UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2020/107

Subjective returns to education

Rational expectations of disadvantaged groups in India

Pooja Balasubramanian*

August 2020

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: This study uses data collected from school students in Mumbai to investigate how they perceive subjective expected returns for different levels of education in an environment that includes labour market discrimination. We are particularly keen to observe subjective returns to education for different social identity groups, such as gender, religion, and caste. Despite lower actual returns to education in the labour market, students from Other Backward Castes and Scheduled Castes and Tribes do not have significantly different beliefs relative to their advantaged peers. To estimate the differences in subjective returns to education, we use a novel statistical technique: distributional regression. This method can be a complement to the mean estimates, enabling the assessment of heterogeneity in expected earnings for different social groups in India. Using the distributional regression, we observe that females and Muslims have left-skewed earnings distributions relative to our reference group (male, upper caste Hindus).

Key words: labour market discrimination, social identity, returns to education, India, heterogeneity, distributional regression

JEL classification: I26, J7, J24, Z13

* University of Goettingen, Goettingen, Germany; pbalasu@gwdg.de

This study has been prepared within the UNU-WIDER project Addressing group-based inequalities.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-864-1

https://doi.org/10.35188/UNU-WIDER/2020/864-1

Typescript prepared by Gary Smith.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Over the last 40 years, countries have shown an increasing trend in secondary and tertiary educational attainment owing to the demand for an educated workforce (Desai and Kulkarni 2008; Fasih et al. 2012). However, at the same time, educational disparities between groups and social strata are increasing (Kabeer 2014). Based on the World Inequality Database on Children, the Global Education Monitoring Report of UNESCO has provided an overview of these inequalities. The report shows that, across different continents, children from lower-income groups, rural areas, and conflict-affected communities have lower mean education and higher likelihood of being out of school. Similarly, the likelihood of completing higher and tertiary education is lowest for females and children from minority tribes and ethnicities (UNESCO 2015). In this paper, we investigate the drivers of persistent gaps in school enrolment and attainment of higher education across children with different social identities in India. Our hypothesis is that there exists biases in perceptions on returns to education for different minority and traditionally discriminated groups (relative to privileged groups). Such biased perceptions can be considered as consequences of the existing inequality traps. We focus on specific groups, including females, disadvantaged castes, and minority religions in India. The earnings expectations are observed for three levels of education: completion of high school, technical education, and college.

The study tests whether Muslims, females, and lower caste groups expect lower returns to education compared to Hindu male and upper caste students (Duraisamy and Duraisamy 2017; Kingdon 2020; Kingdon and Unni 2001; Madheswaran and Attewell 2007). Despite its complexities and varied groups or 'Jatis', the caste system in India is classified into three broad categories for comparison: General (upper caste or advantaged groups), Other Backward Caste (OBC), and Scheduled Caste and Tribes (SC-ST) based on the Census survey of India. Among the two discriminated caste groups (OBC and SC-ST), the inequalities faced by the SC-ST groups are considered to be more severe than those the OBC category face. While the OBCs are in the lower ranks of the caste hierarchy, the SC (Dalits) were considered to be outside of the caste system and deemed 'untouchables'. Similarly, the STs were stigmatized and considered primitive, with an inclination towards crime (Deshpande 2013). Such historical exclusion can be observed today not only in the form of economic deprivation, but also the low standard of living, extreme poverty, and health deprivation (Thorat and Newman 2010). As a result, we expect subjective earnings expectations from education for SC-ST groups to be lower than for OBCs relative to the advantaged upper caste groups. Apart from caste, gender has been an important component when inequality in access to education and equal opportunities is discussed. Studies observing actual earnings in the labour market show that females to have lower returns to education than do males (Duraisamy and Duraisamy 2017; Kingdon 2020; Kingdon and Unni 2001; Madheswaran and Attewell 2007). Considering inequality traps whereby lower returns in the labour market further impede investment in the same, we expect females in our sample to have lower earnings expectations relative to males (Bourguignon et al. 2007; Deininger et al. 2013). Among females, upper castes are expected to have higher earnings expectations compared to Muslims and lower castes. This hypothesis is supported by recent secondary data showing that human development outcomes are far inferior for SC (Dalit) and Muslim women compared to upper caste women. The 2011 Indian Census shows that while 64 per cent of the upper caste Hindu women were literate, only at 56 per cent of SC-ST and Muslim women could say the same.

Stated beliefs have been used to observe decision-making in various domains such as migrant behaviour (McKenzie et al. 2013), environmental concerns (Luseno et al. 2003), and credit constraints (Attanasio and Kaufmann 2009), to name a few. In the context of demand for education and schooling decisions, subjective expectations are shown to differ for students across different economic backgrounds. Studies by Avery and Kane (2004), Dominitz and Manski (1997, 2006), and Attanasio and Kaufmann (2009) find actual returns and subjective returns are not correlated, and low-income students either tend to underestimate expected earnings or overestimate tuition costs. As a result, more credit-constrained students are likely to drop out of education (Attanasio and Kaufmann 2014). Controlling for the stu-

dent's socioeconomic background, Stinebrickner and Stinebrickner (2014) find academic and cognitive abilities of children to explain higher subjective expectations from education. Another mechanism that impacts demand for schooling, through increasing monetary expectations, is parental interest (Attanasio and Kaufmann 2014; Dizon-Ross 2019). Attanasio and Kaufmann (2009) measure perceived earnings for education not only among children but also for their mothers. They find the mother's expectations regarding their child's earning capabilities to be an important indicator of demand for schooling. Dizon-Ross (2019) finds parents in Malawi have inaccurate perceptions regarding their children's academic capabilities, resulting in less investment in education. She emphasizes the effect to be particularly significant among low-income households.

This study contributes to the above literature by focusing on whether horizontal inequalities based on an individual's social group could impact subjective returns. Controlling for economic and educational backgrounds of the parents, this paper estimates stated beliefs regarding returns to education for children from disadvantaged and minority groups, such as females, Muslims, and lower castes (OBC and SC-ST). The closest other work to this study is by Huntington-Klein (2015), who estimated subjective returns at different educational levels among students in the state of Washington (USA), with a particular focus on the heterogeneity across children from different races. Furthermore, a comparison of the subjective and observed data show them to be uncorrelated. While he does not find a gender difference in the subjective and projected earnings, black students have a higher projected return, but seem to have lower subjective expectations. Other studies focusing on race in the USA find similar results (Connor 1999; Cunha and Heckman 2007; Henderson et al. 2011). However, the study has two drawbacks: first, returns are estimated only using a point estimate; and second, since he has cross-sectional data, the author is unable to provide evidence of whether the data is driven by measurement error or if it captures the actual education choices made by the students.[1]

We extend this work and contribute to this scarce literature in three ways. (1) We implement rigorous methods to elicit subjective responses from individuals belonging to different social groups. Following the studies by Delavande (2009), Attanasio and Augsburg (2016), and Dominitz and Manski (2006), the paper collects information on the minimum and maximum earnings (thresholds) that students expected to earn for different levels of education and the likelihood of earning an amount greater than the midpoint of this self-indicated threshold (see Delavande 2009; Delavande and Rohwedder 2008). This method allows us to construct a distribution of subjective earnings for each individual for each level of education. One of the main reasons to use the conditional probability distribution is the need for comparability. Dominitz and Manski (2006) show that individuals formulate expectations of the point estimate based on different thresholds, and there is extensive heterogeneity in how people update their beliefs. The probability distribution for each individual controls for this level of uncertainty.[2] (2) The paper analyses panel data of the students' subjective returns over two years. This dataset allows us to use statistical methods that controls for individual-specific unobservables, thereby testing whether the expectations reported by the students are consistent.

(3) The third contribution of this paper is on the methodological front. By using the novel distributional regression (DR) technique, we are able to analyse the differential effect of belonging to a specific social identity not only on the mean outcome variable (subjective returns to education), but on the entire conditional distribution of the outcome. This method is developed as a complementary technique to ordinary least squares (OLS) estimation that only observes a linear change of explanatory variables on point esti-

[1] Two other studies observe subjective returns to education in India. Maertens (2011) finds subjective earnings expectations to be lower for girls relative to boys. A second paper by Chari and Maertens (2014) finds parents to report lower earnings for girls compared to boys, and explains this disparity based on differing abilities. However, both studies use point estimates to elicit subjective earnings, and are based on a rural sample in India.

[2] Dominitz and Manski (2006) find individuals to have the same point estimate but vary in the level of uncertainty, with overconfidence bias resulting in downweighting of the mean and heterogeneity in each individual's process of belief updating.

mates of the outcome (Hohberg et al. 2017; Stasinopoulos et al. 2018). Finally, as a comparison to the Huntington-Klein (2015) study, I correlate the inflation-adjusted actual earnings for the three levels of education with the subjective earnings reported by the students.

The findings reveal that females expect to gain lower earnings compared to males for all three levels of education. While Muslims expect low returns compared to upper caste Hindus for high school and technical education, the difference is not significant for college. Rejecting our hypothesis, lower caste groups (SC-ST and OBCs) expect higher returns from college education relative to the Hindu upper caste males. Incorporating other moments in the distribution, we find females, Muslims, and SC-ST students have significantly more left-skewed earnings distributions for school and diploma-level education. Past projected earnings are not correlated with current subjective returns. While most subsamples underestimate their earnings relative to actual earnings, male lower caste groups (e.g., SC and ST) have higher subjective returns despite low actual returns. The final section of the paper discuses potential mechanisms explaining this result.

2 Local context

With the increasing violence and exclusion of certain social groups, it is important to study how different social identities perceive opportunities such as education, employment, and economic growth in India. In this study, the focus is specifically on the perception of returns to education for females, disadvantaged caste groups, and Muslims. There is an attempt to extend the analysis to study intersectionality; namely the responses of females from lower castes and Muslim communities.

There is extensive secondary evidence that the above-mentioned disadvantaged groups have been excluded from the economic prosperity that India is experiencing. The Sachar Report in 2006 pointed out that the mean years of schooling for Muslim children has been consistently low, with nearly 25 per cent of Muslim children aged 6–14 years having either never attended school or have dropped out of school. A recent study by Asher et al. (2017) shows decreasing upward mobility for children born into Muslim families. Relative to other social groups, their likelihood of economic mobility fell from 31.5 in 1960 to 29 in 1980. Although overall dropout rates have drastically reduced among Indian children, the difference in dropout between all Indian youth and lower caste youth has significantly increased from 4 per cent in 1989 to 16.21 per cent in 2008. Further disaggregating disadvantaged caste groups as OBCs and SC-ST (Deshpande 2013),[3], the National Sample Survey (NSS) in 2000 showed that 37 per cent of males in the SC-ST groups and 44 per cent of OBCs have never enrolled in formal education, compared to 17 per cent among the upper caste Hindus or the General category (Desai and Kulkarni 2008; Dreze and Sen 2001; Thorat and Newman 2010).

