



# Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender



Orazio P. Attanasio<sup>a,\*</sup>, Katja M. Kaufmann<sup>b,\*\*</sup>

<sup>a</sup> University College London, Gower Street, London WC1E6BT, UK, Institute for Fiscal Studies, BREAD, CEPR and NBER

<sup>b</sup> Department of Economics and IGIER, Università Bocconi, Via Roentgen 1, 20136 Milan, Italy

## ARTICLE INFO

### Article history:

Received 3 April 2012

Received in revised form 4 April 2014

Accepted 5 April 2014

Available online 18 April 2014

### Keywords:

Schooling decisions

Subjective expectations

Perceived earnings and employment risk

Gender differences

Mexico

## ABSTRACT

In this paper we investigate the role of expected returns to schooling and of perceived risks (of unemployment and earnings) as determinants of schooling decisions. Moreover, our data also allow us to analyze whether youths' and/or mothers' expectations predict schooling decisions, and whether this depends on the age and gender of the youth. In particular, we use Mexican data that contain labor market expectations of mothers and youths. We find that expected returns and risk perceptions are important determinants of schooling decisions, the latter in particular from the perspective of the mother. Boys' expectations predict the decision to enter college, but not to enter high school. While girls' own expectations do not predict either of the two educational decisions, mothers' expectations are particularly strong predictors of their daughters' decisions.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

The goal of this paper is to improve our understanding of educational decisions among poor Mexican households in two dimensions. First, we investigate what important determinants of schooling decisions are and whether they differ for male and female youths. We are interested in the role of expectations about monetary returns to schooling and of perceptions of labor market risks (i.e. earnings and unemployment risk) for different schooling choices. Second, we also analyze whether youths' own and/or mothers' expectations about future labor market outcomes for the youth predict schooling decisions and how this depends on the gender and age of the youth.

To address these questions we use a survey of Mexican junior and senior high school graduates that elicits youths' or mothers' beliefs about future labor market outcomes for different scenarios of the highest schooling degree. In addition, we have information about actual schooling choices and an extensive set of controls.

Surprisingly little is known about the importance of gender differences in the determinants of schooling choices and about the role that youths play in making these choices. To what extent do expected returns in the labor market and perceived risks play a different role in

the schooling decisions for boys and girls? Do the expectations of youths and mothers about the return of education in the labor market differ and which of these expectations, if any, matter for schooling decisions? Are there gender differences in terms of whether youths' own expectations and/or their mothers' expectations matter? How does this depend on the age of the youth? These questions are important for the effective design of programs intended to increase schooling among poor households as well as for understanding gender differences in educational attainment.

Much of the existing literature on schooling decisions uses earnings realizations to measure or proxy for expected returns to education. This approach relies on strong assumptions about people's information sets and about how people form expectations. Moreover, such an approach has to deal with the issue that, for every education level, we only observe the earnings of individuals who select that education level. By using data on subjective expectations of earnings under different scenarios, we avoid these problems.

We study two schooling decisions among poor urban Mexican households: that of youths who have just finished junior high school to enroll (or not) in senior high school, and that of youths who have completed senior high school to enroll (or not) in college. We use answers to questions about subjective expectations to construct measures of expected earnings, perceived earnings risk and perceived unemployment risk held by mothers and youths for different schooling scenarios, and study the extent to which these expectations affect the enrollment choices of male and female youths.

\* Corresponding authors.

\*\* Corresponding authors. Tel.: +39 02 5836 3370.

E-mail addresses: [o.attanasio@ucl.ac.uk](mailto:o.attanasio@ucl.ac.uk) (O.P. Attanasio), [katja.kaufmann@unibocconi.it](mailto:katja.kaufmann@unibocconi.it) (K.M. Kaufmann).

We first show that there are important differences in mothers' and youths' own expectations about future earnings of one and the same person, the youth herself. This evidence in itself constitutes a strong justification for using direct measures of expectations: using realized earnings to infer expectations would not allow accommodating this feature without making assumptions in what direction and by how much the expectations of youths and parents differ. Moreover, our data provide us with the possibility of testing whether mothers' and/or youths' expectations are predictive of schooling choices. We therefore proceed to estimate a model of schooling decisions to measure the importance of expected returns to education for boys and girls.

Obviously our exercise is not without problems, arising in part from limitations to the data available to us. One issue that we need to deal with is the fact that the expectation questions are answered either by the youth or by the mother and never by both. In addition to the obvious limitation of not being able to test directly whether both expectations determine schooling choices, observed (average) differences in expected returns between youths and mother could be driven by the selection process determining who answers the questionnaire. In what follows, we discuss what we believe is a plausible instrumental variable approach to deal with this problem. More generally, while the data we use are unique, in that they elicit several points of the subjective distribution function for the returns to education, they are not without a number of limitations, ranging from the fact that we only have three points of the distribution function, to the timing of the survey and education choices. These limitations impose on us the need for some strong assumptions.

We discuss these issues at length in Section 5 (and other sections) and provide a number of necessary caveats. Our results, however, seem to indicate that expected returns and risk perceptions are important determinants of schooling decisions, the latter in particular from the perspective of the mother. Boys' own expectations predict the decision to enter college, but not to enter high school. While girls' own expectations do not predict either of the two educational decisions, mothers' expectations are particularly strong predictors of their daughters' decisions. These results are robust to a number of changes and modeling choices.

### 1.1. Related literature

Our paper is related to a growing literature using data on people's subjective expectations to understand educational decisions.<sup>1</sup> Dominitz and Manski (1996) were the first to elicit subjective expectations of earnings for different schooling degrees, but did not analyze the link between earnings expectations and investment into schooling. More recently, a number of papers investigated the link between subjective expectations of earnings and schooling choices in different contexts and with different methodologies and objectives, including Arcidiacono et al. (2012), Jensen (2010), Kaufmann (forthcoming), Nguyen (2008) and Stinebrickner and Stinebrickner (2012). Relative to these papers, we address several complementary questions making use of the unique features of the Mexican data.

One novel aspect of this paper is our focus on gender differences in the determinants of the decision to enroll in senior high school and in college in urban Mexico. The only paper we know of that uses data on subjective expectations (but not on risk perceptions) to understand gender differences is by Zafar (2009), who analyzes the college major choice of undergraduate students at the Northwestern University. More recently, Reuben et al. (2013) investigate whether behavioral

<sup>1</sup> See the survey paper on the use of subjective expectations data by Manski (2004). Delavande et al. (2011) survey the literature that uses data on subjective expectations in developing countries and find that also in this context individuals are willing to answer the expectations questions and understand them reasonably well (sometimes with the help of visual aids).

biases and preferences explain gender differences in college major choices and expected future earnings.<sup>2</sup>

The second novel aspect of our paper is the emphasis on the differences between mother and youth expectations and the attempt to estimate their different roles. In general, we know little about the role of children in household decisions. Instead children have been modeled as household public goods (see, e.g., Blundell et al., 2005) or incorporated through the "caring preferences" of their parents (see, e.g., Bourguignon, 1999). Notable exceptions, in that they consider the possibility of youths as decision makers, but in different contexts, are Berry (2011), Dauphin et al. (2008), Dinkelman and Martinez (forthcoming) and Giustinelli (2011).

### 1.2. Outline

The rest of the paper is organized as follows: In Section 2, we start with a description of the survey that we use and, in particular, the module on subjective expectations. In Section 3, we present some descriptive evidence on our data and focus, in particular, on the subjective expectations data and their plausibility. We also discuss our identification strategy to compare the subjective expectations of mothers and youths and present our first result which indicates the presence of significant differences in expected returns. In Section 4, we discuss the empirical model used to analyze the role of youths' and mothers' expectations about future labor market outcomes in their education decisions and how this depends on age and gender. Section 5 discusses the paper's main results. In this section, we also discuss some issues related to the use of subjective expectations and present some robustness tests. Finally, Section 6 concludes.

## 2. Subjective expectations about returns to education

The conditional cash transfer program *Oportunidades*, previously known as PROGRESA, has been associated since its inception with attempts to evaluate its impacts. In this spirit, when the program introduced a new component in 2002/3, known as *Jóvenes con Oportunidades*, an evaluation aimed at measuring its impact was started. The data that we use was collected in 2005 as part of that evaluation. In addition to standard variables, the survey contained a detailed subjective expectations module which we use extensively. In this section, we describe the data and its structure. We also describe in some detail the module used to elicit information on subjective expectations and report some evidence on the quality of these data.

