Does Information Shape Borrowing Intentions?

Experimental Evidence on Track-Specific Higher Education Decisions in India

[Preliminary & Incomplete. Please Do Not Cite Without Permission.]

Tanvi Rao[1]

Abstract: This paper investigates the role of credit constraints in determining higher (post-secondary) education enrollment in India and the role of information in mediating the borrowing decision. I use unique data collected from high-school students on their subjective enrollment probabilities for three attendance tracks- technical/professional degrees, general academic degrees and vocational diplomas- and the non-attendance alternative, as well as track-specific expectations about future labor-market outcomes. An "information experiment" measures the effects of providing information on track-specific population returns, on students' own wage-beliefs, borrowing and enrollment decisions. We find that short-term borrowing constraints are a significant determinant of higher education enrollment in this setting. While a select set of individuals are keen to borrow for higher education, they are poorly informed about the labor-market returns to different higher education tracks and on average overestimate returns for all tracks relative to non-attendance; with the largest fraction of overestimation occurring with regards to the vocational education track. Relaxing the information constraint, causes overestimators at baseline to substantially revise earnings downward. We also find evidence to support that baseline overestimators, who are given access to information, borrow smaller amounts and this effect seems to be driven by the distribution of overestimators in the vocational education track. Finally, looking at post-treatment enrollment probabilities, for the group of individuals who borrow, we establish that overestimators exposed to the information treatment are significantly less likely to enroll in vocational tracks. That adjustment is seen along the margin of enrolling in vocational diploma courses is line with nationally representative data on education returns and costs; on average these courses offer earnings premia similar to general-degree tracks but are 4 times as expensive. These findings lend support to the idea that relaxing information constraints causes individuals to alter investment behaviors in a sensible manner, especially when investments are costly.

---

[1] Address: B05 Warren Hall, Cornell University, Ithaca-14853, New York, USA. Telephone: 302-588-0819. Email: tr254@cornell.edu.
I am most grateful to Ravi Kanbur, Jim Berry, Prabhu Pingali and David Just for advice and encouragement. I also thank Basit Zafar and Adeline Delavande for several helpful comments and suggestions on research design and questionnaire construction. The smooth and timely completion of data-collection on ground would not have been possible without institutional support from the TATA-Institute of Social Sciences and in particular Bhaskar Mittra and Maya Nair. An excellent field-team led by Thangkhanlal Thangsing helped manage and implement the survey. Lastly, I thank the entire leadership and administration of Ranchi University and the enthusiastic student-subjects who participated in the survey. Funding from the TATA-Cornell Agriculture and Nutrition Initiative (TCi) is gratefully acknowledged. All errors are mine.

1. Introduction

Across the developing world, not only is the overall attainment of post-secondary education low, but there is also a steep income-gradient in the attendance of post-secondary education. For instance, Kaufmann (2014) documents that in Mexico the poorest 40% represent only 8% of the student body in colleges. Using pooled data from two latest rounds of National Sample Survey (NSS) data[2] in India, I estimate that less than 7% of the sample individuals between 15-59 years attain some form of post-secondary education. Moreover, even though wage-premiums beyond secondary schooling are high in India; ranging from 42% to 113%, depending upon the type of post-secondary training, *more than half* of all individuals who complete secondary schooling do not enroll in further education.

Undoubtedly, individuals face multiple constraints in enrolling for higher education, which regular survey data make impossible to isolate. One standard explanation is the existence of credit or borrowing constraints among poor individuals. However, being poor does not imply being credit constrained- only poor individuals with high expected returns from education, who would like to borrow but cannot do so from existing sources, are credit constrained. Consequently, the measurement of such short-term borrowing constraints and the extent to which they hinder higher education enrollment is not straightforward. Here, I utilize a direct method of eliciting an individual's credit-constrained status for track-specific enrollment, following a method conceptually similar to Stinebrickner and Stinebrickner (2008), to measure the extent to which relaxing credit constraints alter individuals' higher education enrollment decisions. The method allows us to observe individuals' financially feasible choice-sets utilizing existing sources of borrowing and changes in feasible choice-sets on account of individuals' borrowing intentions based on the hypothetical availability of a higher education loan. We find that short-term borrowing constraints are an important but not universal constraint in our context, with a little over half of sample individuals stating that they would like to borrow for higher education enrollment. Interestingly, we find that earnings beliefs are a strong predictor of borrowing intentions, even after controlling for a host of relevant covariates.

[2] Refers to two latest rounds "employment-unemployment" data; i.e. - NSS 66th round (2009-10) and NSS 68th round (2011-12).

While some studies, in a developing country context, find that individuals on average underestimate the returns to different levels of education (Nguyen (2008), Jensen (2010)), beliefs regarding earnings are in fact highly heterogeneous (Wiswall & Zafar (2015)). In our sample of individuals we find that, on average, individuals overestimate earnings at baseline in the attendance tracks relative to the non-attendance alternative. Gamboa and Rodriguez (2014) whose sample of 15-18 year olds in Colombia, is more similar to the one in this paper, also find that students, on average, overestimate the pecuniary returns to higher education. To the extent that individuals overestimate returns on account of insufficient access to information regarding "true" population returns, relaxing a constraint on information can work to alter investment decisions of individuals between costly alternatives. The primary contribution of this paper relates to the effect of providing information regarding track-specific population earnings on (a) the borrowing intentions of students (made in consultation with their families) and (b) subjective enrollment probabilities of students who borrow for enrollment.

While a majority of students overestimate earnings in all three tracks, the largest fraction of overestimation occurs with regards to the vocational education track. Relaxing the information constraint, causes overestimators at baseline to substantially revise earnings downwards, but a symmetric response cannot be established for those who underestimate track-specific earnings at baseline. We also find some evidence to support that baseline overestimators in the information treatment group borrow smaller amounts than those in the control group and this effect seems to be driven by the distribution of overestimators in the vocational education track. Finally, looking at post-treatment enrollment probabilities, for the group of individuals who borrow, we establish that overestimators with access to information on returns, are significantly less likely to enroll in vocational tracks. These findings support the hypothesis that relaxing information constraints cause individuals to alter investment behaviors sensibly, especially when investments are costly.

The three "attendance tracks"- technical/professional degrees, general academic degrees and vocational diplomas/certificate courses- studied in this paper lie at distinct points of the net-return spectrum from post-secondary education in India. The three attendance track are also distinct in the type of educational content they impart and have distinct labor-market implications. Technical degree courses include professional degrees in fields like medicine, engineering and architecture as well as "job-oriented" degrees like Bachelors of Computer

Application, Business Administration, Information-Technology (IT) or Pharmacy. These courses are offered both by government and private institutions and are regulated by the "All-India Council for Technical Education (AICTE)". General degree courses are non-technical and award a bachelor's degree in either the arts, sciences or commerce, further categorized according to subject. Mostly, these are offered by the government via central or state level universities and colleges. Vocational courses are not academic and focus on imparting a set of skills (rather than broader academic knowledge) targeted towards employment in a specific sector. They are offered by both government and private institutes. Under the government, these courses are offered either by Industrial Training Institutes/Centers (ITI/ITC) or by Polytechnics. Recent reports[3] of the NSS 71st round on education expenses, estimates the average yearly costs for technical/professional degrees to be a little over 60,000 rupees (approx. 1,000 dollars) with the expenditure on private institutions being 1.5-2.5 times the cost of government institutes. Average yearly expenses for a general education, in contrast, were found to be around 7,000 rupees (approx. 100 dollars) and for vocational courses, around 30,000 rupees (approximately 450 dollars). While wage premiums for technical degrees are more than a 100% of the wages of those who complete high school, the fact that wage premiums for vocational courses (42% of high-school wage) are around 8 percentage points lower than the wage premiums for general courses, help explain why we see students adjusting enrollment on the vocational education margin.

World over, student loan programs are a popular policy for higher education financing as they balance concerns of cost-sharing with borrowers and of making higher-education more accessible. Recent policies of the Indian government, for higher education, have heavily focused on loans. For instance, the central government's one of two core higher education policies, the "*Pradhan Mantri Vidya Lakshmi Karyakram*", enlists national banks to provide loans for higher education. The Delhi state government also recently launched a higher-education loan scheme for up to 1 million rupees, with the state providing the loan guarantee. However, notwithstanding the recent thrust of the government, overall, the growth rate of education loans has been steadily declining. Data from the Reserve Bank of India (RBI) reveals that in 2014-15, the segment grew just 5.7% year-on-year compared to 9.2% in 2013-14, 10% in 2012-13 and 10.36% in 2011-12. The primary reason for the fall in growth rate of education loans has been attributed to a high

[3] http://pib.nic.in/newsite/PrintRelease.aspx?relid=122881 last accessed on 11.19.2015.

rate of loan defaults. An important aspect of the financial liability of a loan program is its "efficiency index" which broadly signals loan recovery as a fraction of loan repayments, a lending body expects to receive. Comparing across 26 higher education loan programs in different countries, Shen and Ziderman (2009) find India[4] to have a relatively low efficiency index (ranks 8th lowest) due to high default rates on loans. To the extent that individuals' borrow without accurate information on prospects in the job-market, providing information on returns can alter the pool of individuals' demanding loans. This can, potentially, work to alter the composition of loans away from low expected-return individuals. Hence, the bundling of information on expected returns with loan offers, can help not only individuals make better borrowing decisions but can also have implications for the overall supply of education loans.