One aspect that distinguishes the two disadvantaged groups—Muslims and lower caste group such as the OBCs and SC-ST—is the provision of AA or quotas that benefit the latter. With such institutional policies for the SC-ST and OBC, their situation is slightly better than that of Muslims, who do not have institutional policies to overcome their consistent exclusion (Asher et al. 2017; Goel and Deshpande 2016). A common thread across the disadvantaged religious and caste groups is the status of women in the Indian context. Relegated to the role of second-class citizens (Dyson and Moore 1983), women face discrimination across the identity spectrum. From birth, the culturally ingrained parental preference for a son can be observed in the increasing sex ratio in favour of men (Pike 2011; Sharma 2016), decreasing labour force participation (Sarkar et al. 2019), and lower entry into educational institutions

[3] Despite the caste system being complex, with a number of groups under an established hierarchy, most studies follow the categorization used by the National Census and the Affirmative Action (AA) programme that divides the caste groups into four broad categories: the general (upper or privileged) caste, OBCs, SCs (untouchables also known as Dalits), and STs (tribals).

(World Bank 2014). This study explores whether the actual discrimination in education returns is reflected in the minds of youth, particularly children who are yet to complete schooling. The study focuses on highlighting whether these marginalized groups have internalized such overt discrimination and adjusted their beliefs in terms of expected earnings. I hypothesize that the identity groups such as females, Muslims, OBCs, and SC-ST castes expect to earn lower income from higher education compared to the privileged groups (upper caste Hindu males and females).

3 Data and methodology

This section provides a detailed description of the data collection procedures, measurement of the subjective earnings, summary statistics, and the empirical strategy used to estimate the subjective expectations for the different social groups.

3.1 Survey procedure

A survey was conducted in 2017 and 2018 among children aged 12–17 years in public schools in different parts of Mumbai, India. Information on subjective earnings expectations was collected for three levels of education: high school, diploma or technical education, and college. High school includes 10 years of education but does not provide any skills or specialization in the labour market. Technical or diploma education can be considered as an alternative to college or university education. It provides technical skills and focuses on immediate employability. The courses can last from six months to two years.[4]. College or a university degree means 15 years of education within the Indian system.

In both years the exact same procedure was followed. First, all the children completed an 'Education survey' in their classrooms. The survey included questions on socio-demographics and detailed information regarding their earnings expectations for different levels of education. In the subjective earnings schedule of the survey, the students had to answer the following questions: *In the future, when you are 28 years old, do you think you will be working if you completed 12th standard (high school)/diploma (technical education)/college (university)?*. For each education level (high school, diploma, and college), if the student answered 'Yes' they had to answer the follow-up questions regarding earnings:[5]

- In the future, when you are 28 years old, what is the minimum (maximum) monthly income you think you will earn if you complete 12th standard (high school)?
- In the future, when you are 28 years old, what is the minimum (maximum) monthly income you think you will earn if you complete diploma or a technical education course?
- In the future, when you are 28 years old, what is the minimum (maximum) monthly income you think you will earn if you complete college?

For the first question, regarding whether the students expected to work for a given level of education, a total of 18 students said they would not work.[6] These 18 students did not answer the earnings question for the specific level of education. For the remaining students who said they would be working, we calculated the midpoint for the earnings at each education level. The midpoint was based on each

[4] The courses include information technology, medical technician training, accounting, and home economics, to name a few.

[5] If the students wished to respond to the expectations questions they could, but we undertook the analysis conditional on this response

[6] Out of the 18 students, seven said they would not be working after completing high school, four said they would not work after diploma-level education, and eight said they would not work after completing college education.

student's reported minimum and maximum threshold. After two days we went back to the same schools and conducted a one-to-one interview ('exit survey') with the children. In this survey, the children were asked what they thought was the likelihood (from a scale of 0–10) of earning greater than or equal to the midpoint of their self-reported threshold (Attanasio and Kaufmann 2014; Guiso et al. 2002). A final survey was conducted a few weeks after the education and the exit surveys, in which our enumerators made phone calls to the parents of our respondents and asked them a few questions regarding their children and their perceptions of equal opportunities in work and education for all genders, religions, and disadvantaged caste groups.

3.2 Measuring subjective earnings

Subjective earnings expectation was calculated from the education and exit survey based on the following information: range of self-reported earnings expectations, midpoint, and the likelihood for earnings to the right of the midpoint. This paper follows the method used in the recent literature to elicit expectations (Attanasio and Kaufmann 2014; Delavande 2009; Delavande and Rohwedder 2008; Dominitz and Manski 1997). By using the conditional probability distribution, the study is able to control for the variation in the interquartile range and account for heterogeneity across the different social identity groups of interest. The expected earnings are calculated in the following way:

$$E_d(y) = \int_{y_m}^{y_M} \ln(y) f_d(y) dy = \int_{y_m}^{y_{mid}} \ln(y) f_d(y) dy + \int_{y_{mid}}^{y_M} \ln(y) f_d(y) dy \quad (1)$$

where y_m and y_M are the log self-reported minimum and maximum earnings by student i conditional on working full-time with educational degree d (high school, diploma, or college). We make certain assumptions to calculate the individual-specific expected mean of the elicited subjective distribution; first, the likelihood p that earnings are to the right of the midpoint is given by $p = Prob(y \geq (y_m + y_M)/2)$. Second, we assume the probability mass within the two thresholds (i.e. $[y_m, y_{mid}]$ and $[y_{mid}, y_M]$) has a triangular distribution (Attanasio and Kaufmann 2014; Guiso et al. 2002) (see Appendix A2 for details).[7] Following the work by Huntington-Klein (2015), the study additionally compares the projected returns for the three levels of education. The projected weekly earnings were calculated using the NSS (2011–12) for urban Mumbai using the Mincer equation. This was adjusted for the consumer price index of 2018, when the survey took place.

3.3 Data descriptives

This section discusses the descriptives of the sample. Table 1 shows the sample characteristics of the children, who were on average 13 years old; 48 per cent of the students were girls and 75 per cent belonged to the Hindu religious group. The cognitive ability of the students was assessed using the six Raven's matrices. Thus, the score could range from zero (no correct answers) to six (all correct answers). On average, the children answered four questions correctly. Regarding the parent's responses, the father's monthly income within the household was on average Rs.16,000, while the mother's was Rs.8,000. Both parents have completed on average 10 years of education. When asked about their opinions on gender equality and discrimination against minorities, 41 per cent of the parents supported gender equality. However, they believed that when jobs are scarce men should be given priority (65 per cent). When asked whether minorities and caste groups face discrimination in the workplace and at school, on average 75 per cent of the parents agreed that they did. These results are similar to the responses in the nationally implemented World Value survey.

[7] Studies have used both uniform and triangular distributions. Attanasio and Kaufmann (2009) used the triangular distribution, which gives greater weight to the responses closer to the midpoint rather than to the extremes.

Table 1: Summary statistics

	Mean	Std dev.	Min.	Max.	Obs.
Panel A: Children					
Age	13.07	1.29	10	17	408
Female	0.49	0.50	0	1	408
Muslim	0.25	0.43	0	1	408
Cognitive ability	3.89	1.53	0	6	408
Leader caste	2.10	0.59	1	3	187
Panel B: Father					
Income	16,458.96	13,658.63	1,000	90,000	268
Gender equality	0.38	0.49	0	1	269
Men get more jobs	0.64	0.48	0	1	269
Minorities discriminated in work	0.74	0.44	0	1	269
Minorities discriminated in education	0.66	0.47	0	1	270
Assets	0.67	0.47	0	1	270
Education years	10.84	3.27	0	15	386
Quota	0.71	0.45	0	1	115
Panel C: Mother					
Income	8,367.31	7,897.87	500	40,000	52
Gender equality	0.37	0.48	0	1	275
Men get more jobs	0.63	0.48	0	1	275
Minorities discriminated in work	0.75	0.43	0	1	276
Minorities discriminated in education	0.71	0.46	0	1	278
Assets	0.76	0.43	0	1	277
Education years	10.63	3.66	0	17	399

Source: author's construction based on survey results.

In the education survey, apart from collecting information on earnings expectations, we also collected detailed information on the aspirations of the children, such as: *What work would you like to do to earn an income?* and *What is the highest level of education you want to attain?* These two questions on aspirations regarding work and education were not asked in the one-to-one format by the enumerator, but were included in the main survey that the children completed on their own. The question on aspirations is a categorical variable. Similarly, the caste identity of the children is also categorical. Table A1 in the Appendix provides a two-way frequency of the disaggregated caste categories and the levels of education they wish to attain. The distribution of caste in my sample is 38 percent upper caste or general, 49 percent OBC, and 12 percent SC-ST (i.e. the oppressed). Regarding non-monetary aspirations, 38 percent and 33 percent of our sample wish to complete up to diploma or college education, respectively; 13 percent reported wanting to complete high school and 14 percent said they didn't know.

Table 2 presents the raw data of the earnings calculated for each level of education. It shows the minimum, maximum, and midpoint earnings reported by the students. Using the range of variation, log expected income (subjective) and log weekly income (subjective) are calculated from Equation 1. Log weekly projected income is the actual earnings estimated using the NSS (2011–12) for urban Mumbai. Similar to the projected weekly income, the students on average expect highest earnings from a diploma or technical education. The last row is the probability mass to the right of the midpoint earnings. Figure A2 in the Appendix shows the correlation between projected and subjective earnings for school, diploma, and college. Similar to Huntington-Klein (2015) and Jensen (2010), there is no correlation between the projected and subjective earnings specifically for this sample. Following Delevande et al. (2011), we could observe whether the students understood the basic property of probabilities. We asked each child two questions about nested events: (1) What is the likelihood (between 0 and 10) that you will finish your homework in the next two days? (2) What is the likelihood (between 0 and 10) that you will finish your homework in the next two weeks? As a test to understand probability, the students should assign a larger likelihood to finishing the homework in two weeks compared to two days. All except for

39 students obeyed the nesting property and assigned a larger likelihood to completing their homework in two weeks (relative to two days).[8]

Table 2: Raw earnings—subjective (2018) and projected (NSS 2011–12)

	School		Diploma		College	
Minimum	77.29	(566.0)	156.44	(1,135.9)	81.71	(432.5)
Midpoint	93.97	(439.7)	186.77	(857.2)	128.21	(534.4)
Maximum	110.65	(403.9)	217.10	(880.4)	174.71	(671.6)
Interquartile range	33.35	(440.1)	60.66	(1,091.8)	93.00	(366.3)
Monthly income (subjective)	5.44	(3.475)	6.17	(3.769)	6.26	(3.755)
Weekly income (subjective)	1.84	(1.230)	2.19	(1.351)	2.08	(1.292)
Weekly income (projected)	6.03	(0.492)	6.81	(0.413)	6.55	(0.498)
Probability > midpoint	5.10	(2.820)	5.83	(2.911)	5.77	(2.920)
Test for monotonicity	0.90	(0.294)	0.90	(0.294)	0.90	(0.294)
Observations	408		408		408	

Note: all earning variables are presented as logs. Projected log returns were calculated using the NSS data collected in 2011–12 by the National Statistical and Survey Organisation in India. Standard deviations given in parentheses.