### 2.1. The survey

The evaluation survey of *Jóvenes con Oportunidades* was conducted in the fall of 2005. *Jóvenes con Oportunidades* provides an additional grant to youths in the last three years of high school (preparatoria). This grant is deposited into a bank and can be accessed only upon graduation, if the recipient engages in one of several activities (such as going to college or starting a micro enterprise). Alternatively, the recipient has to wait for a year during which time the amount, about US\$300, accumulates at the market interest rate.

The primary sampling units of the evaluation survey are individuals who have just graduated from senior high school or from junior high school and who are eligible for *Jóvenes con Oportunidades*. There are three eligibility criteria for this program: being in the last year of junior high school (9th grade) or attending senior high school (10 to 12th grade), being younger than 22 years of age, and being from a family that benefits from *Oportunidades*.

<sup>2</sup> Two recent papers who link schooling choices and returns to schooling, with some emphasis in gender differences are Abramitzky and Lavy (2012) and Jensen (2012). These papers, however, do not use subjective expectations data.

The survey consists of a family questionnaire and a youth questionnaire for each household member aged 15 to 25. As a consequence, the youths for whom we have data are not only the primary sampling units but also their siblings, provided they are aged 15 to 25. In total we have information about approximately 23,000 youths. The questions in the youth questionnaire were addressed directly to the youth. However, if a specific youth was not present during the interview, the module was answered by the main respondent, who is generally the mother. In what follows, we discuss in detail how we address this issue.

The survey provides detailed information on demographic characteristics of the young adults, their schooling levels and histories, their junior high school GPA, and detailed information on their parental background and the household they live in, such as parental education, earnings and income of each household member, assets of the household and transfers/remittances to and from the household. The youth questionnaire contains a section on individuals' subjective expectations of earnings as discussed next. As with the rest of the questionnaire, the subjective expectations questions are answered by the mother when the youth is not present. We therefore have some information on mothers' expectations of the return to education. We now describe the expectations module.

## 2.2. The expectations module

The subjective expectations module was designed to elicit information on the individual distribution of future earnings and the probability of working for different scenarios about the highest completed school degree. The module starts with a simple explanation of probabilities. In particular, individuals are shown a ruler, graded from zero to one hundred, which is then used to express the probability of future events. The example that is used to illustrate the concept of probability is the event of rain the following day.

After explaining the use of the ruler to express probabilities and having practiced that with the rain example, the interviewer moves on to discuss future earnings and the probability of working under different schooling scenarios. The scenarios differ for students graduating from junior high school and those graduating from senior high school. For the former, the interviewer asks to consider three different possibilities: that the student stops after junior high, that the student goes on to senior high, completes it and stops and that the student goes on to college and completes it. For the latter, only two scenarios are considered: that the students stop at senior high school and that the student goes on to college and completes it.

For each of the relevant scenarios, the youth is then asked questions about the probability of working at the age of 25 and about future earnings at age 25. For example, in the case of the last scenario for a senior high school student, the questions are:

1. Assume that you complete College, and that this is your highest schooling degree. From zero to one hundred, how certain are you that you will be working at the age of 25?
2. Assume that you complete College, and that this is your highest schooling degree. Assume that you have a job at age 25.
  - (a) What do you think is the maximum amount you can earn per month at that age?
  - (b) What do you think is the minimum amount you can earn per month at that age?
  - (c) From zero to one hundred, what is the probability that your earnings at that age will be at least  $x$ ?

where  $x$  is the midpoint between the maximum and minimum amounts elicited from questions (a) and (b) and was calculated by the interviewer and read to the respondent. It is important to remember that these questions were included in the youth questionnaire and therefore are available for every youth in the household aged 15 to 25. Moreover, when a specific youth was not present, these questions were

answered by the mother. This implies that, for some households, we have mother's expectations about returns to education. Notice, however, that we do not have fathers' expectations and that the structure of the data poses a number of selection issues that we discuss at length below.

This type of subjective expectations questions has been used extensively in a variety of contexts. In a companion paper (Attanasio and Kaufmann, 2008), we discuss the internal and external validity of the answers to these questions in our survey. In that paper, we show that respondents seem to have understood the questions reasonably well and that the data pass a number of internal and external validity tests. Below, we report the results which are most relevant for the analysis in this paper.

In what follows, we relate educational decisions to subjective expectations. This is possible because of the timing of the survey. The *Jóvenes* survey was conducted in October/November 2005, that is two or three months after high school and college had started and enrollment decisions had been made.

While the timing of the survey (which occurred just after the school enrollment decision had been taken) allows using the expectations data to estimate an enrollment model, it might also raise a potential concern that individuals might try to rationalize their choices. For example, individuals who decided to enroll in college rationalize their choice by stating higher expected returns to college (that is higher college earnings and/or lower expected high school earnings), and those who decided not to enroll, state lower expected returns to college. This would lead to a more dispersed cross-section of returns after the decision compared to before the enrollment decision.<sup>3</sup>

To check whether ex-post rationalization introduces biases in expectations, we look at the expectations data of the siblings of our main subjects. Let us consider, for instance, the college enrollment decision of senior high school graduates. We can compare the cross sectional distribution of the expectations of our senior high school graduates to the expectations of the cohort of youths who are one year younger (just starting grade 12 at the time of the survey in October/November). If the older youths are ex-post rationalizing their choices, we would expect a cross-sectional distribution of expected earnings and returns that is more spread out for the cohort of senior high school graduates who have decided already whether to enroll in college or not, compared to the distribution of the cohort that is one year younger and just starting grade 12. The same reasoning should hold for perceptions of the probability of working. In Section 2 we discuss this issue at length and present evidence that our main results cannot be explained by ex-post rationalization.

## 2.3. Expected earnings, perceived earnings risk, and expected gross returns to schooling

We use the answers to the three survey questions (2(a)–(c)) described above, together with some additional assumptions, to compute moments of the individual earnings distributions and expected gross returns to high school and college. We are interested in the individual subjective distribution of future earnings  $f(Y^z)$  under the three possible scenarios of final education attainment we are considering: junior high school ( $z = 1$ ), senior high school ( $z = 2$ ) and college ( $z = 3$ ). The survey provides, for each individual, information on the support of the distribution  $[y_{\min}^z, y_{\max}^z]$  and on the probability mass to the right of the midpoint,  $y_{\text{mid}}^z = (y_{\min}^z + y_{\max}^z)/2$ , of the support,  $p = \Pr(Y^z > y_{\text{mid}}^z)$ . Thus we need an additional distributional assumption,  $f(\cdot)$ , in order to

<sup>3</sup> This is true unless people switch positions in the distribution in such a way that the resulting cross-section looks exactly the same as before. This can only be the case if the people who decide to enroll in college are the ones with particularly low expected returns, and they later report high returns to college to justify their decision. And similarly, the people who decide not to enroll in college are the ones with particularly high returns and they later state low expected returns.

be able to calculate moments of these individual earnings distributions, using the three pieces of information on  $y_{\min}^z$ ,  $y_{\max}^z$  and  $p$ . Given such an assumption on  $f(\cdot)$ , we can compute all the moments of any function of future earnings under the different scenarios that we are interested in. For example, the expected value of log future earnings for *each* individual and each scenario ( $z = 1, 2, 3$ ) will be given by  $E(\ln Y^z) = \int_{y_{\min}^z}^{y_{\max}^z} \ln y f_{Y^z}(y) dy$ .

For many of the specifications that we estimate below, we consider the (gross) returns to college and senior high school. We compute these as the difference between expected log college (senior high) earnings and expected log senior high school (junior high school) earnings,  $\rho^z = E(\ln(Y^z)) - E(\ln(Y^{z-1}))$  for  $z = 2, 3$ .<sup>4</sup>

Notice that, as we have no information on the covariance of earnings under different scenarios, we cannot compute the variance of the return. For this reason, to capture the risk aspect of education choices we consider the variance of (log) earnings under the two scenarios.

In Attanasio and Kaufmann (2008), we use three different distributional assumptions, step-wise uniform, bi-triangular and triangular (compare Guiso et al., 2002). The last two assumptions give more weight to the middle of the support and less to the extremes. The first, instead, implies a relatively large value for the total variance. We rule out the possibility that the density function is U-shaped, giving more weight to the extremes. In our companion paper, we show that the first moment of the individual distribution is extremely robust with respect to the underlying distributional assumption, while the second moment is obviously larger for the step-wise uniform distribution that puts more weight on extreme values. In this paper, we present results based on the triangular distribution, but we perform robustness checks using the other two distributional assumptions and point out differences if they occur.