The rest of the paper proceeds as follows. Section 2 briefly reviews some of the relevant literature, section 3 highlights a conceptual framework, section 4 discusses survey and data-collection details, section 5 presents results and section 6 concludes. All figures and tables are at the end, and an appendix contains additional tables to support the core findings of the paper.

2. Literature Review

The primary contribution of this paper is to an existing and prominent literature that uses data on individuals' subjective expectations about future events to study behavior. From a policy viewpoint, an attractive prospect of studying subjective expectations data is to see if an intervention that provides information can alter current decisions via changes in an individuals' information set. While the impact of information provision on measured returns from education has been used to study some aspects of decision-making in education, to my best knowledge this paper is the first attempt to study its effect on demand for higher education financing and individuals' decisions to borrow for enrollment in higher education.

Jensen (2010) finds that 8th grade students in the Dominican Republic substantially underestimate the returns to education and providing information to students in randomly selected schools, on measured returns in the community, to primary, secondary and university education, increased average schooling by 0.20-0.35 years in the four years following the

[4] Refers to India's State Bank of India (SBI) higher education loan program

intervention. Nguyen (2008) also finds that households in Madagascar update their own perceived returns from education in response to information on population returns. She finds that providing statistics to convey such information had positive and statistically significant effects on schooling investments measured by test scores and school attendance. Wiswall & Zafar (2015) show that their sample of students at New York University (NYU) have substantially heterogeneous major-specific earnings beliefs and a non-trivial number of students make both positive and negative errors. They find revisions in own-earnings, in response to information on public earnings, to impact stated college-major choice[5]. In a different vein, Dinkleman & Martinez (2014) find that providing randomly selected students with information about financial aid opportunities in Chile has favorable impacts on school reported absenteeism in treatment schools.

Osman (2014) looks at information and credit constraints in occupational choice in Egypt. By eliciting students stated probabilities of pursuing wage work, self-employment and inactivity after completing vocational high-school, the author finds evidence to support that bundling information and credit can have important compositional effects on the pool of borrowers. Relaxing credit and information constraints separately leads individuals to move towards self-employment but when credit is available information causes individuals to leave self-employment for wage work. This effect is driven by differing perceived returns to self-employment based on whether or not the individual is credit-constrained, and hence differential impacts of information on the two sets of individuals.

A secondary contribution of this paper is an attempt to throw additional light on a highly contested debate[6] on the overall importance of short-term liquidity constraints as a major

---

[5] Some other studies use different methods to gauge the effect of information constraints on stated education choices. Arcidiacono et. al (2012) and Delavande and Zafar (2014) correct for information gaps by adjusting student $i's$ expectation about his/her own earnings by the ratio of the sample median of all students' expectations of the average student's earnings by student $i's$ expectation of the average student's earnings.

[6] Compare Card (2001) and Carnerio & Heckman (2002) for an important line of the debate. The Card argument follows that the consistent empirical finding of IV estimates of wage returns to schooling as exceeding OLS returns is indicative of short run borrowing constraints. That is, ``switchers" induced into schooling, as a consequence of the instrument, face higher marginal costs of schooling as opposed to lower expected returns. This view is contested on grounds of invalid instruments and the fact that the IV>OLS result can hold even in the absence of credit constraints due to evidence on theories of comparative advantage in the labor market. In this case IV estimates may be higher than OLS estimates but lower than "true" rate of return in which case one may wrongly deduce credit constraints.

determinant of post-secondary schooling. Arguments against the importance of short-term liquidity constraints nudge policy towards interventions that relax long-term credit (or financial) constraints in favor of promoting build-up of "lifetime cognitive ability" and "college preparedness" among poor students.

The direct identification of credit constrained individuals, as in this paper, can help further mediate this debate. The approach in this paper follows that of Stinebrickner and Stinebrickner (2008) who study the extent to which the college drop-out decisions at Berea College in the U.S. can be attributed to short-term borrowing constraints. They classify students as credit constrained if they say that they would like to borrow money to increase consumption while in college and are not able to borrow from other sources[7]. Overall, they do not find short-term borrowing constraints to be an important determinant of the drop-out decision. Cameron and Taber (2004) use both IV and structural estimations to establish the absence of short-term constraints in college attendance in the U.S.

Studies that look at the importance of short-term borrowing constraints in developing countries, find them to be an important determinant of post-secondary educational attainment. This makes sense, because, by and large poorer countries do not have the extensive arrangements for higher education financing that are already in place in the U.S. Kaufman (2014) establishes the responsiveness of poor and high expected return individuals to reductions in direct costs of schooling (instrumented for by distance to college)[8] to establish the importance of credit constraints in Mexico. Delavande and Zafar (2014) simulate the effects of a loan policy for higher education in Urban Pakistan and find favorable results for enrollment in the context of the assumptions of their structural model.

Lastly, the paper's emphasis on enrollment along both the quantity and quality margin of higher education enrollment can throw light on the non-uniform effects of information, credit and their interaction for different higher education types. Not accounting for the possibility of differing education qualities can underestimate the importance of credit constraints as it excludes individuals who can afford some types of education but cannot afford all types. Similarly, not

[7] Students at the college already have full tuition subsidy and room and boarding subsidies. They find that a major portion of drop-outs from the college would remain even if credit constraints were removed completely.
[8] The novelty being that she has data on actual (subjective) expected returns of schooling and can compare IV estimates to "true" returns as opposed to comparing with OLS estimates of "true" returns.

accounting for quality can underestimate effects of information provision as individuals may be aware about earnings associated with certain types of higher education that they have a high exposure to, but for not of all education types.

3. Conceptual Framework

Below, we briefly describe how we expect credit and information constraints to operate in the current setting and how they could potentially interact.

An individual i, chooses an education-track d that maximizes the following utility function:

$$max_{d=t,g,v,na} = E[U_d = EPV_{it}(d) - C_i(d), \gamma_i(d) | \Omega_{it}]$$

The four education tracks are technical degrees (t), general degrees (g), vocational degrees (v) and not-attending (na) further education after 12th grade. Individuals seek to maximize the difference between perceived benefit, measured by $EPV_{it}(d)$, and perceived cost $C_i(d)$ from higher education. $\gamma_i(d)$ refers to other observed and unobserved individual-specific inputs that affect U_d. Here, both $EPV_{it}(d)$ and $C_i(d)$ are idiosyncratic. $EPV_{it}(d)$ Is the Expected Present Value[9] of lifetime earnings associated with d and is conditional on an individual's information set at time t- Ω_{it}. Apart from differences in information sets, idiosyncrasies in $EPV_{it}(d)$ arise because of individual-level abilities, preferences and social-networks for occupations related to d and $C_i(d)$ is idiosyncratic because tuition, transportation, room and board costs differ by individuals. Also, for each d, individuals state perceived benefits and costs according to the specific course and institute that they would consider enrolling in within each category d.

At $t = 0$, the individual states baseline enrollment probabilities utilizing baseline beliefs about returns (benefits) associated with education alternatives in their feasible choice-set. An individual's feasible choice-set is defined by subtracting (perceived) yearly costs the individual associates with each track d from the maximum amount an individual states that they and their

[9] We will only have data on an individual's beliefs regarding his/her earnings at age 30, for each education type, and there EPV (d) involves assumptions on how an individual's earnings evolve over his/her life-cycle.

families can pay in a year towards higher education, taking into consideration available sources of borrowing.

Between $t = 0$ and $t = 1$ half of randomly selected individuals (treatment group) participate in an information session that discusses the measured returns in the population to the four distinct education tracks. Also during this period, all individuals are presented with a hypothetical option of taking out a loan, at a fair interest rate, towards their higher education. At $t = 1$, all individuals re-state beliefs about returns (benefits) associated with each track d. For individuals who state that they would like to take out a loan for higher education, the amount that they would like to borrow is added to the amount that they can pay at baseline to define a revised feasible choice-set. Accordingly, individuals re-state enrollment probabilities for choices in their revised feasible choice sets.