Source: author's construction based on survey results.

As we are keen to study how different social identity groups form earnings expectations, an interesting descriptive is a comparison between the projected and subjective earnings for each of the subgroups. The projected earnings are calculated using the NSS data for Mumbai, corrected for selection bias using Heckman's two-step procedure.[9] Figure A1 in the Appendix shows the standardized projected earnings on the *y*-axis and standardized subjective earnings on the *x*-axis. All the points on the red line passing through the origin (at a 45-degree angle) are when the subjective and projected earnings are equal. Points above the line denote underestimation of earnings and those below it reveal overestimation of earnings. For schooling, we find most of the groups tend to be close to the 45-degree line. Muslim females and male lower caste groups who have lower earnings in the labour market tend to overestimate in their expectations. A similar pattern is observed for college education among SC-ST males and females, where they overestimate their earnings relative to actual income. For diploma-level education, females across caste and religious groups overestimate their earnings, unlike males, even though their projected earnings (compared to the mean) are lower.

3.4 Empirical strategy

In this section, we describe two methods: the OLS or point estimates and the DR to estimate the subjective earnings expectations for different social identity groups. Controlling for other covariates such as parental education, income, child's cognitive capacity, and subjective responses of parents regarding equal opportunities for different socioeconomic groups, the OLS and random effects (RE) specifications are described in Equations 2 and 3:

$$E(y)_{id} = \beta_0 + \beta_1 I_i + \beta_2 X_i' + \beta_3 F_i' + \beta_4 M_i' + \beta_5 ProjY_{id} + \varepsilon_{id} \tag{2}$$

$$E(y)_{idt} = \beta_0 + \beta_1 I_i + \beta_2 X_{it}' + \beta_3 F_i' + \beta_4 M_i' + \beta_5 ProjY_{id} + \varepsilon_{idt} + u_i \tag{3}$$

Equation 2 is the OLS model, which uses the pooled sample of all students who participated in the survey over the two years. y_{id} is the subjective expected earnings for each individual i, and is repeated for each education choice d. Equation 3 is the RE model, which includes all the students who participated in the

[8] Delevande et al. (2011) found that 17 per cent of their sample gave a 50–50 likelihood, with the rest following the nesting rule. Their sample included rural boatmen rather than the adolescent urban students in our sample.

[9] Therefore, we compare the subjective earnings to projected earnings that the groups could receive if there was no identity-based discrimination in the labour market.

survey in both 2017 and 2018. y_{idt} is the subjective expected earnings for each individual in time t and the specification is repeated for each education choice d.

The levels of education are three counterfactual situations: completing school, diploma (technical education), and college. I_{idt} is the social identity (I) variable such as gender (male or female), religion (Hindu or Muslim), and upper or lower caste groups (OBC and SC-ST). X'_{idt} comprises a vector of individual-level characteristics such as age, number of siblings, and cognitive capacity. F'_{id} is a vector of covariates for the father, such as income, education, and subjective opinions on equality of opportunity for different social identities in India, such as gender, caste, and religion. M_{id} are the same measures obtained from the mother and $ProjY_{id}$ is the inflation-adjusted weekly projected earnings calculated using the NSS data for Mumbai.[10]

The OLS and RE frameworks imply that the coefficients β_0 and β_i linearly determine the expectation (average) of the dependent variable. A subgroup analysis or interaction term can be used to observe the average linear effect of a specific covariate. However, both of these techniques have a shortcoming—the former reduces the sample size, and the latter is problematic if the variable that determines membership of the individual (such as gender or religion) is also one of the outcomes of interest Hohberg et al. (2017). As an alternative to point estimates or averages, this study uses DR as a methodological contribution. DR (Hohberg et al. 2017; Stasinopoulos et al. 2018) goes beyond the mean and takes into account variation over the entire distribution of the dependent variable for the covariates of interest. Using a maximum likelihood estimation, the distribution of the outcome y_i can be described as a density function $p(y_i|\theta_{i1},\theta_{i2},...\theta_{ik})$, where $\theta_{i1},...\theta_{ik}$ are k different parameters of the distribution, such as mean, standard deviation, and skewness. For each parameter the following equation can be specified:

$$gk(\theta_{ik})_{dt} = \beta_0^{\theta_k} + \beta_1^{\theta_k} I_{it} + \beta_2^{\theta_k} X'_{it} + \beta_3^{\theta_k} F'_{it} + \beta_4^{\theta_k} M'_{it} + \varepsilon_{id} \tag{4}$$

where gk is the link function that models a nonlinear relationship between the parameter θ and the explanatory variables. After selecting a suitable conditional distribution that fits the outcome variable, the model (with relevant covariates) is estimated using maximum likelihood. The model presents effects of the selected covariates on the distribution parameters. If there are n observations, the model produces n distributions of the outcome variable. In this study, after a suitable distribution is selected based on the data, maximum likelihood estimates three parameters of the distribution: mean, standard deviation, and skewness for each observation in the sample. A bootstrap sample is generated which comprises randomly drawn estimated distributions. The DR is re-estimated for this bootstrap sample. Consider any two 'representative agents' that need to be compared (e.g. female Hindus and Muslims): a conditional distributional comprising the three parameters is obtained for the two groups. An estimate that is a combination of all three moments is computed for each group and the difference is called the estimated marginal treatment effect (estimated MTE).This process is repeated for many bootstrap samples and multiple estimated MTEs are obtained, and a bootstrap percentile confidence interval (at 95 per cent) is computed. In the results section, Table 4 shows the estimates for each of the moments' mean, variance, and skewness. Table 5 includes the estimated MTE for each of the groups of interest, including a 95 per cent confidence interval for the estimate.

In this study, since the outcome of interest is not the direct difference in the treatment and control group, but rather the comparison of an entire distribution, DR can be considered a useful complement to the OLS framework. Although distributional measures (moments) can be individually calculated

[10] We run the specification in Equation 3 for all the pooled sample as well as the panel sample over two years (in 2017 and 2018). We have a sample of 120 students who were surveyed over the two years. Taking advantage of a panel format of the data, we rerun Equation 3 using the RE model, incorporating earnings expectations in 2017 and 2018. Eighty children who were in the 2017 sample dropped out and could not be surveyed in 2018. In the last section of the paper, I discuss the impact of dropouts for the results in the study.

and used as dependent variables, the advantage of DR is that it yields one model from which several distributional measures can be obtained. Thus, consistent and comparable measures are estimated since it is from the same model. Finally, as Hohberg et al. (2017) point out, if the aim of the study is to only compare distributions of the outcome variable for the treatment and control, we do not require DR, since we can compare the histograms or kernel densities. However, since this study does not compare only treatment and control groups but estimates the effects of nonlinear covariates in the final analysis, the implementation of the method is a useful contribution to the literature.

4 Results

4.1 Mean effects

In this section, linear regression techniques are used to estimate the mean effect of different identity groups on their subjective earnings expectations. For more robust estimates of how past actual earnings and different social identity groups explain subjective earnings expectations, we estimate OLS and RE models using Equations 2 and 3. The OLS specification includes seemingly unrelated coefficients based on the pooled data,[11] while the RE model uses the panel data composed of students who answered the survey in both 2017 and 2018.[12]

Table 3 shows OLS and RE estimations for each level of education. Considering the OLS models, the inflation-adjusted projected earnings are not correlated with subjective earnings expectations for the children (the descriptives in Figure A2 show a similar result). This finding is also observed in previous studies (Attanasio and Kaufmann 2009; Huntington-Klein 2015; Jensen 2010) across other student samples.

For the OLS models, females expect to earn significantly lower wages after completing a diploma or college. This is not the case for completing high school. For the average estimates, Muslim children expect significantly lower earnings compared to the Hindu males only for diploma. Comparing the two lower caste groups (OBC and SC-ST), we reject the hypothesis that the latter expect lower earnings for higher education levels compared to the upper caste groups. Specifically for college education, we find OBC and SC-ST expect significantly higher subjective earnings relative to the reference (General/upper caste) group. Columns 2, 4, and 6 are RE models that only consider children who participated in the survey in both years. Unlike Huntington-Klein (2015), we are able to control for individual unobservables over a period of two years by using a panel dataset. However, this panel sample entails the risk of not considering the dropouts. If we include only those students who continued their education, we do not find significantly lower earnings expectations for females or Muslims. However, Table A2 in the Appendix is a linear probability model that shows that both females and Muslims are significantly more likely to drop out. Therefore,it is likely that the lower earnings expectations for females and Muslims are driven by the students who dropped out in the next year. These implications of the dropout are discussed in the last section of the study.

[11] Keeping in the mind the error terms of the three models (high school, diploma, and college) can be correlated, we report the coefficients for seemingly unrelated regressions.

[12] We use the RE model (and not fixed effects) since we are interested in studying between-individual time-invariant characteristics such as identity. In fixed effects models, the model controls for these unobservables and only shows the within-individual variation over time.

Table 3: Subjective expectations: OLS and RE

	School		Diploma		College	
	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se	(6) b/se
Female	–0.044	0.106	–0.201***	0.069	–0.154**	0.040
	(0.076)	(0.153)	(0.077)	(0.204)	(0.070)	(0.180)
Muslim	–0.017	–0.104	–0.250**	–0.029	0.019	–0.031
	(0.117)	(0.202)	(0.106)	(0.296)	(0.075)	(0.228)
Age	–0.007	–0.015	0.018	0.106	–0.003	0.027
	(0.023)	(0.092)	(0.028)	(0.132)	(0.026)	(0.112)
OBC (low caste)	0.006	0.145	0.082	0.235	0.253***	0.262
	(0.114)	(0.228)	(0.088)	(0.267)	(0.077)	(0.248)
SC-ST (low caste)	–0.142	–0.343	–0.082	0.023	0.351**	0.137
	(0.203)	(0.358)	(0.137)	(0.322)	(0.141)	(0.352)
Cognition	–0.023	–0.033	0.062**	–0.065	–0.033	–0.014
	(0.021)	(0.039)	(0.025)	(0.053)	(0.027)	(0.051)
Father education	0.021*	0.030	0.016	0.016	–0.005	0.035
	(0.011)	(0.024)	(0.011)	(0.031)	(0.009)	(0.029)
Mother education	0.009	0.018	–0.013	0.021	0.005	0.024
	(0.009)	(0.021)	(0.012)	(0.027)	(0.010)	(0.025)
Actual earnings (NSS)	–0.082	–0.242	–0.059	0.017	0.082	–0.091
	(0.127)	(0.205)	(0.077)	(0.176)	(0.066)	(0.185)
2017		0.093		0.129		0.008
		(0.103)		(0.145)		(0.138)
Constant	1.768**	2.602	1.537**	–0.336	1.042*	1.066
	(0.877)	(1.943)	(0.686)	(2.287)	(0.582)	(2.084)
Observations	376	222	356	208	376	222
R^2	0.016		0.047		0.023	
Adj. R^2	0.004		0.034		0.011	
R^2 B		0.099		0.035		0.078

Note: columns 1, 3, and 5 are pooled OLS models controlling for past actual earnings. Columns 2, 4, and 6 are RE models and include the sample of students who reported earnings expectations in both 2017 and 2018. The dependent variable for the above specifications is the log subjective earnings expectations. The OLS models are seemingly unrelated regressions. Robust standard errors in parentheses. b is the average difference in earnings expectations between each independent variable and the reference group. Robust standard errors are in the parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Source: author's calculations based on survey results.