### 3. Descriptive statistics and subjective expectations

Before using the data on subjective expectations to model schooling choices, we describe the general patterns in the sample that we will be using for our analysis and provide some evidence on the quality of the expectations data. Further details on internal and external validity checks are contained in the companion paper Attanasio and Kaufmann (2008). We also discuss some of the econometric issues involved with the use of these data and in particular the way that we compare mothers' and youths' expectations.

#### 3.1. Characteristics of the youth

While the overall sample includes all youths aged 15 to 25, regardless of their schooling status, in our analysis we focus on students who just finished grade 9 ("secundaria" or junior high school) and decide about enrolling in senior high school, and students who just finished grade 12 ("preparatoria", senior high school) and decide about enrolling in college or not. The data on the siblings sample is used for a number of robustness exercises on the expectations data. From the sample of 23,000 youths, 5700 individuals had just completed grade 9 and 4397 had completed senior high school and decided about enrolling in college.

Dropping individuals for whom we miss information on any of the individual and family background variables which we use in the

<sup>4</sup> As one referee pointed out, considering the difference in expected earnings at 25 with a high school degree and a college degree does not take into account the fact that the youth with the high school degree probably will have longer experience in the labor market, which would be reflected in relatively higher earnings if there are significant returns to experience. We are not aware of studies that have explicitly studied subjective expected returns to tenure, although Dominitz and Manski (1996) report that expected earnings at age 40 are higher than expected earnings at 30, suggesting that individuals do perceive the existence of returns to experience. Given the structure of our data we cannot investigate this issue further.

analysis, we are left with 2981 junior high school graduates and 3321 senior high school graduates. Lastly, we drop those individuals with missing information on the expectation questions (i.e. if any of the questions on the probability of work, the probability of earnings above the midpoint and minimum and maximum earnings for the different schooling scenarios were not answered) or who provided inconsistent answers (i.e. in any of the answers minimum earnings are larger than maximum earnings etc., see detailed discussion in the Online Appendix). This leaves us with the final sample of 2364 junior high school graduates and 2829 senior high school graduates.

In the Online Appendix we report summary statistics of individual and family background characteristics for the "young" cohort (junior high school graduates) and the "old" cohort (senior high school graduates), respectively. We present results separately for male and female youths and for the two samples of mother and youth respondents to investigate whether there are potential sample selection problems when performing the analysis separately for each type of respondent (since it was not randomly determined whether youth or mother responded to the youth questionnaire).

For the sake of brevity, we only briefly summarize our main findings and move a more extensive discussion on the construction of the variables and the comparison of each of the variables for the four subgroups to the Appendix.

One can summarize the evidence by saying that the features of the sample reflect that we are working with families that are beneficiaries of a welfare program targeted to the poorest sector of Mexican society. The differences between girls and boys that we observe conform to our expectations and confirm the findings of other empirical studies, for example the fact that girls outperform boys at school.

There are a few (usually small but) significant differences in individual characteristics and family background variables between the subsample where the earning expectations questions were answered by the youth and the subsample where they were answered by the mother. The selection process is very similar for boys and girls: For the young cohort, the youth sample is from slightly more educated and richer families. For the old cohort, the youth sample is more likely to have fathers who are employees (instead of workers) and to have less siblings. Finding differences between mother and youth sample points toward a potential sample selection in our analysis of these questions, which we address, as discussed below. We now turn to analyze the differences in subjective expectations.

#### 3.2. Youth vs mother respondents: A selection equation

The interviewer visited the primary sampling units and their families in October and November 2005 and interviewed the household head or spouse using the family questionnaire and youths between age 15 and 25 using the "Jovenes" (youth) questionnaire. If a youth was not present, the household head or spouse answered the Jovenes questionnaire as well. As a result, for almost half the sample, the questions on the subjective distribution of future earnings were not answered by the youth herself. Instead mothers stated their expectations about future earnings of her child(ren) that are not present during the interviewer's visit.

While it is unfortunate that we do not have information on the expectations of both mothers and children, the fact that for half the sample the earnings expectations questions were answered by youths, while for the other half the questions were answered by the mother allows asking whether expectations held by mothers and youths about the youths' future earnings are different. If subjective expectations of mothers and youths were objective and rational expectations based on the same information, it would not matter who would answer (and the issue of whose expectations matter would be a moot one). It is therefore interesting to establish, whether the expectations of future earnings are systematically different depending on who answered the question.

In the following section, we compare the expectations of mothers and youths. A straight comparison shows that these expectations are systematically different. Given the structure of the data, however, these differences can arise either because the questions answered by the mother and by the youth are measuring two different and distinct objects (the subjective belief held by the mother and the subjective belief held by the youth) or because the sample of youths absent from the interview (and for whom the question is answered by the mother) is systematically different from those present during the visit.

To take into account the possibility that the observed differences are due to sample selection and to test for the presence of actual differences in expectations, we use a standard Heckman two-step approach (see Heckman, 1979). To achieve non-parametric identification of such a selection model, we need one or more variables that determine whether the question is answered by the youth rather than the mother and that, plausibly, do not affect the expectations directly. For such a purpose, we use information on the timing of the interview (the time of day of the interview, whether it took place on a weekday or weekend and whether it took place during weeks of holiday or not).

We model the probability that the youth answers the questionnaire (rather than her mother) as a function of all individual and family background characteristics and a set of variables capturing the timing of the interview (all variables fully interacted with gender dummies) and estimate the following equation:

$$R = 1 \Leftrightarrow R^* = \delta + X'\kappa + Z'\lambda + \epsilon \geq 0, \tag{1}$$

where  $R = 1$  indicates a youth respondent and the vector  $Z$  includes the timing of the interview variables.

When considering the timing variables that identify the possible differences between the two sets of expectations (and below the schooling equations) we experimented with various sets of variables. We find that, in our sample, the time of day of the interview does not much affect the probability that the youth is present (and therefore answers the expectations questions). However, the day of the week dummies and the indicator for holiday weeks are strongly significant determinants of who the respondent is with F-statistics of 32 and 18 for young and old cohorts, respectively (see Table 1). In particular, both for the young cohort and the old cohort, youths are significantly more likely to respond themselves (because they were present at the moment of the interview) during weekends and during holidays (i.e. weeks 41 and 42 for the young cohort and weeks 40 to 42 for the older cohort).

To be a valid instrument, the timing of the interview should be unrelated to unobserved characteristics of the respondents' (mothers or youths) expectations. While such identification assumption is obviously untestable, we can provide some evidence in its favor. First, conversations with officials in charge of the survey stressed that the timing of the interviews (and the rare event of re-scheduling) was not purposely related to the availability of the youths in the household. Re-scheduling occurred only when nobody who could answer the main survey was present. As for the youth questionnaires, the youths who were present would answer the relevant questions, otherwise the main respondent would.<sup>5</sup>

As an additional check on the validity of the exclusion restriction, we related the timing of the interview to various observables. The timing of the interview can primarily be explained by geographic differences (while accessibility of areas is not a concern given that all households in our sample live in urban areas, see discussion in Section 3.1). The most important predictor of the week in which the interviewer arrives is the size of the locality in which the family lives and that households in larger localities are interviewed during earlier weeks (see the Online Appendix). Importantly, the timing of

**Table 1**  
First-stage regression for whether the youth responds herself.

Dep var:	Youth respondent: Yes/No		
	Instruments	Instr and interact	Instr and controls
	Coeff/(SE)	Coeff/(SE)	Coeff/(SE)
<i>Young cohort</i>			
Interview on a weekday	-0.081*** (0.029)	-0.093** (0.039)	-0.051 (0.039)
Interview in week 41 or 42	0.153*** (0.021)	0.161*** (0.029)	0.139*** (0.029)
Interview in week 45 or 46	-0.127*** (0.032)	-0.089** (0.043)	-0.067 (0.043)
Female		0.040 (0.057)	0.245 (0.383)
Interview on a weekday * Male		0.029 (0.058)	0.010 (0.057)
Interview in week 41 or 42 * Male		-0.017 (0.043)	-0.002 (0.042)
Interview in week 45 or 46 * Male		-0.081 (0.063)	-0.035 (0.063)
Controls	No	No	Yes
Observations	2364	2364	2364
F-statistic	32.452	14.595	7.566
<i>Old cohort</i>			
Interview on a weekday	-0.069*** (0.026)	-0.087** (0.034)	-0.085** (0.035)
Interview in weeks 40 to 42	0.115*** (0.023)	0.129*** (0.031)	0.122*** (0.031)
Interview in week 45 or 46	-0.055** (0.022)	-0.072** (0.030)	-0.080*** (0.031)
Female		0.114** (0.053)	-0.353 (0.374)
Interview on a weekday * Male		0.046 (0.051)	0.048 (0.051)
Interview in Week 40 to 42 * Male		-0.031 (0.046)	-0.011 (0.046)
Interview in Week 45 or 46 * Male		0.039 (0.044)	0.022 (0.045)
Controls	No	No	Yes
Observations	2829	2829	2829
F-statistic	18.284	10.376	4.638

Notes: Table displays coefficients and standard errors in brackets. Included controls are all variables in Section 3.1 and state fixed effects (fully interacted with gender dummies).