At $t = 2$ the individual decides which higher education degree to enroll in. At $t = 0$ and $t = 1$, there is some time remaining for making the final education decision, and hence there is some uncertainty (referred to as "resolvable uncertainty") due to which the individual solves their optimization problem by assigning probabilities to each d. At $t = 2$, the individual makes a discrete choice between the education options, and chooses one that maximizes his utility function. Empirical investigation that follow pertain to $t = 0$ and $t = 1$.

Probabilities assigned to each education option π_d should be positively correlated with $EPV_{it}(d) - C_i(d)$. However, if an education option is entirely (i.e. all courses within d) unaffordable to an individual, then the baseline probability assigned to that education option is zero and $\pi_d = 0$. At baseline, an individual is credit-constrained with regards to d, if $EPV_{it}(d) - C_i(d) > 0$, but they cannot assign a positive, non-zero probability to enrolling in d, because they cannot afford d. Credit constraints can also operate at the intensive margin, if the individual assigns a lower probability to an education alternative than they would if they could borrow. As an example for the intensive margin, consider two courses within d-d_1 and d_2. $C_i(d_1) < C_i(d_2)$ And $EPV_{it}(d_1) - C_i(d_1) < EPV_{it}(d_2) - C_i(d_2)$. $C_i(d_1)$ is affordable to i but $C_i(d_2)$ is not. Being able to borrow, allows i to pursue $C_i(d_2)$ which alters the overall probability of pursuing d. In both cases, the extensive and the intensive margins, credit-constrained individuals would like to borrow but cannot do so. On the other hand, an individual is not credit-constrained with regards to d if either (a) $EPV_{it}(d) - C_i(d) < 0$ or (b) $EPV_{it}(d) - C_i(d) > 0$ but the individual would not like to borrow.

For credit-constrained individuals changes in enrollment probabilities between $t = 0$ and $t = 1$ can be due to alterations purely on account of credit constraints: the possibility of borrowing can allow an individual to assign a positive, non-zero probability to a d which was earlier zero due to unaffordability i.e. a change on the extensive margin, or it can alter probabilities among options with positive, non-zero probabilities at baseline, i.e. a change on the intensive margin. For all individuals, changes in enrollment probabilities between $t = 0$ and $t = 1$ can also be on account of changes in an individual's information set, due to the arrival of potentially new and relevant information, regarding average earnings in the population, and individual updating of their beliefs regarding their own-earnings associated with d (and hence $EPV_{it}(d)$) in response to that information.

Importantly, information and credit constraints can interact. Consider two individuals- one with access to information on measured returns and one without, both of whom are not credit constrained at baseline, with regards to d, on account of situation (a) i.e. when $EPV_{it}(d) - C_i(d) < 0$. For individuals in the treatment group, access to information can raise $EPV_{it}(d)$ and induce them to be in the "credit-constrained" category and hence borrow for enrolling in d. Similarly, for individuals operating at the intensive margin, information can induce borrowing which can cause them to alter enrollment probabilities in $t = 1$, among non-zero probabilities at baseline. This happens when learning about the distribution of earnings associated with d and its magnitude relative to alternatives, encourages an individual to invest in a more expensive alternative within d. In both of these cases, we would expect to see a *larger* increase in the probability for enrolling in d in the treatment group vis-à-vis the control group.

On the flipside, for credit-constrained individuals in the treatment group, who overestimate earnings, information can lower $EPV_{it}(d)$ and induce them to "exit" the credit-constrained category. At the intensive margin, learning that earnings associated with d relative to alternatives, were overestimated at baseline, can induce individuals in the treatment group, to revise $EPV_{it}(d)$ downward and discourage investment in d, which decreases the overall probability of enrolling in d. In both of these cases, we would expect to see *smaller* increase in the probability for enrolling in d in the treatment group vis-à-vis the control group.

4. Data Collection & Experiment Details

4.1 Data collection & Timing

The data for this study was collected from a sample of 1525 students across nine public colleges in the East Indian state of Jharkhand. All nine colleges are constituent colleges of a large state University and the students, at the time of the survey, were studying in the final year of their intermediate degree[10], hereafter referred to as 12th grade. Four of the nine colleges are situated in the capital city of Ranchi, one in a rural block of Ranchi district and four others are in surrounding rural districts. The survey was conducted between October 2014 and February 2015, five-nine months prior to the time when students make actual decisions regarding enrollment in post-secondary education.

Figure 1 highlights the timeline and the structure of the survey. Half of the complete sample was randomly assigned to the information treatment group and the other half to the control group. We drew, approximately, an equal number of students from each college. Further, within each college, students were randomly assigned to survey-sessions of 15 students each. Survey sessions were either a control session or a treatment session, with the latter differing only on account of the feature that it included a 15-20 minute long information session at the end of the collection of baseline data and disbursal of "loan cards" that posed two "borrowing questions" to the students, to be answered the following day. For a given survey-session, round 2 of data collection was conducted the day after the first round. In every college, both rounds of all control sessions were conducted before the treatment survey-sessions, in order to prevent students from the treatment group to share information with students in the control group, in a manner that can influence the results of this paper. Both sets of students answered exactly the same round 1 and round 2 questions.

[10] After completing 10th grade, students decide to attend either a "junior" or an "intermediate" college, for two years of higher-secondary schooling, or to attend a high school which offers 11th and 12th grades. Public "Junior" or "Intermediate" colleges, like the ones surveyed here, are often co-located with public colleges offering undergraduate degrees.

Survey sessions were conducted in classrooms within the students' college and were led by a team of two enumerators. Students answered the questions, posed by the enumerators, on android tablets. The questionnaires were fielded using Open Data Kit (ODK) software.

4.2 Survey Questionnaire & Information Treatment

Round 1 of the survey consisted of questions on (i) socio-economic details including gender, caste, religion, a small "household assets" module, parental education and occupation, older sibling gender & education, scores on previous board examinations and history of grade repetition and (ii) baseline beliefs contingent on each higher education alternative i.e. technical/professional degrees, general degrees, vocational diplomas/certificate courses and the fourth alternative of not attending further education after 12th grade. Since the four education categories for which beliefs were elicited are broad, data collection was preceded by a detailed explanation of possible courses/degrees that are part of every category. Since a majority of the beliefs questions were either probabilistic in nature or required students to express responses on a scale of 0-100, the baseline beliefs module was preceded by a discussion (with examples) on answering probabilistic questions[11]. All scripts and the questionnaires used for this study are available upon request.

In the baseline beliefs module, individuals were asked about certain non-pecuniary[12] and pecuniary beliefs for each education alternative. Pecuniary beliefs included data on expected probability of employment and expected average monthly earnings contingent on completing each higher education alternative. These pecuniary beliefs were collected both for individuals' perceptions regarding their own expected labor market outcomes and outcomes they believe apply to an average individual in the population. Next, stated probabilities of enrollment were elicited for (i) all four higher-education alternatives and (ii) only for higher education alternatives that are affordable to the individual. The affordable choice set for every individual was calculated by subtracting (i) what they think would be the out-of-pocket yearly costs (including fees, boarding, books and other expenses) if they chose to enroll in a degree/course

[11] We ensured that answers to all probabilistic questions sum to 100 by placing the total as a constraint in the questionnaire, without fulfilling which, the survey would not proceed to the subsequent question.

[12] Non-pecuniary beliefs included questions regarding enjoyment of coursework, parental approval, likelihood of graduation

belonging to each of the three higher-education alternatives in a government or private institute of their choosing, from (ii) what they think is the maximum amount that they or/and their parents can pay towards their higher education, without borrowing, if they chose to study after 12^{th} grade. In asking students their out-of-pocket yearly costs, we emphasized that students report costs that they or/and their parents have to pay themselves, over and above any scholarships or borrowing that they are likely to have access to, at the time of making the decision.

At the end of round 1, all students were given a "loan-card" which had two questions related to borrowing for higher education which they had to think about at home and discuss with their family members. The two questions were- a) whether the individual would like to accept a loan, offered at a fair interest rate, for attending higher education, to be repaid only after completion of their studies- "yes" or "no"[13], b) If "yes", keeping in mind the length of their desired degree, how much would they like to borrow on a yearly basis?

Additionally, students part of the treatment survey-sessions also participated in a 15-20 minute information session that discussed the average and the 25th and 75th percentile of the monthly earnings distribution of men and women who have completed each higher education alternative, calculated from two latest rounds of National Sample Survey (NSS) data. Individuals part of the information treatment group also took home a sheet of paper with a graph and some statistics that summarized the contents of the information session that they were part of. The script of the survey-session and the "information sheet" taken home by the students are available upon request.