Other important controls such as higher cognitive capacity measured by the Raven's matrix are positively correlated with earnings expectations for diploma education. Years of education of the father has a positive and significant effect on the earnings expectations for school (only in the OLS models). Actual earnings in the labour market are not correlated with present subjective expectations. As a robustness check, Table A3 in the Appendix includes a pooled model for all three levels of education. We find similar results, whereby females have significantly lower earnings expectations in the OLS model. However, there is no significant difference in earnings expectations for females in the RE model that does not include the dropouts. For Muslim children, we do not find a significant difference in earnings expectations relative to the Hindu male upper caste. Children from the OBC caste group have significantly positive earnings in the pooled OLS and RE models. Thus, both specifications (pooling all the education choices and analysing them separately) show similar patterns for mean estimates. However, given the significance of diploma as a technical course that offers immediate employment relative to college education and has higher actual returns, we think it is important to consider each education level as a counterfactual. Therefore, the study focuses on presenting the mean effects for each education level independently.

4.2 Distributional effects

In the previous section the estimates showed that, on average, females have lower subjective returns in the pooled sample, but we do not find significant mean differences for the panel sample. Other disadvantaged groups such as Muslims do not show significantly different returns. Similarly, OBC and SC-ST groups show significantly higher earnings relative to male upper caste groups for college education in the OLS specification. As described in Section 3.3, when the analysis does not entail a clean treatment and control comparison, and other explanatory variables are used in the final analysis, there are shortcomings where only point estimates are provided (Hohberg et al. 2017). This is true for the current study. As the objective of the paper is to observe the effect of explanatory variables that are not independent of the outcome, one way of conducting robust (complementary) analysis would entail not only the mean but also other moments of the distribution (e.g., standard deviation and skewness).

We include a complementary analysis known as the distributional regression. Based on the gamlss code in R, we first observe which distribution fits the outcome variable (subjective earnings). We compare the log normal and Burr distribution, which are distributions frequently used to measure income. Figures A6 and A7 are the diagnostic plots for each distribution. The plots show the Burr distribution fits the data better than the log normal distribution as the latter has a distribution with a heavier right tail compared to normalized quantile residuals. Using this preliminary evidence as support to continue the analysis, we assuming the data to have a Burr distribution (Hohberg et al. 2017).

As a first step, we obtain the estimates for the moments of the distribution for the groups of interest. Table 4 shows the estimates of the three moments: mean, standard deviation, and skewness for each education choice. Apart from the familiar mean effect, standard deviation shows the uncertainty or variance of the expected earnings and skewness reveals the extent to which the distributions of the two comparison groups are significantly left- or right-skewed. In the second step, we calculate the MTE—that is, the difference in subjective earnings between any two comparative groups evaluated at the mean values for other explanatory variables.[13] The MTE is a combined estimate based on three moments forming a conditional distribution—that is, θ_{ik} from Equation 4. The MTE is shown in Table 5. Both the moments and the MTE based on these moments are calculated for the panel sample of students over the two years.[14] We will first discuss the estimates for each of the moments of the groups we are interested in. Next, we will discuss the combined effect of these moments as the MTE.

In Table 4, we find females have significantly lower mean earnings expectations for school and college. The standard deviation is significant and positive for the females across all groups compared to the Hindu male upper caste group—that is, there is significantly more variation (uncertainty) in the responses of the females. Finally, for school and diploma education choices, females have significantly more left-skewed distributions. Although Muslims do not show significantly different mean earnings expectations (except for diploma, which is positive at 0.194), their earnings expectations are significantly more left-skewed for school and diploma compared to the Hindu male upper caste. Particularly, for school and diploma, both females and Muslims have significantly left-skewed earnings distributions compared to upper caste Hindu males. There is no significant differences in earnings expectations for the OBCs. However, the SC-ST group show, on one hand, higher mean earnings expectations and, on the other hand, significantly left-skewed earnings distributions in the school and diploma education choices.

[13] The other variables includes controls used in the OLS such as parent's education, cognitive ability of the student, and past actual earnings.

[14] The DR with the pooled sample showed similar results. Since the RE model is more robust, we include the latter in the paper.

Table 4: Distributional regression: mean, variation, and skewness

	School			Diploma			College		
	Mean	Std dev.	Skewness	Mean	Std dev.	Skewness	Mean	Std dev.	Skewness
Female	–0.006*	0.39***	–0.43***	–0.105	0.262***	–0.33***	–0.196*	0.131*	–0.174
	(0.06)	(0.05)	(0.092)	(0.07)	(0.052)	(0.099)	(0.081)	(0.052)	(0.102)
Muslim	0.065	–0.11	–0.255*	0.194*	0.08	–0.44***	–0.096	0.057	–0.096
	(0.064)	(0.055)	(0.10)	(0.088)	(0.057)	(0.11)	(0.089)	(0.059)	(0.11)
OBC	0.057	0.051	–0.15	–0.052	–0.073	0.099	–0.126	–0.029	–0.088
	(0.063)	(0.054)	(0.097)	(0.085)	(0.057)	(0.108)	(0.087)	(0.057)	(0.11)
SC-ST	0.243*	0.138	–0.359*	0.402***	0.33**	–0.59***	–0.027	–0.005	–0.015
	(0.108)	(0.088)	(0.165)	(0.119)	(0.094)	(0.166)	(0.14)	(0.089)	(0.173)
Constant	0.26*	1.65***	–0.983***	0.508***	1.233***	–0.54*	0.532	1.249***	–0.609**
	(0.106)	(0.172)	(0.234)	(0.132)	(0.137)	(0.215)	(0.145)	(0.151)	(0.234)
Observations	652			632			652		
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Note: dependent variable is the log earnings expectation. Robust standard errors are in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$

Source: author's calculations based on survey results.

Table 5: Distributional regression

	Difference MTE	95 per cent confidence intervals	
Gender	Male–Female	LB 0.025	UB 0.975
School	0.224	0.007	0.437
Diploma	0.342	0.059	0.507
College	0.326	0.074	0.399
Religion	Hindu–Muslim	LB 0.025	UB 0.975
School	0.224	0.0013	0.306
Diploma	0.209	0.027	0.403
College	0.18	–0.0427	0.459
Caste	General–OBC	LB 0.025	UB 0.975
School	0.433	–0.048	1.384
Diploma	0.357	–1.397	1.609
College	–0.325	–1.72	1.639
Caste	General–SC-ST	LB 0.025	UB 0.975
School	0.282	–0.166	1.38
Diploma	0.358	–1.526	1.557
College	–0.559	–1.867	1.639

Note: the difference MTE is the difference in log earnings expectations between male–female, Hindu–Muslim, and General–OBC and SC-ST, respectively. This difference is calculated for school, diploma, and college. Columns 2 and 3 are 95 per cent confidence intervals for the MTE. LB and UB are the upper and lower bounds of the CI. The difference in MTE is statistically significant at the 95 per cent confidence interval only if the bounds do not contain 0.

Source: author's calculations based on survey results.

In Table 5, we estimate the MTE, which is a combined estimate of the three moments. In the first panel, we compare males and females, where column 1 is the MTE of being a male and column 2 is the 95 per cent bootstrap confidence intervals for the MTE. We can interpret the MTE as follows: for a male (at the mean values of all other control variables), the subjective earnings expectation is higher compared to a female by 0.224 units. This MTE is statistically significant at the 95 per cent confidence interval for all three education levels. Similarly, considering panel 2 (religion), we find the subjective expectations are significantly higher among Hindus for school and diploma, but not at the college level. The third and fourth panels compare OBC and SC-ST caste groups to the General caste. We find the MTEs are not significantly different across all education choices. Table A4 in the Appendix further disaggregates the groups for males and females, respectively.

Females show significantly lower MTEs across all levels of education, and this result is driven by not only the mean but also the skewness of the distribution. For Muslims, even though the mean earnings expectations are not significantly different, the MTE is lower for Muslims due to their left-skewed distributions. For the OBC, there is no significant difference in the MTE, and this is also observed across the estimates of the moments. However, for the SC-ST, while we do not find significant differences in the MTE, we observe their mean earnings expectations to be higher while their earnings distribution is significantly left-skewed. Comparing the OLS and DR techniques, we find combining three moments that form a distribution provides more robust estimates compared to the linear OLS. Especially for females, we now observe their subjective earnings expectations to be significantly lower than those of males across all three education levels. Similarly, for Muslims we observe significantly lower MTEs compared to Hindus for school and diploma. For low caste groups (OBCs and SC-ST), the results do not change.

5 Discussion and mechanisms

The primary objective of this study is to provide a descriptive picture of how students from social identities that are historically discriminated in the labour market tend to form subjective expectations regarding their returns from human capital investment. Particularly when estimating different moments of the distribution, we find females and Muslims to have a lower earnings expectations for school and diploma, which no longer exist at the college level for Muslims. The discriminated caste groups, on the other hand, do not expect significantly different earnings relative to the upper caste group. In this section, potential moderating factors are discussed, such as parents' opinions on gender equality, non-monetary aspirations for education, knowledge of policies such as AA, and the strong presence of local leaders that could impact children's expectations of earnings. In addition, concerns regarding the study are also highlighted.

5.1 Gender

Controlling for cognitive ability and parents' education, girls in the sample consistently have significant and lower earnings expectations for school and diploma education compared to boys. The MTE from the DR also shows that females have significantly lower earnings expectations compared to males (Table 5). Contrasting this, Table A1 in the Appendix shows a greater percentage of girls have higher educational aspirations (44 per cent of girls compared to 31 per cent of the boys aspire to complete college education, $p = 0.000$).

For this sample, we find girls to have higher non-monetary aspirations from education and schooling, but they do not expect this education to translate into higher monetary benefits. Women in India are found to have a 12 per cent lower likelihood of being employed compared to men. Even those in employment earn annual incomes that are 36 per cent lower than those of men, *ceterius paribus* (Bhandari and Bordoloi 2006; Mitra 2019). Studies by Maertens (2011), Dréze and Kingdon (2001), Kingdon (2020), and Kingdon and Theopold (2008) show that the social norm of women getting married between the ages of 18 and 23 further deters their parents from investing in their education. This, in turn, may likely impact the girls' perspectives on lower benefits from completing higher education, but seeing a greater non-monetary value of education (e.g. gains in the marriage market).