\* p < 0.1.  
\*\* p < 0.05.  
\*\*\* p < 0.01.

the interview is not correlated with individuals' educational decisions and is correlated only in few cases with individual and family background characteristics.

The main assumption underlying our analysis, therefore, is that – after controlling for the long list of these individual and family background characteristics and geographic variables such as locality size, distance from universities and state fixed effects – the timing of the interview is not correlated with youths' expectations (which are unobserved for half of the samples) and not correlated with mothers' expectations (again unobserved for the other half of the sample).

In what follows, we will be using this selection model (which fully interacts all variables with gender dummies) and its underlying assumption to test both for differences between youths' and mothers' expectations and to estimate the school decision model, as we discuss below.

### 3.3. Are mothers' and youths' expectations different?

In Tables 2 and 3, we present summary statistics for the variables derived from the subjective expectations questions for the young and old cohorts. The top two panels of the tables report expected log earnings and expected (gross) returns. The bottom two panels, instead, focus on perceived earnings and employment risk for the three different

<sup>5</sup> An anonymous referee pointed out the possibility that rescheduling to obtain the presence of the youth would invalidate our instrument. This comment prompted us to contact the survey officials.

**Table 2**  
Subjective expectations of future earnings: junior high school graduates.

Junior HS graduates resp:	Boys			Girls		
	Youth	Mother	Corr diff	Youth	Mother	Corr diff
	Mean/(SE)		(Y – M) (P-val)	Mean/(SE)		(Y – M) (P-val)
Exp log earnings						
Junior HS	7.059 (0.608)	7.232 (0.489)	–0.295 (0.125)	7.005 (0.589)	7.217 (0.451)	–0.705 (0.001)
Senior HS	7.614 (0.536)	7.764 (0.425)	–0.092 (0.581)	7.585 (0.550)	7.724 (0.421)	–0.549 (0.005)
College	8.243 (0.502)	8.361 (0.416)	–0.228 (0.163)	8.209 (0.521)	8.31 (0.438)	–0.336 (0.064)
Exp return						
Senior HS	0.555 (0.321)	0.532 (0.307)	0.188 (0.106)	0.579 (0.315)	0.506 (0.269)	0.163 (0.147)
College	0.628 (0.368)	0.597 (0.304)	–0.146 (0.240)	0.625 (0.355)	0.586 (0.326)	0.210 (0.106)
Std dev of log earn						
Junior HS	0.077 (0.048)	0.071 (0.045)	0.053 (0.004)	0.076 (0.047)	0.07 (0.045)	0.049 (0.008)
Senior HS	0.067 (0.043)	0.059 (0.037)	0.022 (0.136)	0.063 (0.039)	0.062 (0.043)	0.009 (0.546)
College	0.054 (0.037)	0.052 (0.033)	0.012 (0.358)	0.055 (0.037)	0.053 (0.037)	0.016 (0.265)
Prob of Work						
Junior HS	0.495 (0.210)	0.515 (0.218)	0.106 (0.175)	0.476 (0.214)	0.523 (0.209)	–0.046 (0.552)
Senior HS	0.677 (0.173)	0.674 (0.171)	0.120 (0.064)	0.661 (0.192)	0.675 (0.169)	0.017 (0.809)
College	0.826 (0.166)	0.824 (0.148)	0.080 (0.172)	0.826 (0.164)	0.812 (0.157)	0.023 (0.698)
Observations	638	473		759	494	

Notes: The “Corrected difference” between the expectations of mother and youth corrects for sample selection by instrumenting for who responds to the expectation questions. As instruments we use variables that capture the timing of the interview, which are strong predictors for who responds (see Table 1). In addition we correct for potential differences in the composition of the mother and youth samples.

schooling degrees, junior and senior high school and college (or the latter two in the case of the old cohort). We summarize these measures separately for male and female youths, and separately for mother and youth respondents. In columns 3 and 6 of these tables, we test the hypothesis that the youth's and mother's expectations are the same.

We use the selection (Eq. (1)) that we have just discussed to correct for the possibility that the observed means differ because youths who answer the questionnaire might be systematically different, in some non observable dimension, from youths whose mother answers the questionnaire and control for compositional differences.

**Table 3**  
Subjective expectations of future earnings: senior high school graduates.

Senior HS graduates: resp:	Boys			Girls		
	Youth	Mother	Corr diff	Youth	Mother	Corr diff
	Mean/(SE)		(Y – M) (P-val)	Mean/(SE)		(Y – M) (P-val)
Exp log earnings						
Senior HS	7.618 (0.497)	7.648 (0.444)	–0.165 (0.403)	7.525 (0.507)	7.637 (0.418)	–0.263 (0.091)
College	8.267 (0.646)	8.317 (0.440)	–0.065 (0.739)	8.236 (0.485)	8.308 (0.426)	–0.136 (0.380)
Exp return						
College	0.648 (0.359)	0.669 (0.365)	0.082 (0.615)	0.711 (0.407)	0.671 (0.364)	0.126 (0.343)
Std dev of log earn						
Senior HS	0.065 (0.039)	0.059 (0.036)	–0.003 (0.845)	0.063 (0.039)	0.060 (0.038)	0.028 (0.042)
College	0.054 (0.033)	0.053 (0.033)	–0.001 (0.968)	0.054 (0.035)	0.050 (0.030)	0.123 (0.277)
Prob of work						
Senior HS	0.661 (0.173)	0.653 (0.180)	–0.098 (0.223)	0.664 (0.188)	0.644 (0.178)	0.112 (0.083)
College	0.821 (0.161)	0.813 (0.154)	–0.054 (0.062)	0.825 (0.161)	0.813 (0.156)	–0.111 (0.837)
Observations	583	737		778	731	

Notes: The “Corrected difference” between the expectations of mother and youth corrects for sample selection by instrumenting for who responds to the expectation questions. As instruments we use variables that capture the timing of the interview, which are strong predictors for who responds (see First-stage Tables). In addition we correct for potential differences in the composition of the mother and youth samples.

Not surprisingly, but reassuringly, expected log earnings increase in schooling level. Gross returns to schooling – measured as the difference between expected log earnings of two consecutive schooling degrees – are large and larger for college than high school. Comparing male and female youths when youths respond themselves, males expect slightly higher earnings and lower returns. Mothers on the other hand expect surprisingly similar earnings and returns for boys and girls.<sup>6</sup>

Standard deviations of log earnings are one possible measure of (perceived) earnings risk and commonly used for measuring risk in the case of observed earnings. It is important to point out that having information on the individual earnings distributions (in the form of minimum, maximum and probability above the midpoint) allows us to derive a measure of “true” risk, as perceived by the individual, while using the variability of observed earnings data will confound risk with unobserved heterogeneity. Moreover, observed variability will not distinguish between predicted and unpredicted changes.