The next day, for round 2, students were (i) re-asked about their stated enrollment probabilities for all four higher education alternatives, (ii) expected average monthly earnings for each higher education alternative. In addition, their response to the questions posed in the "loan card" were recorded. For individuals who answered in the affirmative for wanting to borrow, the amount that they would like to borrow on a yearly basis was used to re-calculate their "affordable" choice set and stated probabilities of enrollment for this choice set were further re-elicited.

[13] The question was worded as: *"Suppose that someone (bank or non-bank) offers you a loan to enroll in a higher education course of your choice. This loan is available at a fair interest rate and you have to repay the loan only after you complete your higher education. Do you think you would accept such a loan"?*

5. Results

Table 1 summarizes the key background variables of sample individuals and checks for balance in these characteristics across control and treatment groups. Control and treatment individuals do not differ statistically on account of almost all relevant socio-economic and demographic characteristics, at baseline. However, we see that control individuals are more likely to own land (p-value=0.04) and individuals in the treatment group have a slightly higher index of household assets[14] (p-value=0.10). Nevertheless, two other variables that are also indicative of the individual's household's well-being, namely the "HH Facility Index"[15] and the number of education tracks (out of 3) affordable to the individual at baseline, do not statistically differ between control and treatment groups. More importantly, baseline differences in land ownership and household assets, do not manifest in statistically different baseline enrollment probabilities or substantial differences in baseline earnings beliefs.

5.1 Impact of the information treatment on own-earnings beliefs

In this section we present evidence to establish that the information treatment conveyed new and relevant information on earnings to our sample of individuals and caused a substantial revision in own-earnings beliefs. This is indicative of the fact that the treatment was successful in addressing substantial information-gaps at baseline.

We see that while perceptions about own-earnings are highly heterogeneous, a substantial majority of individuals in the sample overestimate earnings at baseline. Figure 1 plots the track-wise distribution of beliefs regarding average monthly own-earnings of individuals, separately for males and females, overlaid on reference lines that represent each gender's "true" average earnings. Idiosyncrasies in individuals' beliefs regarding their own-earnings arise not only on account of differences in information sets, but also on account of individual's perceptions about

[14] The household asset index is a simple sum of 19 dichotomous items measuring household possessions.
[15] The household facility index is a simple sum 6 household characteristics including household level access to piped indoor water, a separate kitchen, a flush toilet, electricity, Liquefied Petroleum Gas (LPG) for cooking, and a "pucca" house.

their own abilities, and due to differences in preferences and occupational social-networks. To establish that overestimation of own-earnings in the sample is largely on account of inaccuracies in information sets, Figure 2 plots the track-wise distribution of beliefs regarding the average monthly-earnings of an average individual in the population. Both Figures 1 & 2, indicate that a majority of individuals overestimate earnings in all three attendance tracks, with overestimations for technical and vocational education being most substantial. Appendix Table A1 summarizes errors in beliefs in the three attendance tracks *relative* to the non-attendance alternative. Here, an individuals' informational error is measured as:

$$rel\_error_{id} = (actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA}) \tag{1}$$

With $rel\_error_{id} < 0$ indicating an overestimation of track-specific earnings relative to non-attendance and $rel\_error_{id} > 0$ indicating an underestimation of track-specific earnings relative to non-attendance. In (1) $actual_d$ indicates the "true" earnings-average for track d in the population and $perceived_{id}$ indicates individual-specific beliefs regarding the earnings-average for an average individual in the population. The subscript NA denotes the non-attendance alternative. Two points are of importance in Table A1. One, is that the largest fraction of individuals overestimate earnings in the vocational education track, with the fraction of overestimators in this track exceeding the fraction of overestimators in technical and general tracks by 9 to 15 percentage points, respectively. Secondly, the average extent of overestimation across tracks is about *twice* as much as the extent of underestimation.

Next we examine the effect of the information treatment on individuals' beliefs regarding their own earnings. Similarities in the distributions of own-earnings beliefs and population-earnings beliefs in our sample, suggest that information regarding population earnings are an important input in individuals' beliefs about their own earnings. This is further confirmed by examining the distribution of post-treatment own-earnings beliefs of individuals' part of the information treatment with that of control individuals in Figure 4. It is clear that the own-earnings distribution of treatment individuals is shifted downward (to the left) and significantly differs from that of control individuals (combined K-S p-value = 0). The two distributions are identical at baseline.

Tables 2 and 3 further test for systematic differences in own-earnings revisions, between treatment and control groups. Panel A of Table 2 establishes that a significantly larger proportion of individuals in the treatment-group revise relative earnings downward in the general and vocational education tracks. Here, earnings revisions for a track d is defined relative to earnings revisions for the non-attendance alternative as:

$$revise_{id} = (wage'_{id} - wage'_{iNA}) - (wage_{id} - wage_{iNA}) \quad (2)$$

With $wage'$ indicating round 2 (post-treatment) earnings beliefs and $revise_{id} < 0$ indicating a downward revision in relative earnings for track d. Panel B relates the direction of own-earnings revisions to the direction of baseline errors measured as in equation (1). Comparing across overestimators at baseline, individuals in the treatment group, are significantly more likely to revise earnings downward in comparison to their peers in the control group. This magnitude of this effect is largest for the vocational track.

Panel A of Table 3, regresses relative earnings revisions measured as in equation (2), and transformed using an inverse-hyperbolic sine (IHS) transformation (Burbidge et. al, 1988), on a treatment group dummy. Downward earnings revision is largest in magnitude for the vocational education track. In Panel B, the interaction of $revise_{id}$ and rel_error_{id}, with the latter also being IHS transformed, yields an elasticity interpretation for the rel_error_{id} coefficients. The interaction of the rel_error_{id} term with the treatment group dummy, indicates that a 1% error in population earnings results in 0.09-0.13 percentage point larger revision in the treatment group, depending on the track. Among alternative theories of updating described in the literature (Wiswall & Zafar, 2015a) the positive coefficient on the treatment-error interaction term in Panel B provides evidence to support skill-price updating in the sample. Individuals believe that earnings are a product of their level of skill and the price per unit of skill. While they are relatively certain about their level of skill, they are unsure about the price of skill in the economy. Upon learning that average population earnings are lower than believed, individuals revise downward their estimate of the price of skill in the economy and also their beliefs regarding their own future earnings.

Finally, in Appendix Table A2 we establish that beliefs about own-earnings associated with the completion of a higher education track are indeed a significant predictor of an individual's subjective probability of enrollment in the track. Specifically, we regress the log subjective probability of enrolling in track d (relative to the non-attendance probability) on the log expected average own-earnings in track d (relative to the non-attendance earnings belief). The first two columns examine this relationship cross-sectionally, with column (2) adding in a host of baseline controls. Expected earnings are an important but relatively inelastic predictor of enrollment, with a choice-elasticity of approximately 0.5. However, cross-sectional estimates are still expected to be biased on account of unobserved tastes that are likely correlated with earnings and enrollment probabilities. To account for this, and following Wiswall & Zafar, (2015b), in columns (3)-(4) we take the sample of individuals in the treatment group of the sample, for whom we see substantial revisions in earnings beliefs, and regress individual level *changes* in enrollment probabilities on earnings beliefs. The choice-elasticity is 0.33 percentage points smaller in magnitude, but continues to predict enrollment probabilities. This suggests, that revisions in own-earnings beliefs, can potentially affect choices made by individuals among costly alternatives in their higher-education choice-sets. We explore this issue subsequently, by studying borrowing decisions in section 5.2 and changes in stated enrollment probabilities on account of borrowing in section 5.3.

5.2 Does Information Affect Borrowing Decisions?

In this section, we first explore the "baseline" effect of relaxing credit-constraints on higher education enrollment decisions in the sample. Simply being poor does not imply being credit constrained; only poor individuals with high expected returns, who would like to borrow to finance the costs of enrolling in higher education, and cannot borrow from available sources are credit-constrained. This theoretically appealing definition of credit-constraints stresses the importance of borrowing intentions. Borrowing intentions are shaped by expectations regarding the future stream of returns from investments under consideration. To the extent that expectations regarding future earnings are made on account of inaccuracies in an individual's information set, the information treatment should influence borrowing intentions. Therefore, we then look at the effect of the information treatment on borrowing intentions that characterize the "credit-constrained" status of sample individuals.

We modify the approach used in Stinebrickner and Stinebrickner (2008) to directly measure credit-constrained individuals. The primary aspects of the data used to measure classify individuals as credit-constrained, on the enrollment margin, is if they state that they would like to accept the "offered" loan for attending higher education. Further, we measure the amounts that individuals state that they would like to borrow[16], changes in individuals' financially feasible choice-sets and revised enrollment probabilities. Because enrollment probabilities for education alternatives feasible at baseline take into account the possibility of borrowing from existing sources, revisions in enrollment probabilities on account of borrowing, satisfy the theoretical requirements of measuring credit-constrained individuals described earlier.