In the survey of the parents, we correlated the opinions of the parents on gender equality and the student's self-reported earnings expectations. We asked the parents of the children in a phone survey what were their opinions on a positive and negative statement on gender equality. They had to answer whether they agreed or disagreed with each statement. Based on the study by Dhar et al. (2018) that observed a positive correlation between parents' attitudes towards gender equality and their children's attitudes,

we expected the girls whose parents support gender equality would have higher subjective earnings expectations across all three levels of education. Figure A3 in the Appendix shows in the *y*-axis the mean expected earnings for boys and girls. However, we do not find a correlation between opinions on equality and higher earnings expectations for girls.

5.2 Religion

The DR shows Muslim children have lower earnings expectations compared to Hindu children for schooling and diploma education. While this might not be evident in the point estimates (Table 3), the DR reveals significantly lower earnings for Muslims, driven by their left-skewed distributions (Table 4). Muslims, unlike the OBC and SC-ST groups, do not benefit from targeted programmes such as AA. A recent study by Asher et al. (2017) shows that, in terms of intergenerational mobility, while SC-ST and OBC groups have experienced better upward mobility over the last 30 years, Muslims have experienced a drastic decline. In this study a similar trend is observed for the sample of students in Mumbai, whereby earnings expectations and aspirations of Muslims are lower than those of children from disadvantaged caste groups.

5.3 Caste

Comparing the earnings expectations of disadvantaged caste groups—namely OBCs and SC-ST—to the General category, we find both OBC and SC-ST students have marginally higher earnings expectations when considering the pooled sample. However, the RE and DR models show no difference in earnings expectations. Despite facing extensive discrimination in the labour market not only in terms of earnings gap but also the type of employment opportunities (Deshpande 2012; Madheswaran and Attewell 2007),[15] we find students from the SC-ST caste groups to have no difference in earnings expectations. Furthermore, the SC-ST groups overestimate their earnings compared to the inflation-adjusted projected earnings.

Although widely debated (for a detailed discussion, see Deshpande (2012) and Deshpande (2013)), the study postulates a potential role of AA in favour of the OBCs and SC-ST groups to explain their high earnings expectations despite glaring labour market discrimination. Figure A4 plots the log earnings expectations for all caste groups depending on whether they are aware of the policy of AA and its benefits in education and the labour market. However, this is not statistically significant.[16]

Another explanation for high earnings expectations of lower caste students can be attributed to the strong presence of OBC and Dalit (SC-ST) political movements in Mumbai. Many parts of Mumbai have experienced extensive Dalit and lower caste mobilization, such as the Dalit Panthers in the 1970s (Collins 2017; Deshpande 2013; Pawar 2018),[17] and the emergence of local leaders.

---

[15] The projected earnings calculated using the NSS for 2011–12 shows SC-ST have significantly lower rates of return across all levels of education.

[16] Table A5 in the Appendix is an OLS result and shows the total effect of the interaction term of awareness regarding AA and the caste category. The first two rows are the total effect of the interaction between knowledge of AA policies for a OBC student relative to the General caste. The third and fourth rows are the mean estimates from knowledge about AA for an SC-ST student compared to the General caste. Controlling for individual and parent covariates, I find that both SC-ST and OBC students whose parents have knowledge about AA policies report higher levels of earnings expectations compared to the general or upper caste category. However, these estimates in Table A5 are not statistically significant.

[17] The movement emerged as a voice against the oppression of the Dalits (SC-ST) and the need for change.

Figure A5 plots the log earnings expectations of students from each caste group depending on the caste of their local leader.[18] We do find a positive correlation ($p = 0.05$) between earnings expectations of SC-ST students and the caste identity of the local leader being the same. Similarly, when the local leader is from the general caste, the earnings expectations of SC-ST students are significantly lower ($p = 0.02$). Future studies can explore these mechanisms by including diverse samples such as low caste groups in rural areas and urban cities that have not benefited from strong Dalit movements.

5.4 Further concerns

In this section, I point out some concerns regarding this study. As described earlier, the survey was implemented for students in Mumbai over two years: 2017 and 2018. In an attempt to create a panel dataset, we found 80 students out of the 324 surveyed in 2017 to have dropped out. Table A2 in the Appendix is a linear probability model on the likelihood of dropping out. Over two years, the likelihood of dropping out was higher for females, Muslims, and OBC females. A child's father's education significantly decreased the likelihood of dropping out for the pooled sample and for boys. There is a significant decrease in dropping out for older male students. Particularly in this sample, it could reflect the opportunity cost of dropping out, which is higher for older students as they are closer to completing their high-school degree.

A notable concern that is likely to cause a downward bias in the earnings expectations of females is the working age stated in the question. All students were asked how likely they are to work at the age of 28 after completing education. It could be that most girls do not expect to be working at 28 years old (on account of marriage and family responsibilities). One way to observe this is to see how many females said they do not expect to work, for a given education level, at 28 years old. Prior to asking the earnings question, I had asked if the students were likely to work. In this sample, nine females said they would not be working after completing any of the education levels.[19]. Although the results hold without including the nine students, the high working age stated in the survey for females must be considered when looking at their earnings expectations.

6 Conclusion

Based on both the OLS and DR methods, females and Muslims (particularly at school and diploma levels) expected significantly lower earnings compared to upper caste males. However, students from disadvantaged caste groups such as the OBCs and SC-ST did not have different earnings expectations compared to the upper caste groups. The DR revealed other moments of the earnings distribution, such as uncertainty (variance) and left-skewness, to be prominent estimates driving the lower earnings expectations for females and Muslims.

Additionally, comparing projected and subjective earnings for each of the subgroups, we observe SC-ST students to overestimate their earnings compared to what they could earn in the labour market. On the other hand, advantaged groups such as male upper caste students, who experience above-average projected earnings, underestimate what they could earn in the labour market. An important take-away from this result is reassessment of demand-side interventions that provide information about earnings to households and individuals as a potential policy option. While females and Muslims (who underestimate

[18] The local leader is not at the individual level. Given that many students come from a similar locality, they also have a common local leader or MLA (Member of the Legislative Assembly). As a result, our sample size is small.

[19] Out of the nine females, four said they would not work after school, two each would not work after diploma and college, and one claimed not to be working after both diploma and college

their earnings in the labour market) could benefit from such information interventions, it may not be beneficial to disadvantaged caste groups who have high aspiration levels.

This study shows that existing educational inequalities have been internalized by females and Muslim students. In order for these groups to expect gains and to actively participate in the education and labour markets, they will require significantly higher aspirations. On the other hand, lower caste groups, which have been historically discriminated, do not seem to update their earnings expectations based on the discrimination in the labour market. Particularly in this study, the disadvantaged caste students belong to urban areas where the discrimination might not be explicit at the school or household levels. Additionally, SC-ST and OBC groups have been able to assert their rights through collective political and social mobilization over recent years. Future studies could undertake a similar exercise among lower caste students from rural areas. Finally, in order to move closer to the Sustainable Development Goals, such as 'Target 4: Ensure inclusive and equitable quality education for all', it is also essential to urge for policy interventions on the supply side that encourage inclusive and accessible education from the primary levels, and remove barriers for entry into higher levels of education and employment.

References

Asher, S., P. Novosad, and C. Rafkin (2017). 'Estimating Intergenerational Mobility with Coarse Data: A Non Parametric Approach'. Working Paper. Washington, DC: World Bank.

Attanasio, O., and B. Augsburg (2016). 'Subjective Expectations and Income Processes in Rural India'. *Economica*, 83(331): 416–42. https://doi.org/10.1111/ecca.12190

Attanasio, O., and K. Kaufmann (2009). 'Education Choices, Subjective Expectations and Credit Constraints'. NBER Working Paper 15087. Cambridge, MA: NBER. https://doi.org/10.35648/20.500.12413/11781/ii120

Attanasio, O., and K. Kaufmann (2014). 'Education Choices and Returns to Schooling: Intrahousehold Decision Making, Gender and Subjective Expectations'. *Journal of Development Economics*, 109: 203–16. https://doi.org/10.1016/j.jdeveco.2014.04.003

Avery, C., and T.J. Kane (2004). 'Student Perceptions of College Opportunities: The Boston COACH Program'. In: C.M. Hoxby (ed.) *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*. Chicago, IL: Chicago University Press.

Bhandari, L., and M. Bordoloi (2006). 'Income Differentials and Returns to Education'. *Economic and Political Weekly*, 41(36): 3893–900.

Bourguignon, F., F.H.G. Ferreira, and M. Walton (2007). 'Equity, Efficiency and Inequality Traps: A Research Agenda'. *Journal of Economic Inequality*. https://doi.org/10.1007/s10888-006-9042-8

Chari, A.V., and A. Maertens (2014). 'Gender, Productive Ability and the Perceived Returns to Education: Evidence from Rural India'. *Economics Letters*, 122(2): 253–57. https://doi.org/10.1016/j.econlet.2013.10.029

Collins, M. (2017). 'Writing Dalit Assertion: Early Dalit Panther Politics and Legal Advocacy in 1980s Tamil Nadu'. *Contemporary South Asia*, 25: 238–54. https://doi.org/10.1080/09584935.2017.1350142

Connor, C.O. (1999). 'Race, Class, and Gender in America: Narratives of Opportunity Among Low- Income African American Youths'. *Sociology of Education*, 72(3): 137. https://doi.org/10.2307/2673226

Cunha, F., and J.J. Heckman (2007). 'Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling'. *Labour Economics*, 14: 870–93. https://doi.org/10.1016/j.labeco.2007.06.002

Deininger, K., S. Jin, and H. Nagarajan (2013). 'Wage Discrimination in India's Informal Labor Markets: Exploring the Impact of Caste and Gender'. *Review of Development Economics*. https://doi.org/10.1111/rode.12020

Delavande, A. (2009). 'Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence'. World Bank Policy Research Working Paper 4824. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-4824

Delavande, A., and S. Rohwedder (2008). 'Eliciting Subjective Probabilities in Internet Surveys'. *Public Opinion Quarterly*, 72(5): 866–91. https://doi.org/10.1093/poq/nfn062

Delevande, A., X. Gine, and D. McKenzie (2011). 'Eliciting Probabilistic Expectations with Visual Aids in Developing Countries: How Sensitive Are Answers to Variations in Elicitation Design?'. *Journal of Applied Econometrics*, 212: 193–212. https://doi.org/10.1596/1813-9450-5458

Desai, S., and V. Kulkarni (2008). 'Changing Educational Inequalities in India in the Context of Affirmative Action'. *Demography*, 45: 245–70. https://doi.org/10.1353/dem.0.0001

Deshpande, A. (2012). *The Grammar of Caste*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198072034.001.0001

Deshpande, A. (2013). *Affirmative Action in India*. Oxford: Oxford University Press.

Dhar, D., T. Jain, and S. Jayachandran (2018). 'Intergenerational Transmission of Gender Attitudes: Evidence from India'. *Journal of Development Studies*, 55(12): 2572–92. https://doi.org/10.1080/00220388.2018.1520214

Dizon-Ross, R. (2019). 'Parents' Perceptions and Children's Education: Experimental Evidence from Malawi'. *American Economic Review*, 109(8): 2728–65.

