducing the vulnerability of agricultural households to climate variability and change. Fiszbein, et al., (2009) suggested that conditional cash transfers (CCT) can drive the actual household choice of a child's education towards optimal level.

As gender heterogeneity in the dropout choice of households are also plausibly emanating from underlying stereotyped gender roles in the society, we suggest that policies incentivizing greater participation of women in the labour force might also alleviate the existing condition.

Substantial improvement in enrolment and dropout rate in elementary education had only occurred when education up to the elementary level was taken up as an obligation of the government. A more demanding Right to Education in India that includes school education may be an essential step towards reducing high school dropouts and achieving the Sustainable Development Goals of the United Nations.

Acknowledgement

We thank the IHDS and ICRISAT team for making their dataset publicly available. Authors bear full responsibility for any remaining errors.

References

Colmer, J. (2020). Rainfall Variability, Child Labor, and Human Capital Accumulation in Rural Ethiopia. American Journal of Agricultural Economics, 103(3), 858-877. Retrieved from https://doi.org/10.1111/ajae.12128

Cox, D. R. (1972). Regression models and life -tables (with discussion). Journal of the Royal Statistical Society, series,B 187-220.

Desai, S., Vanneman, R., & and National Council of Applied Economic Research, N. D. (2005). India Human Development Survey (IHDS). 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, S., Vanneman, R., & and National Council of Applied Economic Research, N. D. (2011-12). India Human Development Survey-II (IHDS-II). 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, J., & Kingdon, G. G. (2001). School Participation in Rural India. Review of Development Economics, vol. 5,issue 1,1-24.

Fang, Q., & van de Schoot, R. (2019, October 14). Intro to Discrete-Time Survival Analysis in R. Retrieved from www.rensvandeschoot.com: https://www.rensvandeschoot.com/tutorials/discrete-time-survival/

Ferreira, J. A., Santos, E. J., Fonseca, A. C., & Haase, R. F. (2007). International Journal of Educational Development A 10-year follow-up study. Journal of Vocational Behavior, 70, 61-77.

Fiszbein, A., Schady, N., Ferreira, F. H., Grosh, M., Keleher, N., Olinto, P., & Skoufias, E. (2009). Conditional cash transfers: A World Bank Policy Report. Washington, D.C.: World Bank. Retrieved from https://doi.org/10.1596/978-0-8213-7352-1

Fulford, S. (2014). Returns to Education in India. World Development, 59, 434-450. doi:10.1016/j.worlddev.2014.02.005

Gelli, A., Masset, E., Diallo, A. S., Assima, A., Hombrados, J., Watkins, K., & Drake, L. (2014). Agriculture, nutrition and education: On the status and determinants of primary schooling in rural Mali before the crises of 2012. International Journal of Educational Development, 39, 215-225.

Goel, S., & Zakir, H. (2018, December). Gender, caste, and education in India: A cohort-wise study of drop-out from schools. Research in Social Stratification and Mobility, 58, 54-68.

Jenkins, S. (2005). Survival Analysis. Unpublished Manuscript. Mimeo: University of Essex. Retrieved from https://www.iser.sx.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf

Kearney, M. S., & Levine, P. B. (2016). Income Inequality, Social Mobility, and the Decision to Drop Out of High School. Brookings Papers on Economic Activity, 1, pp. 333-396.

Kleinbaum , D. G., & Klein, M. (2008). Survival Analysis: A Self -Learning Text Second Edition. New Delhi: Springer (India) private Limited.

LaFave, D., & Thomas, D. (2016). Farms. Families, and Markets: New Evidence on Completeness of Markets in Agriculrural Settings. Econometrica, 84(5), 1917-1960.

Liu, X., & Hannum, E. (2017). Early poverty exposure predicts young adult educational outcomes in China. China Economic Review, 44, 79-77.

Marchetta, F., Sahn, D., & Tiberti, L. (2019). The Role of Weather on Schooling and Work of Young Adults in Madagascar. American Journal of Agricultural Economics, 101(4), 1203-1227. Retrieved from https://doi.org/10.1093/ajae/aaz015

Marshall, J. H. (2011). School quality signals and attendance in rural Guatemala. Economics of Education Review, 30, 1445-1455.

Momo, M. S., Cabus, S. J., Witte, K. D., & Groot, W. (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

Singh, I., Squire, L., Strauss, J., & (Editors). (1986). Agricultural Household Models: Extensions, Applications,and Policy. Baltimore and London: Published for the World Bank , The Johns Hopkins University Press.

Sowmya, D., Paul, C. M., & Gade, S. (2019). Household income dynamics and investment in children: Evidence from India. Education Economics, 27(5), 507-520. doi:10.1080/09645292.2019.1599325

Tamusuza, A. (December 2011). Leaving School Early: The Quest for Universal Primary Education in Uganda. Journal of Statistique Africain, 13, 110-151. Retrieved from https://www.afdb.org/fr/documents/document/african-statistical-journal-vol-13-25569

Taylor, J. E., & Adelman, I. (2003). Agricultural Household Models: Genesis, Evolution, and Extensions. Review of Economics of the Household, 1, 33-58.

Thomas, M. K., Singh, P., & Klopfenstein, K. (November 2015). Arts education and the high school dropout problem. Journal of Cultural Economics, 39(4), 327-339.

Valdivieso, P. (2015, September). Survival Analysis: Exploring the dropout motives in a panel of Peruvian children, using the Young Lives program dataset. Student paper, MSc Dissertation, University of Edinburgh. Young Lives International Study of Childhood Poverty ,University of Oxford. Retrieved from https://w)ww.younglives.org.uk/sites/www.younglives.org.uk/files/YL-SP_Valdivieso_Drop-out%20in%20Peru.pdf

Zhang, H. (2017). Opportunity or new poverty trap: Rural-urban education disparity and internal migration in China. China Economic Review, 44, 112-124.

Zimmermann, L. (2020). Remember when it rained--Schooling responses to shocks in India. World Development, 126. Retrieved from https://doi.org/10.1016/j.worlddev.2019.104705

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