Table 4 looks at the effect of relaxing credit-constraints, or making the hypothetical loan offer, with regards to the control group of the experiment. We consider this to be the "baseline" importance of credit-constraints in the sample; that is for the representative student not exposed to the information treatment. While short-term borrowing constraints are important, they are not universal. In Panel A, out of a total of 718 individuals in the control group, who completed both survey rounds, 403 individuals (56.13%) of the sample answered *"yes"* to the question of whether or not they would like to accept the offered loan towards their higher education enrollment. This represents an upper-bound on the proportion of credit-constrained individuals in the control-arm of the sample. The distribution of amounts that individuals state that they would like to borrow are highly dispersed. Among credit-constrained individuals, the mean borrowing amount is 1, 11,461.40 rupees per year (approx. 1,671 dollars) and the median borrowing amount is 50,000 rupees per year (approx. 750 dollars).

Panel B of the same table explores changes in enrollment probabilities among individuals who borrow. As described in section 3, we conceptualize an enrollment increase on the extensive margin as an increase in the probability of enrollment in a track unaffordable to the individual at baseline, which is now affordable on account of borrowing. 65% of individuals who borrow, increase enrollment along this margin. An enrollment increase on the intensive margin refers to increase in the probability of enrollment in a track which was affordable to the individual even at baseline. Here, the possibility of borrowing can potentially increase the overall probability of pursuing the track by enabling an individual to attend a more expensive degree/course within the

[16] Question posed only to individuals who say that they would like to borrow

track, increasing the overall probability of enrolling in the given track. Approximately 23% of individuals increase enrollment along the intensive margin. This suggests that a majority of students (88% of borrowers) did interpret the borrowing question correctly, and applied the loan amounts for higher education enrollment and not simply consumption.

In Appendix Table A3, we make a brief digression to establish that, indeed, individuals who overestimate at baseline, are significantly more likely to want to borrow for higher education. A simple correlation reveals that, overestimators are approximately 12.8 percentage points more likely to borrow, than those who do not overestimate. This effect is robust to the inclusion of several controls. This indicates that our concerns regarding borrowing intentions being shaped, in part, by information-gaps are valid.

Table 5 looks at borrowing differences across treatment and control groups, interacted with an indicator for baseline overestimators. Here, we see again, that not only are overestimators much more likely to borrow, but they also intend to borrow much larger amounts. The overall effects of treatment among those who overestimate at baseline, are inconclusive, though it does appear that overestimators in the treatment are less likely to borrow and borrow smaller amounts. However, these overall effects (columns 1-2) are not statistically significant. However, when we break down the sample by track, which in effect amounts to weighing differences in logged borrowing amounts between treatment and control groups, by the distribution of overestimators in each track, we find that borrowing amounts in the vocational education track are significantly smaller among treatment group overestimators. This is consistent with our earlier findings that the vocational education track had the largest proportion of downward revisions and saw downward revisions of the largest magnitude.

5.3 Does Information Impact Enrollment on Account of Borrowing?

Finally, in this section, we compare the enrollment decisions of students who borrow (also termed as the credit-constrained students in our sample), across treatment and control groups and by status of earnings-overestimation at baseline. In previous sections, we established that compared to other tracks, the largest proportion of students overestimate earnings in vocational tracks, and the strongest effect of the information treatment was on downward wage revisions in the vocational education track. Differences in borrowing amounts between overestimators in control and treatment groups are also driven by the distribution of overestimators in the vocational track. In Table 6, the coefficient on the treatment-overestimate dummy, indicates differences in post-treatment enrollment probabilities of overestimators in treatment versus control tracks. Here, we clearly see that treatment group overestimators have a significantly lower subjective likelihood of enrolling in vocational tracks (p-value=0.059). This indicates that the information treatment not only led individuals to revise beliefs about their future own-earnings but also impacted their future choice of higher education enrollment.

6. Conclusion

We study the effect of providing information on track-specific population returns, in an experimental setting, on students' track-contingent earnings beliefs, borrowing and enrollment decisions, using unique data collected from a sample of 12^{th} grade students in a large public state university in the Indian state of Jharkhand. As our outcome of interest, we study students' subjective enrollment probabilities for three attendance tracks- technical/professional degrees, general academic degrees and vocational diplomas- and the non-attendance alternative. Revisions in earnings expectations help explain difference in borrowing behavior between treatment and control groups, as well as revisions in enrollment probabilities.

While a majority of students overestimate earnings in all three tracks, the largest fraction of overestimation occurs with regards to the vocational education track. Relaxing the information constraint, causes overestimators at baseline to substantially revise earnings downwards. We also find evidence to support that baseline overestimators in the information treatment group borrow smaller amounts than those in the control group and this effect seems to be driven by the distribution of overestimators in the vocational education track. Finally, looking at post-treatment enrollment probabilities, for the group of individuals who borrow, we establish that overestimators with access to information on returns, are significantly less likely to enroll in vocational tracks. These findings support the hypothesis that relaxing information constraints cause individuals to alter investment behaviors sensibly, especially when investments are costly. Moreover, adjustment along the margin of enrolling in vocational diploma courses is line with nationally representative (NSS) data on education returns and costs; on average these courses offer earnings premia similar to general-degree tracks but are 4 times as expensive.

These findings have implications, not only in highlighting the important role of information in helping resource-constrained individuals make better decisions between costly alternatives but also suggest possible implications for increasing the overall supply for higher education loans in the country, by altering the composition of borrowers away from low-expected return individuals, who mistakenly anticipate substantial returns on account of inaccurate access to information.

References

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, *166*(1), 3-16.

Attanasio, O., & Kaufmann, K. (2009). *Educational choices, subjective expectations, and credit constraints* (No. w15087). National Bureau of Economic Research.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, *83*(401), 123-127.

Cameron, S. V., & Taber, C. (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of political Economy*, *112*(1), 132-182.

Delavande, A., & Zafar, B. (2014). University choice: the role of expected earnings, non-pecuniary outcomes, and financial constraints. *FRB of New York Staff Report*, (683).

Dinkelman, T., & Martínez A, C. (2014). Investing in schooling in Chile: The role of information about financial aid for higher education. *Review of Economics and Statistics*, *96*(2), 244-257.

Gamboa, L. F., & Rodríguez, P. A. (2014). *Do Colombian students underestimate higher education returns?* (No. 012050).

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, *125*(2), 515-548.

Kaufmann, K. M. (2014). Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns. *Quantitative Economics*, *5*(3), 583-630.

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from Madagascar. *Unpublished manuscript*, *6*.

Osman, A. (2014). Occupational choice under credit and information constraints. *Available at SSRN 2449251*.

Shen, H., & Ziderman, A. (2009). Student loans repayment and recovery: international comparisons. *Higher education*, *57*(3), 315-333.

Stinebrickner, R., & Stinebrickner, T. (2008). The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study. *American Economic Review*, *98*(5), 2163-84.

Wiswall, M., & Zafar, B. (2015). How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital*, *9*(2), 117-169.

Wiswall, M., & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, *82*(2), 791-824.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, *48*(3), 545-595.