Dominitz, J., and C.F. Manski (1997). 'Using Expectations Data to Study Subjective Income Expectations'. *Journal of the American Statistical Association*, 92(439): 855–67. https://doi.org/10.1080/01621459.1997.10474041

Dominitz, J., and C.F. Manski (2006). 'Measuring Pension-Benefit Expectations Probabilistically'. *Labour*, 20(2): 201–36. https://doi.org/10.1111/j.1467-9914.2006.00343.x

Dréze, J., and G.G. Kingdon (2001). 'School Participation in Rural India'. *Review of Development Economics*. https://doi.org/10.1111/1467-9361.00103

Dreze, J., and A. Sen (2001). *Basic Education as a Political Issue in India*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199257492.003.0005

Duraisamy, P., and M. Duraisamy (2017). 'Social Identity and Wage Discrimination in the Indian Labour Market'. *Economic and Political Weekly*, 52(4).

Dyson, T., and M. Moore (1983). 'On Kinship Structure, Female Autonomy, and Demographic Behavior in India'. *Population & Development Review*, 9(1): 35–60. https://doi.org/10.2307/1972894

Fasih, T., G. Kingdon, H. Patrinos, C. Sakellariou, and M. Soderbom (2012). 'Heterogeneous Returns to Education in the Labor Market'. World Bank Policy Research Working Paper WPS6170. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-6170

Goel, D., and A. Deshpande (2016). 'Identity, Perceptions and Institutions: Caste Differences in Earnings from Self-Employment in India'. IZA World of Labor Working Paper 10198. Bonn: IZA.

Guiso, L., T. Jappelli, and L. Pistaferri (2002). 'An Empirical Analysis of Earnings and Employment Risk'. *Journal of Business & Economics Statistics*, 20: 241–53. https://doi.org/10.1198/073500102317351985

Henderson, D.J., S.W. Polachek, and L. Wang (2011). 'Heterogeneity in Schooling Rates of Return'. *Economics of Education Review*, 30(6): 1202–14. https://doi.org/10.1016/j.econedurev.2011.05.002

Hohberg, M., P. Peter, and T. Kneib (2017). 'Treatment Effects Beyond the Mean: A Practical Guide Using Distributional Regression'. University of Göttingen Working Paper. Göttingen: University of Göttingen.

Huntington-Klein, N. (2015). 'Subjective and Projected Returns to Education'. *Journal of Economic Behavior and Organization*, 117: 10–25. https://doi.org/10.1016/j.jebo.2015.05.005

Jensen, R. (2010). 'The (Perceived) Returns to Education and the Demand for Schooling'. *Quarterly Journal of Economics*, 125(2): 515–48. https://doi.org/10.1162/qjec.2010.125.2.515

Kabeer, N. (2014). 'Social Justice and the Millennium Development Goals: The Challenge of Intersecting Inequalities'. *Equal Rights Review*, 13: 91–116.

Kingdon, G.G. (2020). 'The Private Schooling Phenomenon in India: A Review'. *Journal of Development Studies*. https://doi.org/10.1080/00220388.2020.1715943

Kingdon, G., and N. Theopold (2008). 'Do Returns to Education Matter to Schooling Participation?' Working Paper GPRG-WPS-052. Oxford: ESRC Global Poverty Research Group.

Kingdon, G.G., and J. Unni (2001). 'Education and Women's Labour Market Outcomes in India'. *Education Economics*, 9(2): 173–95. https://doi.org/10.1080/09645290110056994

Luseno, W.K., J.G. McPeak, C.B. Barrett, P.D. Little, and G. Gebru (2003). 'Assessing the Value of Climate Forecast Information for Pastoralists: Evidence from Southern Ethiopia and Northern Kenya'. *World Development*, 31(9): 1477–94. https://doi.org/10.1016/S0305-750X(03)00113-X

Madheswaran, S., and P. Attewell (2007). 'Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey'. *Economic and Political Weekly*, 42(41): 4146–53.

Maertens, A. (2011). 'Does Education Pay Off? Subjective Expectations on Education in Rural India'. *Economic and Political Weekly*, 46(9): 58–63.

McKenzie, D., J. Gibson, and S. Stillman (2013). 'A Land of Milk and Honey with Streets Paved with Gold: Do Emigrants Have Over-Optimistic Expectations about Incomes Abroad?' *Journal of Development Economics*, 102: 116–27. https://doi.org/10.1016/j.jdeveco.2012.01.001

Mitra, A. (2019). 'Returns to Education in India: Capturing the Heterogeneity'. *Asia and the Pacific Policy Studies*, 6(2): 151–69. https://doi.org/10.1002/app5.271

Pawar, J. (2018). *Dalit Panthers: An Authoritative History*. New Delhi: Forward Press.

Pike, D.W. (2011). *Crimes Against Women*. New York: Nova Science.

Sarkar, S., S. Sahoo, and S. Klasen (2019). 'Employment Transitions of Women in India: A Panel Analysis'. *World Development*, 115: 291–309. https://doi.org/10.1016/j.worlddev.2018.12.003

Sharma, S. (2016). 'Achieving Gender Equality in India: What Works, and What Doesn't'. Availale at: https://theconversation.com/achieving-gender-equality-in-india-what-works-and-what-doesnt-67189.

Stasinopoulos, M.D., R.A. Rigby, and, F. De Bastiani (2018). 'GAMLSS: A Distributional Regression Approach'. *Statistical Modelling*, 18(3–4): 248–73. https://doi.org/10.1177/1471082X18759144

Stinebrickner, R., and T.R. Stinebrickner (2014). 'A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout'. *Review of Economic Studies*, 81: 426–72. https://doi.org/10.1093/restud/rdt025

Thorat, S., and K. Newman (2010). *Blocked by Caste: Economic Discrimination in Modern India*. New Delhi: Oxford University Press.

UNESCO (2015). 'Global Education Monitoring Report'. Technical Report. Paris: UNESCO.

World Bank (2014). 'Republic of India—India: Women, Work and Employment'. Technical Report. Washington, DC: World Bank.

Appendix

A1 Tables and figures

Table A1: Highest education level by groups

	Highest class									
	Don't know		High school		Diploma		College		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Gender of student										
Male	122	21.2	90	15.7	170	29.6	193	33.6	575	100.0
Female	70	9.5	86	11.6	340	46.0	243	32.9	739	100.0
Total	192	14.6	176	13.4	510	38.8	436	33.2	1,314	100.0
Caste										
General	59	10.4	56	9.9	311	54.9	140	24.7	566	100.0
OBC	105	18.4	90	15.7	159	27.8	218	38.1	572	100.0
SC-ST	28	15.9	30	17.0	40	22.7	78	44.3	176	100.0
Total	192	14.6	176	13.4	510	38.8	436	33.2	1,314	100.0
Religion										
Muslims	74	14.6	108	21.3	137	27.1	187	37.0	506	100.0
Hindus	118	14.6	68	8.4	373	46.2	249	30.8	808	100.0
Total	192	14.6	176	13.4	510	38.8	436	33.2	1,314	100.0

Source: author's construction based on survey results.

Table A2: Likelihood of dropping out in 2018

	(1) b/se	(2) b/se	(3) b/se
Female	–0.092**		
	(0.042)		
Age	–0.076***	–0.133***	0.033
	(0.024)	(0.033)	(0.033)
Muslim	0.175***	0.167**	0.123**
	(0.046)	(0.082)	(0.054)
Cognitive abilities	0.005	0.024**	–0.060**
	(0.012)	(0.011)	(0.026)
Aspiration	0.027*	0.014	0.034
	(0.014)	(0.019)	(0.023)
OBC (ref: General)	–0.036	0.013	–0.084*
	(0.047)	(0.081)	(0.048)
SC-ST (ref: General)	–0.027	–0.083	0.136
	(0.052)	(0.075)	(0.092)
Education father	–0.011*	–0.021***	–0.006
	(0.006)	(0.008)	(0.008)
Education mother	–0.009	–0.007	0.005
	(0.006)	(0.006)	(0.008)
Minimum earnings	–0.024	–0.019	0.025
	(0.031)	(0.032)	(0.095)
Midpoint	–0.127	–0.088	–0.245
	(0.085)	(0.094)	(0.232)
Maximum earnings	0.140**	0.108	0.181
	(0.061)	(0.070)	(0.158)
Constant	1.365***	2.157***	–0.020
	(0.332)	(0.499)	(0.401)
Observations	351	213	138
R^2	0.182	0.297	0.229
Adj. R^2	0.153	0.259	0.161

Note: the linear probability model estimates the likelihood of dropping out based on the outcomes used in the final analysis. The dependent variable for all three models is a dummy (1 = drop, 0 = did not drop). Column 1 is pooled over gender, Column 2 and 3 are separate estimates for males and females, respectively. Robust standard errors in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Source: author's calculations based on survey results.

Table A3: Subjective earnings—pooled

	Model 1	Model 2	Model 3
	(1)	(2)	(3)
	b/se	b/se	b/se
Female	–0.123**	–0.099*	0.058
	(0.052)	(0.056)	(0.136)
Diploma (ref: school)	0.154**	0.213***	0.265***
	(0.061)	(0.068)	(0.089)
College(ref: school)	0.185***	0.245***	0.253***
	(0.060)	(0.065)	(0.081)
Muslim	–0.044	–0.082	–0.040
	(0.061)	(0.066)	(0.118)
Age of student	0.013	0.013	0.034
	(0.019)	(0.019)	(0.064)
OBC (ref:General)	0.185***	0.141**	0.256**
	(0.057)	(0.063)	(0.120)
SC-ST (ref:General)	0.074	–0.035	0.028
	(0.072)	(0.103)	(0.175)
Cognitive ability	0.001	0.001	–0.036
	(0.018)	(0.018)	(0.044)
Education father	0.020**	0.020**	0.026
	(0.008)	(0.008)	(0.023)
Education mother	0.002	0.002	0.023
	(0.008)	(0.008)	(0.020)
Actual earnings (NSS 2011)		–0.090	–0.049
		(0.058)	(0.080)
2017			0.075
			(0.088)
Constant	0.919***	1.508***	0.628
	(0.278)	(0.477)	(1.219)
Observations	1473	1473	652
R^2	0.029	0.030	
Adj. R^2	0.022	0.022	
R^2 B			0.093

Note: the dependent variable for all three model specifications is log earnings expectations. Models 1 and 2 are OLS (with and without controlling for projected earnings). Model 3 is an RE model for students who participated in the survey in 2017 and 2018. It controls for projected earnings. Robust standard errors in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Source: author's calculations based on survey results.