Perceived earnings risk, as measured by the standard deviation of logs, decreases with education and is lowest for the expectations of earnings conditional on having a college degree. At the same time the probability of work increases with education. Thus lower income risk could be another important motivation, in addition to higher expected earnings, for achieving higher schooling. Male and female youth respondents perceive a very similar level of earnings risk, while males perceive a slightly higher probability of working, where the difference decreases again in schooling level.<sup>7</sup>

The third and sixth columns of Tables 2 and 3 present the results of the test of the differences between youths' and mothers' expectations, for each of the subjective expectations measures, i.e. for expected earnings, expected returns, perceived probability of working, and earnings risk. Table 2 shows the results for the young cohort. Mothers' expectations about future earnings of her children are higher than the expectations for the youth sample. For girls, mothers expect 70% higher junior high school earnings, 55% higher senior high school earnings and 34% higher college earnings, where all three differences are significant. For boys, mothers' earnings expectations are between 10 and 30% higher than those of the male youths, though the difference is not significant. Differences for the older cohort are smaller (14–26% for girls and 7–17% for boys) and only significant on conventional levels for girls' senior high school earnings. In terms of expected returns, mothers generally expect lower high school and college returns than youths (though differences are not significant on conventional levels).

Mothers perceive a lower earnings risk than the youths (with the exception of male senior high school graduates, though not all differences are significant). The perceived probability of working is similar for mother and youth respondents (with a few exceptions).

Our results suggest that there are genuine differences in mothers' and youths' expectations. Differences between mother and youth respondents are particularly important for the young cohort. For example, girls' expectations of earnings are consistently lower than mothers' expectations for all schooling scenarios and differences are large.

These results underline the importance of understanding whose expectations are relevant for school choices, the ones of the mothers or the ones of the youths themselves. We pursue this issue in Section 5, after presenting some results to support the validity of the expectation data that we use in the rest of this paper.

<sup>6</sup> In the next section, we present results of several validity checks including a comparison of our data on subjective expectations to earnings data from the Census, see the Online Appendix.

<sup>7</sup> In addition to second moments, one can consider higher moments of the distribution. Looking at the skewness, one finds that, on average, individual earnings distributions are left-skewed (i.e. the probability to have earnings above the midpoint is larger than 0.5) and increasingly left-skewed with increasing schooling level.

### 3.4. Data validation: Expected earnings and individual characteristics

As briefly discussed in the Introduction, Dominitz and Manski (1996) show in their seminal paper on subjective expectations of earnings for different schooling degrees that, in the case of a small sample of Wisconsin high school and college students, people are willing and able to answer subjective expectations questions in a meaningful way. Manski (2004), Attanasio (2009) and Delavande et al. (2011) survey the literature that uses data on subjective expectations (the latter two in developing countries) and also find that individuals are willing to answer the expectations questions and understand them reasonably well (sometimes with the help of visual aids). They conclude that data on people's subjective expectations can be a useful tool for understanding people's behavior also in the context of developing countries.

In the following we provide some evidence that also in the case of our sample of Mexican junior and senior high school graduates, the individuals and their mothers are willing and able to respond meaningfully to subjective expectations questions. For the sake of brevity, we report all result tables of this section in the Online Appendix.

First we present evidence on the fraction of missing and inconsistent answers, separately for the four subsamples of interest, that is male and female youth respondents and their mothers. An additional interesting exercise consists of relating the expected value of future earnings to a number of observable variables, reflecting individual and family background characteristics of the youths. While our goal is not to analyze how people form expectations, we can check as a validation exercise, whether expected earnings vary with observables in a way similar to actual earnings. We expect people to draw inferences about their own potential earnings from what they observe from others. Thus finding that expectations vary with observable characteristics in a way similar to observed earnings lends support to their validity. Lastly, we also compare our data on subjective expectations to Census data on earnings (see the detailed discussion in the Online Appendix).

In the Online Appendix, we show that individuals were indeed willing to answer the expectation questions. The fraction of missing responses is below 2% for both cohorts and very similar for all subgroups. We define “inconsistent” answers as replies in which minimum earnings are below or exactly equal to maximum earnings for the same schooling scenario, in which the probability of earnings above the midpoint is indicated as being zero or one (elicited probabilities were never larger than one or below zero given the elicitation method, see Section 2) or answers for which the implied returns are negative. The younger cohort exhibits larger fractions of inconsistent answers than the old cohort, as one would expect given their lower level of schooling. In particular, the probability that an individual of the young cohort answers any of the expectation questions (i.e. probability of working, minimum and maximum earnings and probability above the midpoint for the three different schooling scenarios) in an inconsistent way is around 20%, the analogue probability for an individual from the old cohort is around 14%. Again there are no significant differences in the fraction of inconsistent responses between youth and mother respondents, neither for boys nor girls.<sup>8</sup>

Furthermore, we also show a comparison within subgroups between the youths who decided to continue schooling with those who decided to stop school. Results in the Online Appendix show that both in terms of missing and inconsistent responses and for all four subgroups and both cohorts, there are no significant differences between youths stopping or continuing school to the next higher schooling level (with one exception, that is, for the young cohort, female youth respondents are less likely to respond to the expectation question when stopping school (2.6%) compared to when they continue school (0.9%), where the difference is significant at 10%).

<sup>8</sup> Our findings in terms of fraction of non-responses for expectation questions and fraction of inconsistent responses are in line with, for example, the findings of Dominitz and Manski (1996), Delavande et al. (2011) and Mahajan et al. (2011).

In the following we relate the expected value of future earnings to a number of observable variables (individual and family background characteristics). Finding that expectations vary with observable characteristics in a way similar to observed earnings lends further support to their validity. At the same time, we stress again that this exercise should not be interpreted as providing evidence on how people form expectations, since our data are not suitable for such an analysis.

We first discuss results for the young cohort (see the Online Appendix). For boys, having a father who is an employer significantly increases expected senior high school and college earnings. Expectations of junior and senior high school earnings are decreasing in the number of brothers that a boy has, possibly because this decreases the likelihood that he can take over the (small-scale) business of the father. For girls on the other hand, earnings expectations with senior high school and college are significantly lower when they have a father who is an employer, while expected earnings are larger when the girl comes from a richer family. Maybe surprisingly, expected junior high school earnings are significantly lower for girls whose mother is slightly more educated (on 10%), while expected college earnings are higher when the father is more educated. For the young cohort a higher GPA is not associated with higher expected earnings in contrast to what we find for the older cohort.

For the older cohort we find that a higher GPA is associated with significantly higher earnings, in particular for college as the highest degree. For boys expected earnings decrease with age (consistent with lower expected earnings for those who had to repeat a grade), and increase with the number of sisters, and expected high school earnings are larger if the boy is from a richer family. Expectations about girls' earnings on the other hand are lower if the mother is single, separated or divorced and there is no father in the household and lower if the father is a family worker. Again girls' expected earnings are lower if the mother has a junior high school degree instead of less education and higher if the father has a college degree.

Lastly, we compare expected earnings to observed (Census) earnings for different schooling degrees and show that both are in the same ballpark. In the Online Appendix we discuss the relevant data sources, and how we conduct the comparison and we discuss in detail why this comparison cannot be interpreted as a test of how "rational" individuals' expectations are (one of the reason is that earnings realizations refer to individuals who are between 25 and 30 in 2000, while the individuals whose expectations we elicited about earnings at age 25 will only be 25 in 2015 (in the case of the young cohort) or they turned 25 in 2012 (in the case of the old cohort)).

To conclude, the evidence of this section suggests that our sample of Mexican youths and their mothers understand the expectation questions and give meaningful answers. At the same time there are important differences between mothers' and youths' expectations about earnings of one and the same person – the youth herself – suggesting that conventional approaches using earnings realizations and strong assumptions on rationality and information sets could be problematic. In particular it is impossible, with such approaches, to address the question of whose expectations matter for schooling choices.

#### 4. An empirical model of schooling decisions

In this paper, we see school enrollment as an investment decision and, as such, we model it as a function of the expected return to that investment and its risk. In this section, we discuss the simple empirical framework that we use to relate schooling decision to the subjective expectations of youths and their mothers.

##### 4.1. An empirical model of schooling decisions

One possible approach to analyze the relationship between enrollment and expected returns would be the construction of a full dynamic optimization model where individuals choose current activities taking

into account the current and future benefits and the costs of the alternative choices. This type of models has been proposed, for instance, by Keane and Wolpin (1997) and used in a variety of contexts (see, for instance, Attanasio et al., 2011; Todd and Wolpin, 2006). Kaufmann (forthcoming) shows how data on people's subjective expectations can be used in a simple model of college enrollment choice. In this paper, rather than following this route, we present probit regressions that relate the probability of enrollment to several control variables, and to subjective expectations of earnings. In a fully specified model, schooling decisions are determined by the entire probability distribution of future earnings under alternative scenarios. Here we assume that the effect of this distribution can be summarized by a few moments of the distribution of earnings at age 25. In addition, we control for current labor market conditions through state dummies and for family background and ability through several variables that we observe in our data set.