Figures & Tables

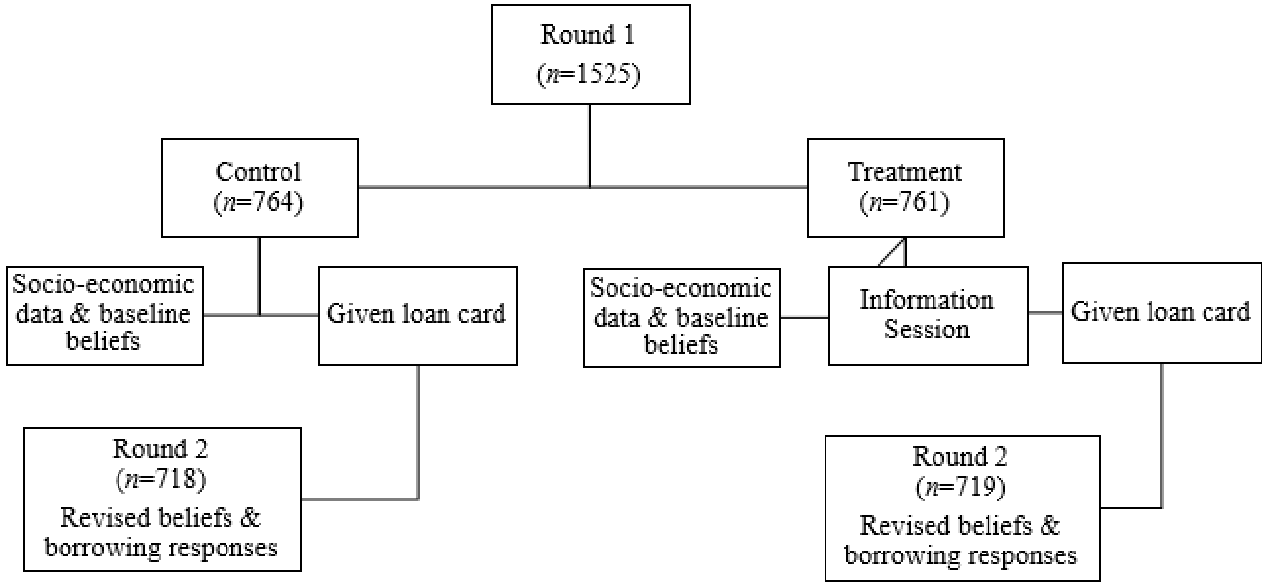


Figure 1: Survey Structure & Experimental Design

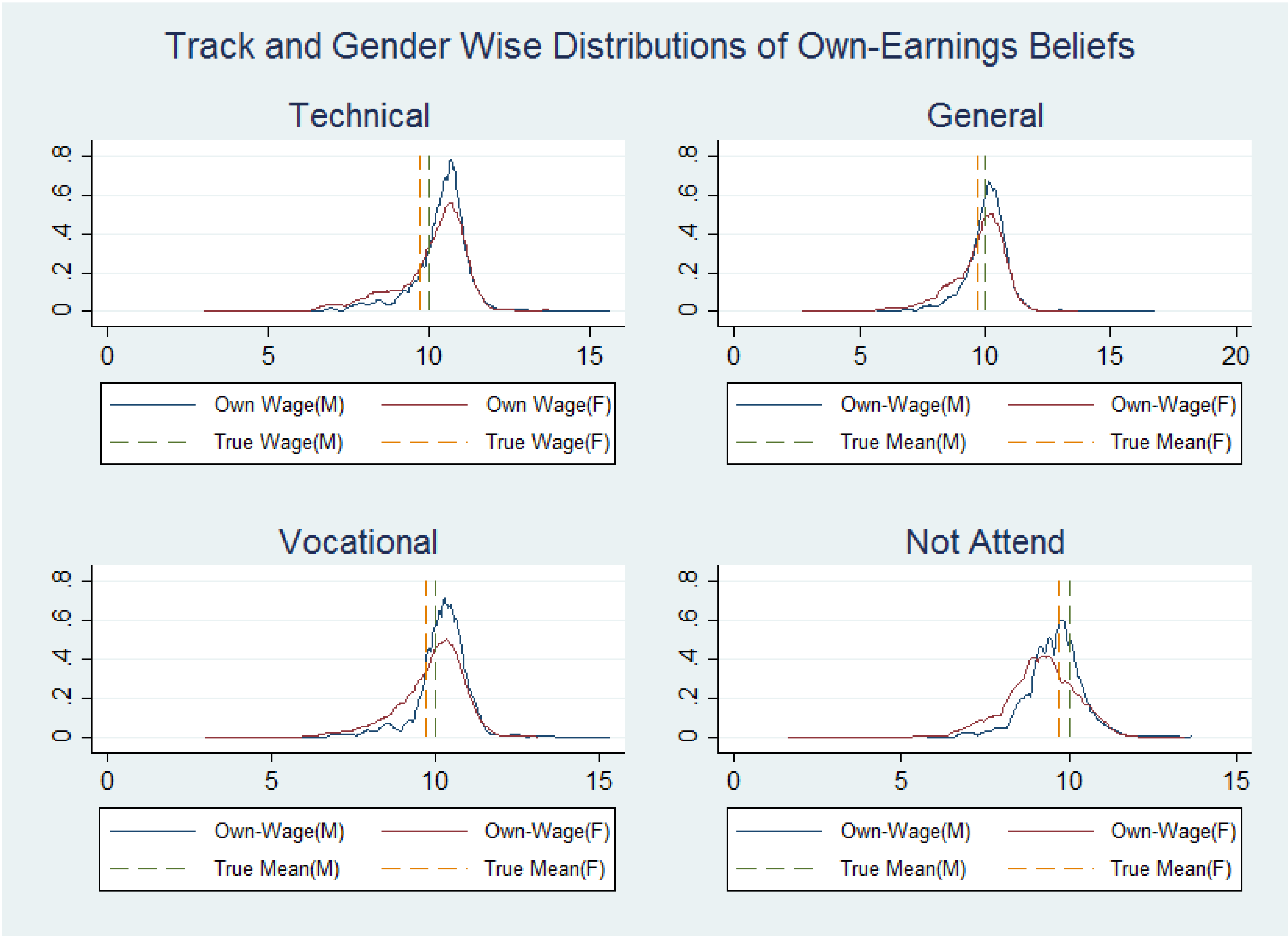


Figure 2: Track and Gender-Wise Distributions of Own Earnings Beliefs at Baseline

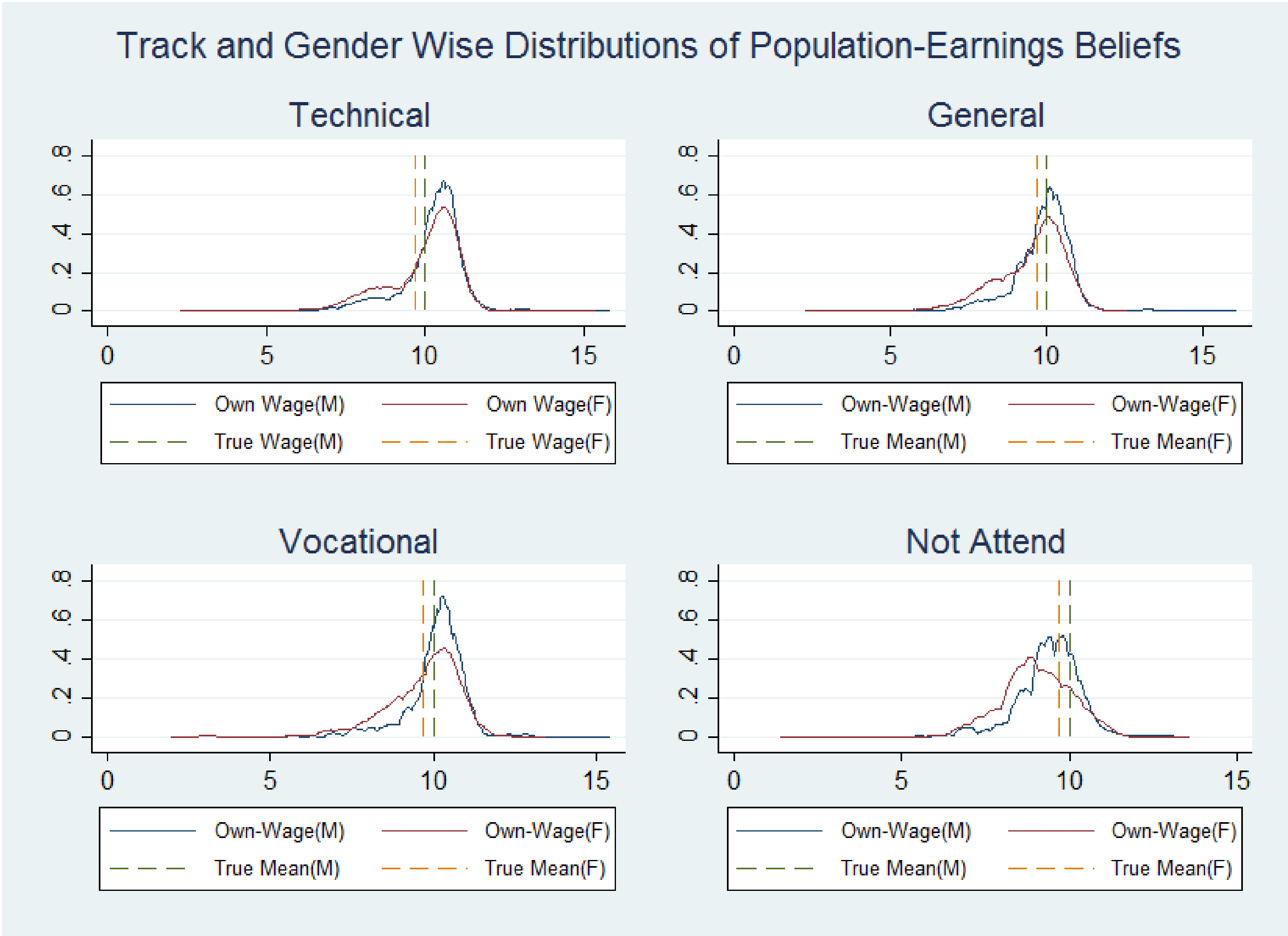
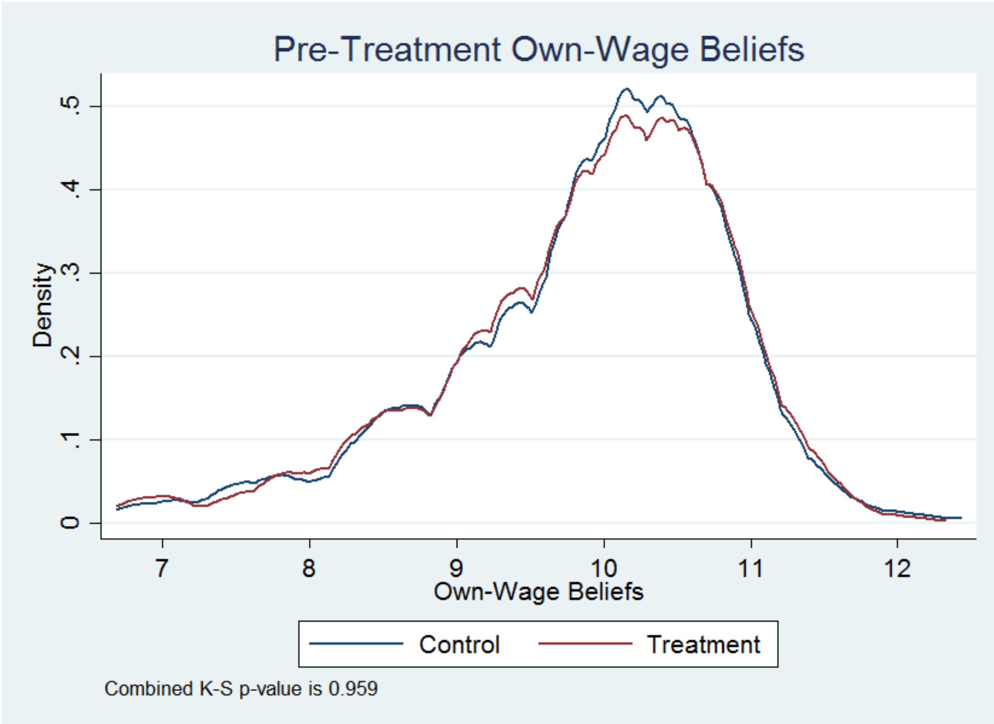