Table A4: Distributional regression

	Male			Female		
Religion	H – M	LB 0.025	UB 0.975	H – M	LB 0.025	UB 0.975
School	0.0003	–0.011	0.056	–0.002	–0.01	0.01
Diploma	0.192	0.04	0.28	0.125	0.024	0.16
College	0.25	0.048	0.34	0.145	0.026	0.23
Caste	H – OBC	LB 0.025	UB 0.975	H – OBC	LB 0.025	UB 0.975
School	–0.0004	–0.0009	0.004	0.0001	–0.0001	0.0001
Diploma	0	–0.0001	0.00013	0.00012	–0.00017	0.0001
College	–0.0003	–0.0001	0.00013	–0.00015	–0.0008	0.00045
Caste	H – SCST	LB 0.025	UB 0.975	H – SCST	LB 0.025	UB 0.975
School	0.0003	–0.001	0.0012	–0.0005	–0.001	0
Diploma	–0.0037	–0.023	0.0025	–0.001	–0.0035	0.0003
College	0.00024	–0.0034	0.0013	–0.0016	–0.0077	0.0003

Note: MTE for subsamples at 95 per cent confidence intervals. Each panel is the estimated difference between upper caste and Muslims or disadvantaged castes.

Source: author's calculations based on survey results.

Table A5: Affirmative action policies: OLS

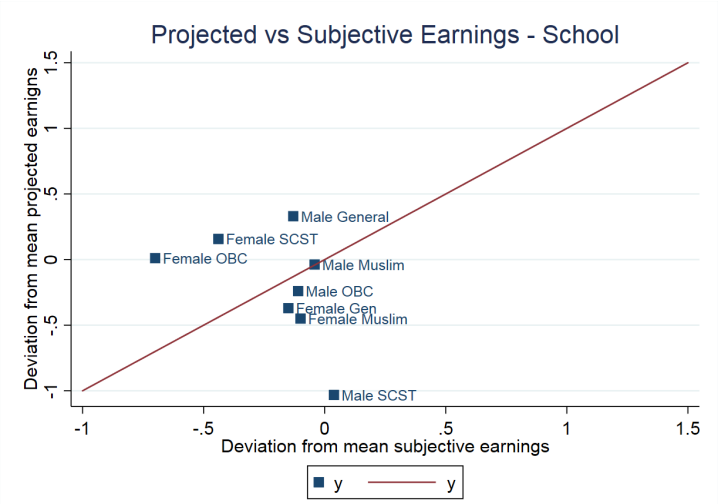
	School	Diploma	College
	(1)	(2)	(3)
No benefit (OBC vs General)	–0.925	0.823	0.868
	(1.61)	(0.719)	(0.599)
Benefit (OBC vs General)	0.143	0.459	0.740
	(0.521)	(0.513)	(0.74)
No benefit (SC-ST vs General)	–1.151	–0.106	0.538
	(1.616)	(0.635)	(0.500)
Benefit (SC-ST vs General)	0.401	0.084	0.901
	(0.691)	(0.681)	(1.056)
Observations	119	119	119
R^2	0.104	0.163	0.074
Controls	Yes	Yes	Yes

Note: dependent variable is the log earnings expectation. Robust standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

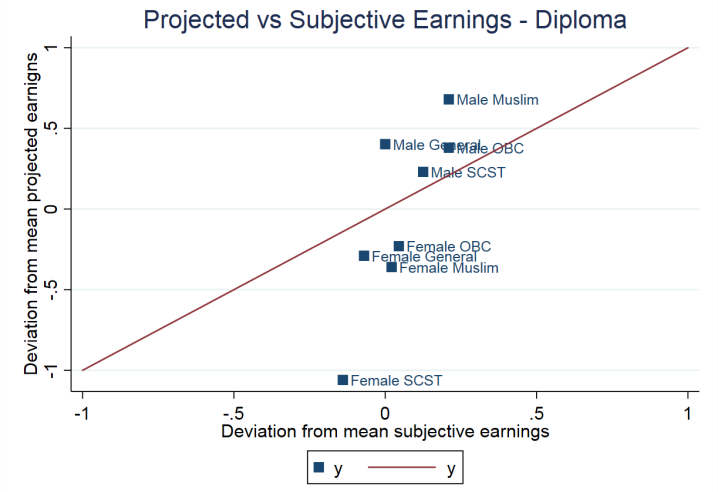
Source: author's calculations based on survey results.

Figure A1: Subjective vs projected earnings: subsample analysis of returns to school, diploma and college education

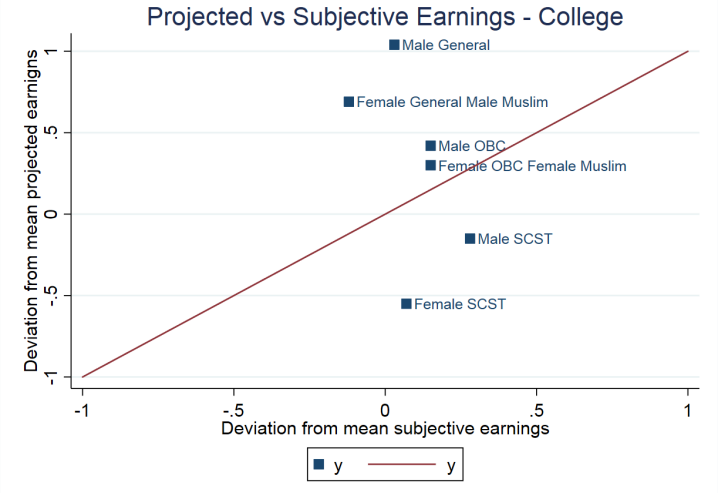
(a) School



(b) Diploma

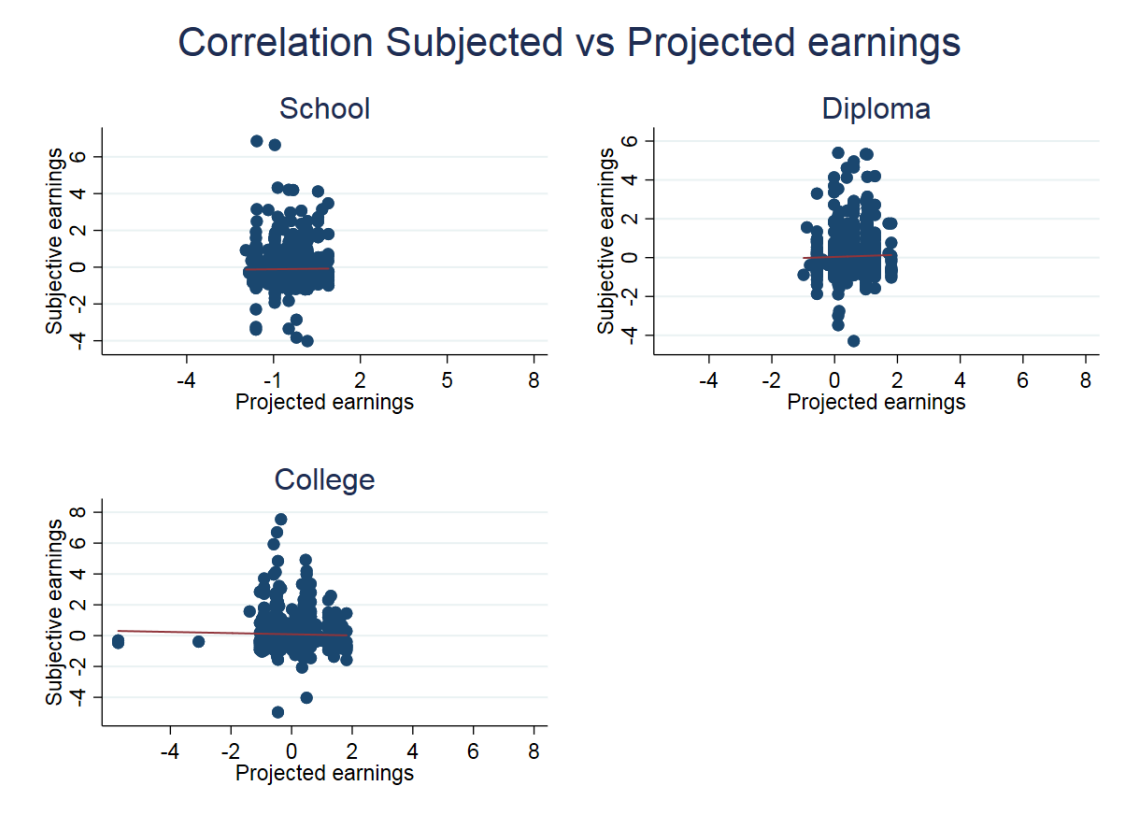


(c) College



Source: author's construction.

Figure A2: Projected vs subjective earnings



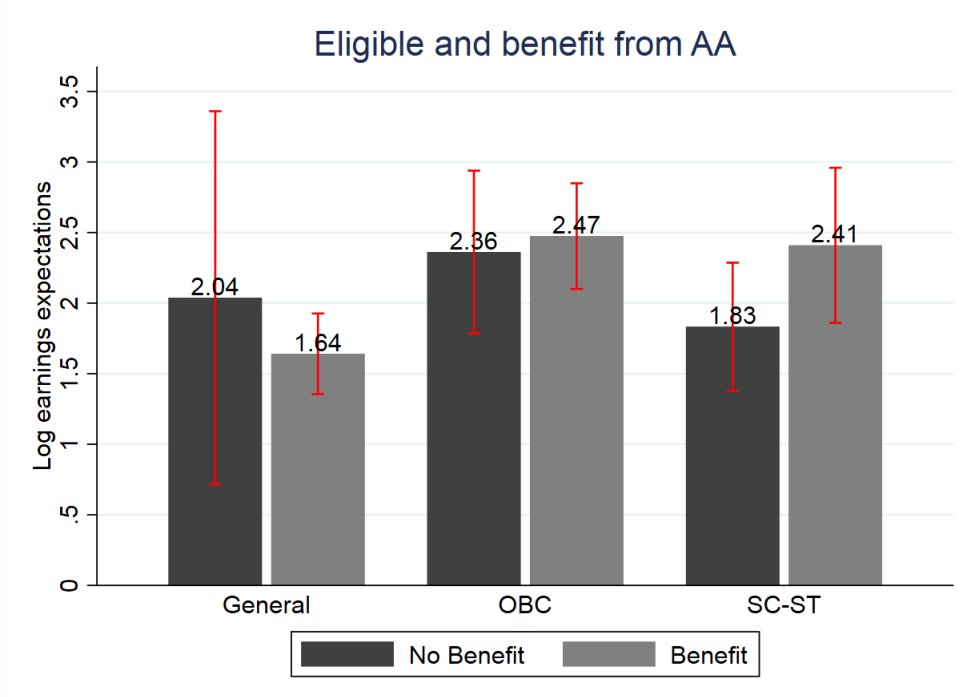
Source: author's construction.

Figure A3: Equal opportunities: gender

Note: Panel A: all girls and boys should have equal opportunities in education and job markets. Panel B: when there are fewer jobs, men have more rights in getting the job than women.