Specifically, to model the decision to enroll in senior high school, having completed junior high, we use a latent index model which we estimate on the sample of junior high school graduates. Denoting with  $S$  the enrollment decision ( $S = 1$  if the individual decides to attend and  $S = 0$  otherwise) we have:

$$S = 1 \Leftrightarrow S^* = \alpha + \sum_{z=2}^3 \beta_z * \rho_z + \sum_{z=1}^3 \gamma_z * \text{Var}(\log Y_z) + \sum_{z=1}^3 \delta_z * P_z^W + X'\theta + U > 0. \quad (2)$$

where  $z = 1, 2, 3$  denotes junior high school, senior high school and college, respectively. The vector  $X$  contains a number of control variables that are likely to affect the schooling decision, ranging from measures of individual ability to parental background and state fixed effects and  $\theta$  denotes the vector of corresponding parameters. In terms of subjective expectations measures  $\rho_z$  denotes the expected (gross) returns to senior high school ( $z = 2$ ) and college ( $z = 3$ ) defined as  $\rho_z = E(\log Y_z) - E(\log Y_{z-1})$ , where  $E(\log Y_z)$  is the expected value of the distribution of (log) earnings at age 25 for the scenario that degree  $z$  ( $z = 1, 2, 3$ ) is the highest completed by the youth. We expect the return to senior high school relative to junior high school, as perceived by the individual, to have a positive effect on the decision to attend senior high school.  $\text{Var}(\log Y_z)$  denotes the variances of future earnings under different schooling scenarios and captures the possibility that the riskiness of a given investment might affect schooling decisions. Finally,  $P_z^W$  is the subjective probability of employment under different scenarios, which we consider as an additional risk factor because the questions on future expected earnings are conditional on working.

Notice that one would expect a high perceived earnings risk with a junior high school degree to have a positive effect on the probability of continuing to senior high school, and a high variance of log earnings with a senior high school degree to have a negative effect. On the other hand, for the decision to continue to senior high school, a high variance of log earnings after college increases the option value of continuing to senior high school. By enrolling in senior high school, one can wait for additional information while still having the option to go to college. The specification in Eq. (2) is flexible enough to be able to capture these aspects.

In a complete model, individual decisions would be determined by the entire distribution of all future returns. Our choice to consider only two moments of the various distributions (the mean and the variance) is driven mainly by the available data. First, although the data we use is unusually rich, we only have information on the expectations under different scenarios at age 25. Second, for each scenario, in addition to the probability of employment, we only observe three points of the CdF of subjective expectations: the min, the max and the probability that actual earnings are above the mid-point. With an assumption about the functional form of the CdF, one can then derive all the moments of the distribution, but this practice would be quite arbitrary. Of course, we



could have entered directly the three variables that we observe, but thought that reporting the results in terms of mean and variances would be an intuitive and useful way to present the results.<sup>9</sup>

In Attanasio and Kaufmann (2008), we notice that the mean of the subjective probability distribution is not too sensitive to the specific assumptions on the distribution, while the variance changes substantially. However, our robustness analysis shows that our main results are unaffected by the choice of specific distributions, in terms of the significance or sign of the risk variables.

Although the subjective expectations should capture individual perceptions of ability, in the specifications that we have estimated, we also control for other proxies of ability, such as GPA and parental education, that are often used in the literature to capture differences in the ability to benefit from high school or college through higher expected returns (see, e.g. Cameron and Heckman, 1998, 2001) as well as differences in psychological costs/benefits of attending college. In our specification, these variables can proxy for preferences for education and for the probability of completing senior high school or college (conditional on enrollment).

To model the decision to enroll in college – taken by youths who have just completed senior high school – we use an equation analogous to Eq. (2). The only modification we need to make is that we consider only the distributions of earnings under the two relevant scenarios for a youth who has just graduated from senior high school: to stay with the degree that she received already or to get a college degree ( $z = 2,3$ ). We therefore get:

$$S = 1 \Leftrightarrow S^* = \alpha + \beta_3 * \rho_3 + \sum_{z=2}^3 \gamma_z * \text{Var}(\log Y_z) + \sum_{z=2}^3 \delta_z * P_z^W + X'\theta + V > 0. \quad (3)$$

As in the case of the high school enrollment decision, we include measures of expected returns and perceived risks, while controlling for individual and family background characteristics and for state fixed effects.

Finally, to examine the question of gender differences in the determinants of schooling, we completely interact each regressor with a male and a female dummy to test for differential effects.

#### 4.2. Whose expectations matter for schooling decisions, parents' or youths'?

Whose expectations matter for education choices is likely to depend on the age and potentially also on the gender of the child/youth. One is therefore likely to obtain different results when modeling the decision to attend primary, secondary and higher education. Especially for youths who have finished high school and are deciding whether to enroll in college, the assumption that all decisions are taken by parents might be too strong.

Of course, if youths' and parents' expectations were fully rational and based on the same information, they would coincide. However, if either the information set or the way that it is processed differs, the subjective expectations of the different actors might differ and in the previous section, we have shown that these expectations are systematically different.

If one had data on both parents' and youths' expectations one could address these issues estimating an equation similar to Eqs. (2) or (3) but including the expectations of mothers, fathers and youths and hence establish whose expectations inform enrollment decisions. Neglecting the

variance terms, the model we would want to estimate for both school attendance decisions is as follows:

$$S = 1 \Leftrightarrow S^* = \alpha + X'\beta + \gamma_p * \text{Parents' expectations} + \gamma_Y * \text{Youths' expectations} + W > 0. \quad (4)$$

Clearly the set of parameters  $\gamma_p$  and  $\gamma_Y$  are separately identified only if the two sets of expectations differ.

Unfortunately, we have two problems in implementing Eq. (4). First, we do not have information on fathers' expectations. We discuss and interpret our results in the light of this drawback and present evidence for around 20% of households in which there is no father present, because the mother is single, separated, divorced or widowed. Obviously, these types of households are different from households with the father present, so we cannot (and do not want to) generalize our findings to the general role of mothers' and youths' expectations. Nevertheless it is interesting to see results for families without father present, since in those cases we do not need data on fathers' expectations.

Second, even focusing only on the role of mothers' vs youths' expectations, as discussed above, we never observed both of the expectations for the same individual. Therefore, even neglecting fathers' expectations, we cannot estimate Eq. (4) as such. Instead, given the identification assumption that we used to test for differences in mothers' and youths' expectations, we can estimate two versions of Eq. (4), one where  $\gamma_p$  is constrained to zero and one where  $\gamma_Y$  is constrained to zero and check which fits the data better. While this is not ideal, it offers suggestive evidence about whose expectations are relevant for enrollment decisions at different ages and for boys and girls. We discuss this issue further below.

## 5. Schooling decisions and returns to schooling

In this section, we present our main results. First, we analyze whether expected monetary returns and perceived risks (of unemployment and earnings) are important determinants of schooling decisions. Second, we analyze whether youths' own and/or mothers' expectations predict schooling decisions and how this depends on the age and gender of the youth. In the last part of the section, we present additional robustness checks for our main analysis.

We report the results that we obtain modeling the schooling decisions of the two cohorts that we consider: youths who have completed junior high school and decide whether to enroll in senior high school, and youths who decide whether to enroll in college, having completed senior high school. For each of these two cohorts, we estimate the schooling decisions jointly for boys and girls. However, to study gender differences, we interact all regressors with gender dummies (both in the main equation and in the selection equation). In the tables, we present the estimation results in two separate columns.

### 5.1. Mother vs youth decision

Having shown that mothers' expectations are significantly different from those of their children, we can sensibly address the question discussed in Section 2 about whose expectations predict education choices. With data on mothers' and youths' expectations, one could model schooling choices as potentially depending on both variables and check which, empirically, affect the enrollment decision.

Unfortunately, as we discussed above, this strategy is not feasible because mothers' and youths' expectations are observed in different subsamples. Instead, we run separately two regressions, one with the youths' expectations and one with the mothers' expectations on the two different samples. We take into account that the two subsamples were not randomly selected (compare Section 3) using the same selection model that we discussed above, and correct our estimates using a Heckman-type selection equation. We use once again the exclusion

<sup>9</sup> One referee suggested including the average of minimum and maximum earnings as a measure of location of future expected earnings and the range as a measure of uncertainty. One could make an argument for following such an approach, if one had evidence that the answers to the probability questions are of lower quality than the answers to the min and max questions. We have tried such a specification (and several others along the same lines) and the results are reported in Tables 22 to 29 in the Online Appendix. Relative to our preferred specification used in this paper, results are very similar.

restriction that the timing of the interview does not affect education choices and is not correlated with individuals' expectations, while it determines significantly to which of the two samples an individual observation belongs.