Figure 3: Track and Gender-Wise Distributions of Population Earnings Beliefs at Baseline



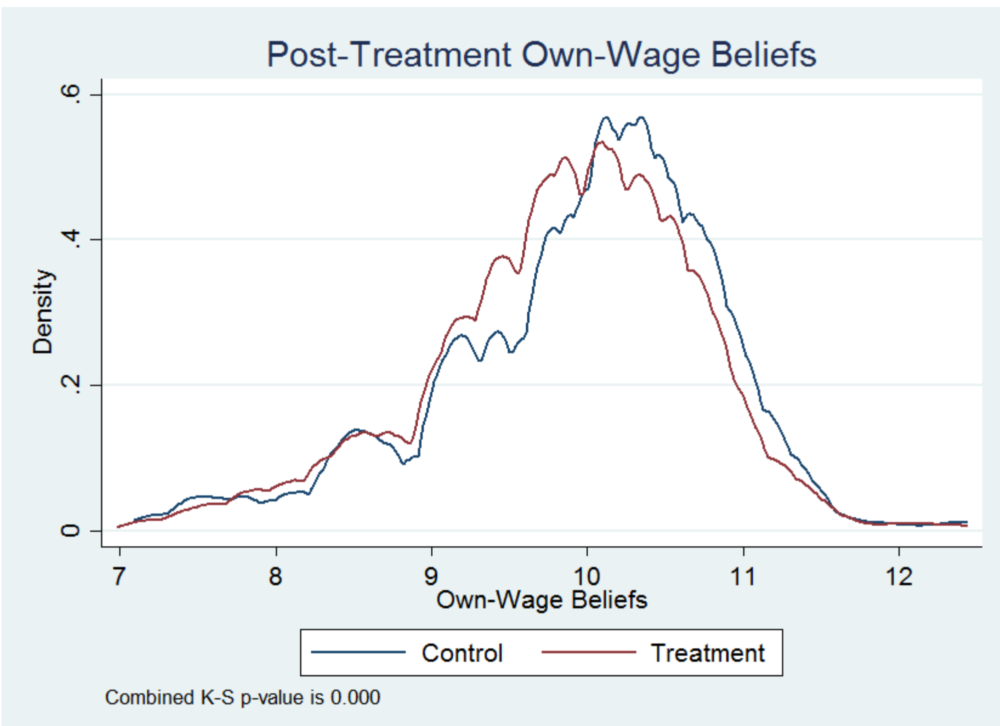


Figure 4: Pre & Post Treatment Distribution of Own-Earnings Beliefs (All Tracks Pooled)

Table 1: Balance of Baseline Variables

	Control	Treatment	p-value
	(1)	(2)	(3)
Age	17.24	17.30	0.19
% Male	0.53	0.55	0.45
% Scheduled Tribe	0.33	0.34	0.70
% Hindu	0.65	0.63	0.39
Asset Index	7.52	7.82	0.10
HH Facility Index	2.62	2.63	0.87
% Own Land	0.74	0.69	0.04
Board Exam Score	61.16	61.17	0.98
% Father High School	0.18	0.20	0.22
% Mother High School	0.08	0.09	0.91
% Father Family Business	0.11	0.14	0.11
% Father Salaried Job	0.21	0.21	0.88
% Mother Housewife	0.60	0.61	0.51
Average Older Sibling Edu.	5.59	5.50	0.48
Edu. Tracks Affordable	1.04	1.08	0.49
Enroll Probability (Tech)	10.92	11.80	0.47
Enroll Probability (Gen)	21.36	22.28	0.60
Enroll Probability (Voc)	12.78	11.97	0.51
Enroll Probability (NA)	54.94	53.96	0.68
% Overestimate Earnings	0.62	0.59	0.22
% Arts Stream	0.34	0.34	0.84
% Commerce Stream	0.31	0.31	0.99
% Science Stream	0.35	0.35	0.86

Notes: Columns (1) and (2) show sample means of individuals in the Control and Treatment group, respectively. Column (3) shows p-values of OLS regressions of each variable on a treatment group dummy.

Table 2: Track-Wise Proportion of Downward Revisions Across Control & Treatment

	Revise Relative Earnings Downward?		
	Technical	General	Vocational
Panel-A			
Treatment	0.0397	0.0662**	0.0717***
Constant	0.440***	0.380***	0.447***
Panel-B			
Treatment	-0.00636	0.018	-0.0104
Overestimate	0.038	0.0553	0.0531
Treat#Overestimate	0.0818	0.0932*	0.122**
Constant	0.416***	0.349***	0.410***
Observations	1,437	1,437	1,437

Notes: (1) In all columns, the outcome variable is a dummy for individuals who revise beliefs about track *d* earnings, relative to non-attendance earnings, downward.
That is, *outcome=1* if $(wage'_{id} - wage'_{iNA}) - (wage_{id} - wage_{iNA}) < 0$ where $wage'$ refers to round 2 wage belief.

(2) 'Overestimate' is a dummy for individuals who overestimate, at baseline, beliefs about population earnings in track *d*, relative to non-attendance earnings.
Thus, *Overestimate=1* if $(actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA}) < 0$

(3) Standard errors are clustered at the survey-group level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 3: Track-Wise Earnings Revisions Across Control & Treatment

	Earnings Revision		
	Technical	General	Vocational
Panel-A			
Treatment	-0.279	-0.649	-1.112**
Constant	-0.381	0.597**	-0.489
Panel-B			
Treatment	-0.141	-0.574	-0.654
Error	0.0643	0.0538	0.0904**
Treat#Error	0.0944*	0.115**	0.127**
Constant	-0.199	0.680**	-0.108
Observations	1,437	1,437	1,437

Notes: (1) In all columns, the outcome variable is a continous variable of revision in beliefs about track *d* earnings, relative to non-attendance earnings. Earnings revisions have been transformed using an inverse hyperbolic sine transformation, which works similar to a log-transform, but is capable of handling zero and negative values (Burbidge et. al, 1988)
That is, $IHS((wage'_{id} - wage'_{iNA}) - (wage_{id} - wage_{iNA})) < 0$

(2) 'Error' is a continous variable of error in beliefs about population earnings in track *d*, relative to error in beliefs about non-attendance earnings. This variable has also been IHS transformed.
That is, $IHS((actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA})) < 0$

(3) Standard errors are clustered at the survey-group level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 4: Effect of Relaxing Credit Constraints in Control Group

Panel A- Summary of responses to borrowing questions	
% Accept loan	56.13%
Mean loan amount (in rupees)	1,11,461.40
Median loan amount (in rupees)	50,000.00
Panel B- % of individuals who change probability of enrollment in at least one track relative to non-attendance (n=403)	
% increase (extensive margin)	65.01%
% increase (intensive margin)	22.83%
% no increase	12.16%

Notes: (1) This table only considers individuals in the control group of the study in order to assess the "baseline" importance of credit-constraints in the sample; that is for the representative student not exposed to the information treatment.