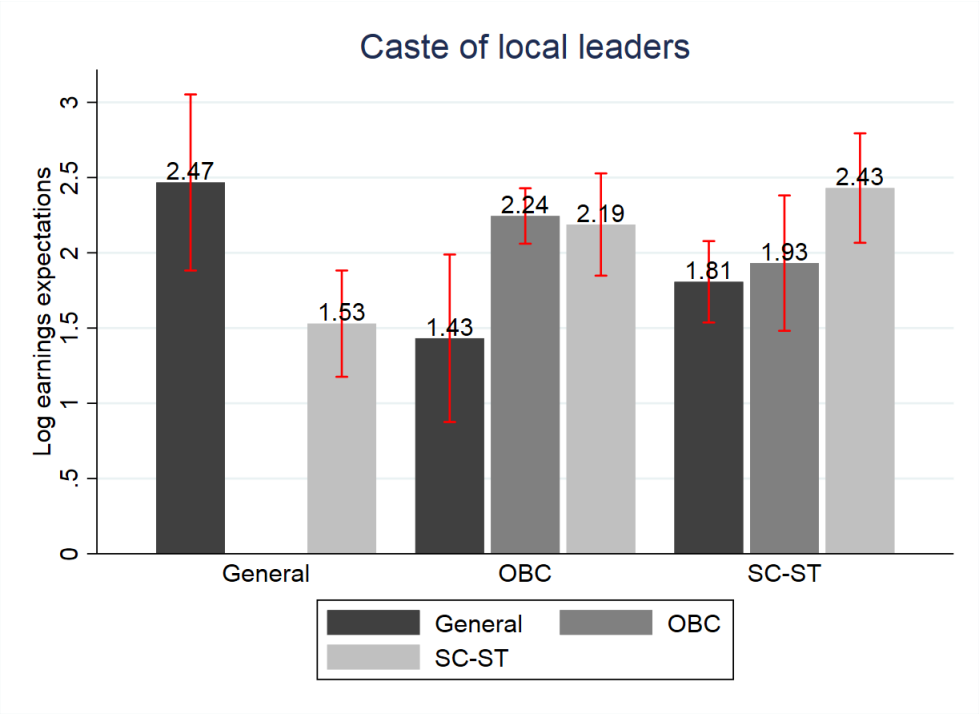
Source: author's construction.

Figure A4: Affirmative action and earnings expectations



Source: author's construction.

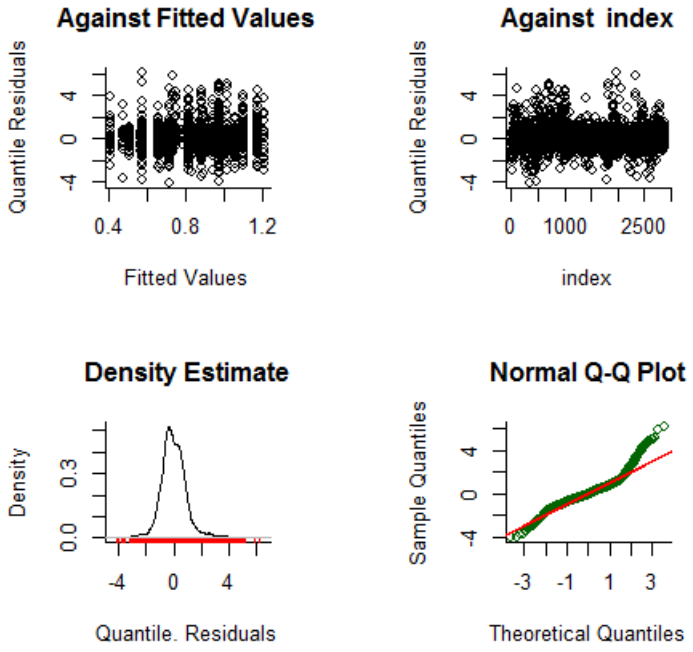
Figure A5: Caste of the local leaders



Note: the bars represent the caste identity of the students. Each of the subgroups in the *x*-axis denotes the caste identity of the local leader. Leader General: General vs SC-ST $p = 0.02$**, Leader OBC: General vs OBC $p = 0.002$***, Leader SC-ST: General vs SC-ST $p = 0.05$*.

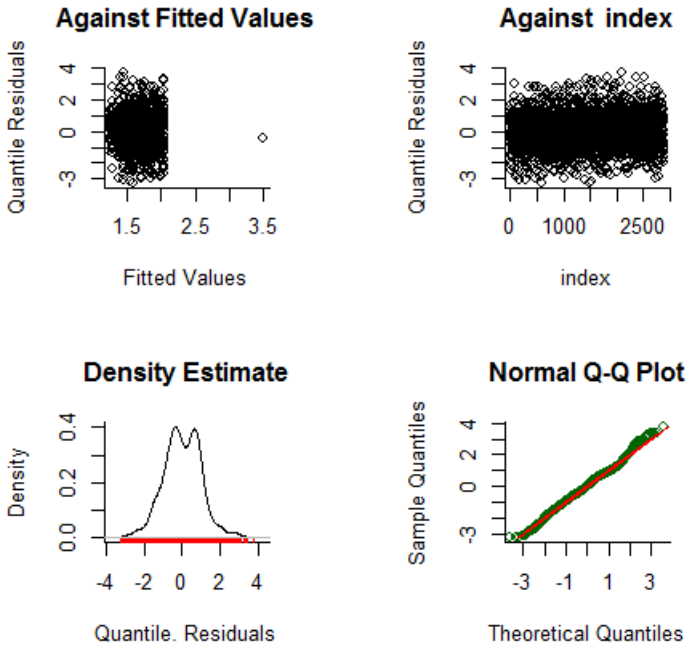
Source: author's construction.

Figure A6: Log normal distribution



Source: author's construction.

Figure A7: Burr distribution



Source: author's construction.

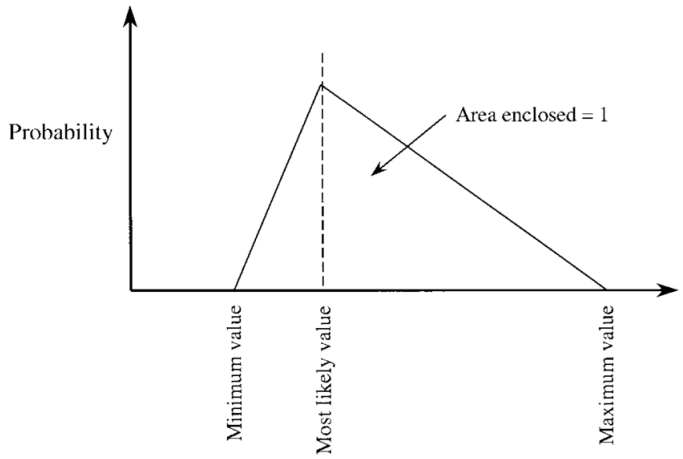
A2 Measuring subjective earnings expectations

We assume the probability distribution for the expected earnings to have the asymmetric triangular distribution given by Figure A8. In our case, the most likely value is assumed to be the midpoint calculated from the minimum and maximum thresholds (if the triangular distribution was symmetric, the most likely value would be equal to the mean). The expected value of a left triangular distribution is given as $E(y) = p\frac{2y_{mid}+y_M}{3}$, and for a right triangular distribution is $E(y) = (1-p)\frac{2y_{mid}+y_m}{3}$.

Therefore, the expected mean can be written as:

$$E_d(y) = (1-p)\frac{2y_{mid}+y_m}{3} + (p)\frac{2y_{mid}+y_M}{3} \tag{5}$$

Figure A8: Triangular distribution



Source: author's construction.

A3 Questionnaires

Education survey

1. **School ID**:

   _____

2. **Your name**:

   _____

3. **Father's name**:

   _____

4. **Last name**:

   _____

5. **How old are you?**:

   _____

6. **Are you a boy or girl?**

   ☐ Boy

   ☐ Girl

7. **What is your caste?**:

   _____

8. **What is your home address?**:

   _____

9. **What is your phone number?Landline or parent's mobile phone**:

   _____

10. **What has your father studied?**

    ☐ 1-4 (Primary school)

    ☐ 5-10 (High school)

    ☐ 11-15 (Graduate)

    ☐ Post graduate

    ☐ Others (Specify)

    ☐ No education

☐ I do not have a father

11. What does your father work as?

☐ Office job

☐ Daily wage laborer

☐ Small business

☐ Big business

☐ At home

12. What has your mother studied?

☐ 1-4 (Primary school)

☐ 5-10 (High school)

☐ 11-15 (Graduate)

☐ Post graduate

☐ Others (Specify)

☐ No education

☐ I do not have a mother

13. What does your mother work as?

☐ Office job

☐ Daily wage laborer

☐ Small business

☐ Big business

☐ At home

14. How many younger siblings do you have?

☐ Brothers

☐ Sisters

15. How many of you siblings go to school? (excluding you)

☐ Brothers

☐ Sisters

16. What is your position among your brothers and sisters?

☐ Oldest

☐ Youngest

☐ Other

17. How do you travel to school?

☐ Walk

☐ Cycle

☐ Auto

☐ School Bus

☐ Public Bus

☐ Van

☐ Other _____

18. How much time does it take for you to come to school?

_____ minutes

19. Is this the nearest school to your house?

☐ Yes

☐ No

20. What is the highest class you would like to complete?
☐ Class 7 ☐ Class 8 ☐ Class 9 ☐ Class 10 ☐ Class 12 ☐ Technical course ☐ College ☐ I do not know

21. In the future what work would you like to do to support yourself and your family?:

_____

22. In the future, when you are 28 years old, do you think you will be working if you completed 12th standard (high school)

☐ Yes

☐ No [Skip to 24]

23. When you are 28 years old, what monthly income do you think you will earn if you complete 12th standard?:
Minimum: _____ Maximum: _____

24. In the future, when you are 28 years old, do you think you will be working if you completed diploma or technical education?

☐ Yes

☐ No [Skip to 26]

25. When you are 28 years old, what monthly income do you think you will earn if you complete diploma or technical education?:
Minimum: _____ Maximum: _____

26. In the future, when you are 28 years old, do you think you will be working if you completed college or graduation?

☐ Yes

☐ No [Skip to 28]

27. When you are 28 years old, what monthly income do you think you will earn if you complete college or graduation?:
Minimum: _____ Maximum: _____

28. What are the reasons for coming to school?

☐ To study

☐ Meet my friends

☐ To get a job

☐ Parents told me to go

29. Do you want to do the same work as your father?

☐ Yes

☐ No

30. Do you want to do the same work as your mother?

☐ Yes

☐ No

Exit survey

Note: The exit survey includes questions on the probability distribution of earnings expectations for different levels of education.

I. Control questions

On a scale from 0 to 10, 0 being not possible at all and 10 being completely possible:

31. Between 0 to 10 what is the likelihood that you will complete your homework in the next two days?: ☐

32. Between 0 to 10 what is the likelihood that you will complete your homework in the next two weeks?: ☐

II. Subjective earnings

33. Between 0 to 10 what is the likelihood that you will earn greater than _____ the midpoint if you completed high school education?

0	1	2	3	4	5	6	7	8	9	10

34. Between 0 to 10 what is the likelihood that you will earn greater than _____ the midpoint if you completed diploma or technical education?

0	1	2	3	4	5	6	7	8	9	10

35. Between 0 to 10 what is the likelihood that you will earn greater than _____ the midpoint if you completed college education?

0	1	2	3	4	5	6	7	8	9	10