Our estimates, however, are not necessarily related to Eq. (4). Suppose that only the mothers' expectations matter for education decisions. In this case, the results from the sample where mothers' expectations are available would provide a consistent estimate of the effect of these expectations on school enrollment. Instead, the results from the sample where only the youths' expectations are available would provide inconsistent estimates, that is the coefficient on youth expectations might be zero or, to the extent that youths' expectations are somewhat correlated with the mothers' expectations, could attract a significant coefficient, albeit probably smaller in size, due to attenuation bias. On the other extreme, if only the youths' expectations matter, we would obtain consistent estimates only from the sample with youth respondents and inconsistent estimates from the sample with mother respondents. If both expectations matter, one would obtain biased estimates in both samples, except in very special circumstances.

Thus what can we learn from our results? If we found that both mothers' and youths' expectations are significant, this would imply that either both pairs of expectations matter for the decision or that their expectations are correlated. While we would not be able to separately identify these two explanations, learning whether youths' and/or mothers' expectations matter is interesting in itself. Furthermore, if we were to find that only the coefficient on mothers' expectations is significantly different from zero and significantly larger than the coefficient for the youths, this evidence would suggest that only mothers' expectations matter for the schooling decision (and vice versa).

#### 5.1.1. College attendance choice

About 29% of male high school graduates enroll in college compared to 31% of girls (see table on summary statistics in the Online Appendix). We model the college attendance decision using Eq. (3), which relates the probability of enrollment to a set of control variables and expected (gross) returns to college, perceived unemployment and earnings risk under the two different schooling scenarios.

In Table 4, we present estimates for boys and girls in separate columns. We report our estimates separately for youth respondents (Columns 1 and 2) and mother respondents (Columns 3 and 4) taking into account the selected nature of the different samples used (also all regressors in the selection equation are interacted with gender dummies). For the sake of brevity, Table 4 only displays the coefficient on the main variables of interest, that is expected labor market outcomes, while we present a table including all significant controls in the Online Appendix. As controls we use all individual and family background characteristics discussed in Section 1 and state fixed effects. We briefly discuss our findings in terms of these characteristics before moving on to a discussion of the main results.

In terms of individual characteristics, academic performance, as measured by the GPA, is an important determinant of the decision to attend senior high school for both boys and girls. Past academic performance is both a measure of the psychological costs or benefits of getting further education and also captures the likelihood of being able to complete senior high school (compare Stinebrickner and Stinebrickner, 2012). Boys are less likely to enroll in college the older they are, most likely because age is proxy for whether they had to repeat past grades. As expected, also parental education is an important determinant of the probability of going to senior high school.

Interestingly, higher parental income appears to be more important for girls than boys in the decision to attend college. Having a father who is a family worker or self-employed decreases the likelihood of boys to enroll in college, while the likelihood is increasing in having a mother in the labor force. Cost variables, such as distance from a university and the level of tuition fees in nearby universities are also important

**Table 4**  
College attendance choice: youth and mother expectations.

Dependent variable	College attendance decision			
	Youth respondent		Mother respondent	
	Boys	Girls	Boys	Girls
<i>Pooled data</i>				
Expected return – College	0.384** (0.168)	0.185 (0.132)	–0.002 (0.143)	0.198 (0.143)
Prob of work – Sr HS	–0.501 (0.452)	0.274 (0.355)	–0.437 (0.368)	0.101 (0.352)
Prob of work – College	0.108 (0.480)	0.052 (0.416)	0.194 (0.432)	1.094*** (0.402)
Var of log earn – Sr HS	–5.041 (9.147)	–8.245 (7.177)	13.837 (8.965)	5.594 (8.101)
Var of log earn – College	–7.377 (12.521)	7.773 (8.484)	–18.657* (10.830)	13.389 (11.729)
<i>By presence of father</i>				
Expected return – college *	1.318*** (0.409)	0.217 (0.318)	–0.389 (0.425)	0.444 (0.372)
No father in household	–0.000 (0.918)	0.945 (0.951)	–1.219 (1.047)	0.848 (0.842)
Prob of work – college *	–0.502 (1.100)	–0.683 (1.095)	0.401 (1.271)	2.280** (1.066)
No father in household	0.404 (24.533)	–36.804 (27.107)	–2.860 (22.830)	14.687 (20.441)
Var of log earn – Sr HS *	–17.356 (35.725)	–1.807 (17.793)	19.396 (36.115)	8.706 (24.441)
No father in household	0.274 (0.196)	0.071 (0.157)	–0.022 (0.157)	0.015 (0.158)
Expected return – college *	–0.819 (0.551)	–0.000 (0.398)	–0.543 (0.397)	0.028 (0.405)
Father in household	0.075 (0.507)	–0.203 (0.389)	0.450 (0.393)	0.280 (0.396)
Prob of work – college *	–8.182 (10.536)	–8.916 (8.183)	14.041 (10.153)	–0.453 (8.862)
Father in household	–4.958 (13.816)	8.184 (10.720)	–22.514* (12.521)	10.707 (13.294)
Var of log earn – college *	2829		2829	
Father in household	1468		1361	
Observations				
Censored obs				

Notes: Table displays coefficients and standard errors in brackets. Included controls are all variables discussed in Section 3.1 and state fixed effects (fully interacted with gender dummies).

\* p < 0.1.

\*\* p < 0.05.

\*\*\* p < 0.01.

determinants of the decision, even though the level of tuition fees in the nearby university seems to affect only girls.

The key results of Table 4 (upper panel) are three: First, boys' own expectations about returns to college are a strongly significant predictor for their decision to enroll in college. For girls on the other hand, the coefficient on their own expectations is not significantly different from zero. This result is not driven by a larger standard error for the girls' coefficient, i.e. it is not the case that the impact of returns in girls' choices is estimated less precisely.<sup>10</sup> Instead the coefficient for girls is less than half the size of the boys' coefficient (though we cannot reject that they are the same, since standard errors are relatively large).

Second, mothers' expectations are significant predictors in girls' decisions to enroll in college, while girls' own expectations are not significant. In particular, if the mother perceives a high likelihood of her daughter finding a job with a college degree, this significantly increases the probability of her daughter attending college. For boys on the other hand, only their own expectations about returns to college predict their schooling decisions.<sup>11</sup>

Third, our data enable us to look at two potential determinants of schooling decisions which have often been neglected in this analysis,

<sup>10</sup> Also both boys and girls expect high returns to schooling.

<sup>11</sup> While the gender difference between the role of mothers' perceived probability of working with a college degree is not significant on conventional levels with a p-value of 0.12, the coefficient for girls is about five times larger than the one for boys.

that is the perceptions of unemployment and earnings risk for different schooling levels. Since we found that people perceive lower unemployment and lower earnings risk for higher schooling degrees (e.g. having a college degree reduces labor market risk compared to a high school degree, see Tables 2 and 3), risk perceptions could be an important additional determinant of schooling decisions in addition to expected returns. Data on people's subjective expectations allow us to derive a measure of "true" risk, as perceived by the individual, while if one were to use the variability of observed earnings this would confound risk with unobserved heterogeneity. Moreover, observed variability will not distinguish between predicted and unpredicted "shocks".

We find that perceptions about unemployment risk are an important determinant for college attendance decisions from the perspective of the mother. In particular, perceiving a high probability of employment with a college degree increases the probability of girls to attend college. Also the perception of earnings risk is significant in the case of a college degree, again from the perspective of the mother, this time for boys' likelihood to attend college. As predicted, perceiving a higher earnings risk with college degree leads to a lower probability to attend college. It is interesting to see that mothers' risk perceptions predict schooling decisions, but not the ones of boys (or girls, though for them none of the coefficients on expectations are significant). This finding is consistent with the literature on gender differences in risk aversion (with a higher level of risk aversion for women, see, e.g. Borghans et al. (2009) and Schubert et al. (1999)). At the same time, we do not want to make stronger claims about what is driving this result, since in this case it would be more appropriate to compare mothers' risk perceptions with fathers' risk perceptions, which we unfortunately do not observe.