(2) An increase on the extensive margin refers to $(\pi'_{id} - \pi'_{iNA}) - (\pi_{id} - \pi_{iNA}) > 0$ where π'_d refers to round 2 enrollment probability in track *d* and *d* was not affordable to the individual without a loan.

(3) An increase on the intensive margin refers to $(\pi'_{id} - \pi'_{iNA}) - (\pi_{id} - \pi_{iNA}) > 0$ where π'_d refers to round 2 enrollment probability in track *d*, and *d* was affordable to the individual even without a loan.

Table 5: Borrowing Across Treatment & Control

	Overall	Overall	Technical	General	Vocational
	borrow (0/1)	borrow_amt	borrow_amt	borrow_amt	borrow_amt
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0827	0.0601	-0.0165	0.131	0.24
Overestimate	0.140***	0.426***	0.385***	0.465***	0.532***
Treat#Overestimate	-0.0156	-0.152	-0.0454	-0.292	-0.387*
Constant	0.470***	10.60***	10.65***	10.61***	10.49***
Observations	1,437	852	852	852	852

Notes: (1) borrow (0/1) is a dummy variable for whether or not an individual says they would like to borrow for higher education enrollment. Borrow_amt is the log of the amount the individuals states they would like to borrow (i.e. among those who "borrow").

(2) In columns (1) and (2) 'Overestimate' is a dummy for individuals classified as 'overestimators' i.e. those individuals who overestimate earnings in track *d*, relative to non-attendance earnings, for 2 or more (out of 3) tracks.

(3) In columns (3)-(4) 'Overestimate' is a dummy for individuals who overestimate, at baseline, beliefs about population earnings in track *d*, relative to non-attendance earnings.
That is, *Overestimate=1* if $(actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA}) < 0$

(4) Standard errors are clustered at the survey-group level. ***, **, and * indicate significance at the 1, 5, and 10 percent level, respectively.

Table 6: Probability of Enrollment Among Credit-Constrained Individuals Across Control & Treatment

	Log-Odds Relative to Non-Enrollment		
	Technical	General	Vocational
Treatment	-0.606	-0.839	0.687
	(0.451)	(0.244)	(0.344)
Overestimate	-0.0522	0.462	0.629
	(0.948)	(0.471)	(0.298)
Treat#Overestimate	0.188	0.295	-1.690*
	(0.852)	(0.738)	(0.059)
Baseline Controls	YES	YES	YES
Observations	782	782	782

Notes: (1) The outcome variable for all three columns is the round-2 log-odds of enrolling in track *d* relative to non-attendance.
That is,$(log(\pi_{id}) - log(\pi_{iNA}))$ measured post-treatment.

(2) 'Overestimate' is a dummy for individuals who overestimate, at baseline, beliefs about population earnings in track *d*, relative to non-attendance earnings.
That is, *Overestimate=1* if $(actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA}) < 0$

(3) Baseline controls include age, gender, father's education and ownership of farm land.

(4) Standard errors are clustered at the survey-group level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels respectively. P-values are in parentheses.

Appendix

Table A1: Track-wise over & under estimation of population earnings

	Technical	General	Vocational
Panel-A			
% Overestimate	60.50%	54.86%	69.88%
% Underestimate	39.50%	45.14%	30.12%
Panel-B			
Mean Overestimation (in Rs.)	39880	32801	23731
Mean Underestimation (in Rs.)	18029	14548	15719
Panel-C			
Median Overestimation (in Rs.)	17954	10193	11697
Median Underestimation (in Rs.)	8546	5307	4522

Notes: (1) In *Panel A*, 'Overestimate'/'Underestimate' is a dummy for individuals who overestimate/underestimate, at baseline, beliefs about population earnings in track *d*, relative to non-attendance earnings.

(2) In *Panel B*, the extent of overestimation/underestimation, is a continous variable of error in beliefs about population earnings in track *d*, relative to error in beliefs about non-attendance earnings.

(3) Overestimation is defined as: $(actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA}) < 0$ and underestimation is: $(actual_d - perceived_{id}) - (actual_{NA} - perceived_{iNA}) > 0$

Table A2: Earnings Elasticity of Enrollment

	Log-Odds Relative to Non-Enrollment			
	(1)	(2)	(3)	(4)
Relative Log-Earnings	0.548***	0.489***	0.167**	0.167^{+}
Constant	3.355***	0.0331	0.461***	0.461***
Baseline Controls	NO	YES	NO	NO
Individual FE	NO	NO	YES	YES
Standard Errors	Clustered	Clustered	Robust	Clustered
Observations	4,563	4,065	2,148	2,148

Notes: (1) The outcome variable in all four columns is the round-1 log-odds of enrolling in track d relative to non-attendance.
That is, $(log(\pi_{id}) - log(\pi_{iNA}))$ where $log(\pi_{id})$ is the log-probability of enrolling in track *d*.

(2) Enrollment probabilities used here were elicited for *all* choices/tracks, regardless for affordability to the individual.

(3) 'Relative Log-Earnings' is the log-earnings belief for track *d* relative to log-earnings for non-attendance.
That is, $(logwage_{id} - logwage_{iNA})$

(4) Baseline Controls include controls for age, gender, caste, religion, father's education, father's occupation, mother's education, mother's occupation, ownership of farm land, score in class 10 examinations, college, subject stream and household facility index.

(5) Columns (3) and (4) only include individuals from the treatment group of the sample, as we see revisions in earnings beliefs for this group.

(6) Clustered Standard errors are at the survey-group level. ***, **, *, and + indicate significance at the 1, 5, 10 and 15 percent levels respectively.

Table A3: Earnings Overestimation & Borrowing Likelihood

	Borrow(0/1)	Borrow(0/1)
Overestimate	0.128***	0.113***
age		0.0118
gender		0.0505
2.caste		-0.0413
3.caste		-0.0399
4.caste		-0.0212
2.religion		0.551***
3.religion		0.667***
4.religion		0.645***
5.religion		0.887***
6.religion		0.673***
7.religion		-0.069
2.father_edu		-0.027
3.father_edu		-0.025
4.father_edu		0.027
5.father_edu		-0.0184
6.father_edu		-0.0258
7.father_edu		-0.0349
8.father_edu		-0.144
9.father_edu		0.384***
2.father_occ		-0.0524
3.father_occ		0.0526
4.father_occ		-0.0915
2.mother_edu		-0.0376
3.mother_edu		0.0126
4.mother_edu		0.0406
5.mother_edu		0.0436
6.mother_edu		0.203
7.mother_edu		0.0685
8.mother_edu		-0.22
9.mother_edu		-0.537
2.mother_occ		-0.0151
3.mother_occ		-0.00601
4.mother_occ		-0.0448
1.farm_land		-0.00177
ten_score		0.000501
101.college		0.193*
102.college		0.136
103.college		0.134
104.college		0.193*

105.college		0.335***
106.college		0.168
107.college		0.0976
108.college		0.221**
109.college		0.121
111.college		0.102
2.stream		0.00467
3.stream		0.127***
1.facility_index		-0.0283
2.facility_index		0.0242
3.facility_index		0.0119
4.facility_index		-0.0573
5.facility_index		-0.111
6.facility_index		-0.0555
Constant	0.513***	-0.503
Observations	1,437	1,274

Notes: (1) The outcome variable borrow (0/1) is a dummy variable for whether or not an individual says they would like to borrow for higher education enrollment.

(2) 'Overestimate' is a dummy for individuals classified as 'overestimators' i.e. those individuals who overestimate earnings in track *d*, relative to non-attendance earnings, for 2 or more (out of 3) tracks.

(3) Standard errors are clustered at the survey-group level. ***, **, and * indicate significance at the 1, 5, and 10 percent levels respectively.