While it is interesting to see that boys' own expectations are correlated with schooling decisions, and girls' expectations are not, it would be interesting to analyze whether boys' expectations are significant because they are correlated with the ones of fathers'. Unfortunately, as discussed before, we do not observe fathers' expectations. The exercise we conduct instead is to see whether boys' own expectations matter for schooling decisions in the special case of families without a father present (where there is obviously no need to have data on fathers' expectations). Of course, we do not claim that our results can be generalized to the whole population, since these families are very different in terms of their characteristics and very likely also in terms of how decisions are made.

We repeat the analysis above to compare our results to the results for the subsample of households headed by single, separated, divorced or widowed women (which is the case for about 16% of senior high school graduates). We report the coefficient estimates on the main variables of interest in the lower panel of Table 4.

Also for the subsample of families without a father, we find that boys' own expectations are strongly significant predictors of their college attendance decision (significant at 1%). Mothers' expectations are insignificant in the decision of boys and the coefficient on expected returns is basically zero. For girls on the other hand, only mothers' expectations matter. In particular, the coefficient on mothers' perceptions of the probability of work with college is significant (and significantly larger than for girls). Second, for these types of families the coefficient on expected returns is significantly larger for boys than for girls (at 2%).

Although we control for potential self-selection problems because of the non-randomness of who responds to the expectations questions, it is worthwhile to point out that self-selection does not appear to be a problem in the case of the old cohort (we never reject that the correlation between the error terms of selection equation and main equation is zero).<sup>12</sup>

To summarize, our evidence shows that in the context of college enrollment decisions, there are important gender differences in the

determinants of school choices. In particular, boys' own expectations predict their schooling decisions, while girls' own expectations do not. Mothers' expectations on the other hand are predictors of their daughters' educational choices, but not of the ones of their sons. Interestingly, even in households without a father present, boys' own expectations matter for schooling decision but not the ones of mothers', thus providing suggestive evidence that boys are involved in the decision to attend college in these types of families.

### 5.1.2. High school attendance choice

We now turn our attention to the decision to enroll or not in senior high school for youths who have just finished junior high school. Around 78% of boys and 81% of girls graduating from junior high school decide to enroll in senior high school (see the Online Appendix). In Table 5, we report our estimates separately for youth respondents (Columns 1 and 2) and mother respondents (Columns 3 and 4), taking into account the selected nature of the sample (again fully interacting all regressors with gender dummies). Again we present estimates for boys and girls in separate columns, although they are estimated simultaneously.

As in the case of the college attendance decision, Table 5 only displays the coefficients on the main variables of interest, that is expected labor market outcomes, while we present a table including all significant controls in the Online Appendix. As controls we use all the individual and family background characteristics discussed in Section 1 and state fixed effects. We briefly discuss our findings in terms of these characteristics before moving on to a discussion of the main results.

Individual and family background variables have the expected signs: the probability of senior high school attendance increases with GPA and with parental education. A larger number of siblings decrease the likelihood of continuing schooling, in particular the presence of brothers. The negative effect is stronger on boys than girls. One potential explanation might be that even though girls might be forced to occasionally skip school to look after her siblings, boys might be forced to work and leave school completely to contribute to the families' household income. Being from a family in the lowest per capita parental income category decreases the likelihood of enrolling in high school (significant only for girls). Having a father who is a family worker decreases the likelihood for girls to attend high school and having a father who is self-employed increases the likelihood for boys.

The key results of Table 5 (upper panel) are: First, the expectations of the 15-year-old youths do not predict the decision to enroll in high school. Mothers' expectations on the other hand are a predictor of girls' decision to enroll in high school, but not of the decision of boys. In particular, if (in the case of girls) a mother perceives a higher probability that her daughter will be employed with a junior high school degree, this lowers the probability of the daughter to enroll in senior high school. If the mother on the other hand perceives a high probability of employment with a college degree, this increases the likelihood of enrollment in senior high school, as one would expect. The latter result suggests that having the option to enroll in college with a high school degree can be important for the high school enrollment decision. Lastly, mothers' perceptions about earnings risk with a senior high school degree matter for girls, with the expected negative sign. For boys on the other hand, mothers' expectations are not significant.<sup>13</sup> Second, our results show that risk perceptions (of unemployment and earnings) can be very important in schooling decisions. Again, as in the case of the college attendance decision, they matter mostly from the perspective of the mother.

When we focus our attention to the subsample of families without a father, we allow the coefficients on the expectations to be different in this subsample, and report the results in the lower panel of Table 5. In

<sup>12</sup> Results remain basically the same when running the regressions without selection correction (see the Online Appendix).

<sup>13</sup> Differences between the coefficients of boys and girls are significant in the case of the perceived probability of work with a junior high school degree (at 5%) and close to significant in the case of the perceived earnings risk with a junior high school degree (p-value 0.11).

**Table 5**  
High school attendance choice: youth and mother expectations.

Dependent variable	High school attendance decision			
	Youth respondent		Mother respondent	
	Boys	Girls	Boys	Girls
<i>Pooled data</i>				
Expected return – Sr HS	–0.059 (0.201)	0.141 (0.188)	0.175 (0.238)	–0.053 (0.268)
Expected return – College	0.208 (0.167)	0.236 (0.173)	–0.001 (0.234)	–0.223 (0.232)
Prob of work – Jr HS	0.089 (0.371)	0.036 (0.326)	0.498 (0.428)	–0.803* (0.456)
Prob of work – Sr HS	–0.060 (0.494)	–0.210 (0.434)	–0.878 (0.620)	–0.156 (0.603)
Prob of work – College	0.650 (0.398)	–0.157 (0.450)	0.886 (0.545)	1.107** (0.525)
Var of log earn – Jr HS	4.431 (6.374)	9.887 (7.936)	11.426 (8.022)	–6.476 (7.990)
Var of log earn – Sr HS	11.812 (9.310)	–10.159 (8.054)	–7.449 (10.717)	–17.492** (8.677)
Var of log earn – college HS	5.934 (11.026)	7.737 (10.736)	–11.015 (12.238)	8.595 (10.282)
<i>By presence of father</i>				
Expected return – Sr HS *	0.500 (0.496)	–0.111 (0.461)	0.131 (0.643)	–0.384 (0.600)
Expected return – College *	0.907** (0.444)	0.280 (0.399)	0.371 (0.697)	–0.011 (0.511)
Prob of work – Jr HS *	–0.827 (0.796)	–0.534 (0.762)	–1.272 (1.279)	–0.026 (1.064)
Prob of work – Sr HS *	1.641* (0.930)	1.311 (1.036)	–2.095 (2.137)	–2.624* (1.418)
Prob of work – college *	0.312 (0.910)	–2.320* (1.348)	0.985 (1.870)	2.422** (1.174)
Var of log earn – Jr HS *	–0.791 (21.235)	14.693 (21.792)	15.255 (20.122)	–19.453 (19.893)
Var of log earn – Sr HS *	–1.201 (27.531)	15.146 (31.630)	36.600 (47.002)	–5.700 (31.236)
Var of log earn – college *	92.069* (48.860)	–4.134 (24.867)	24.736 (40.142)	84.584 (61.050)
Expected return – Sr HS *	–0.348 (0.241)	0.092 (0.216)	–0.009 (0.298)	–0.165 (0.334)
Expected return – college *	0.023 (0.192)	0.197 (0.210)	–0.200 (0.284)	–0.461 (0.287)
Prob of work – Jr HS *	0.499 (0.481)	0.070 (0.404)	1.138** (0.540)	–1.351** (0.617)
Prob of work – Sr HS *	–1.038 (0.670)	0.002 (0.555)	–1.235 (0.758)	0.484 (0.873)
Prob of work – college *	0.749 (0.481)	–0.244 (0.475)	0.970 (0.609)	0.194 (0.714)
Var of log earn – Jr HS *	3.794 (8.004)	12.040 (9.889)	20.103* (11.552)	–3.968 (10.676)
Var of log earn – Sr HS *	11.935 (11.128)	–9.817 (9.492)	–11.611 (12.769)	–22.208** (11.103)
Var of log earn – College *	–2.851 (11.605)	9.979 (13.777)	–24.522 (17.163)	–2.572 (13.357)
Observations	2364		2364	
Censored obs	967		1397	

Notes: Table displays coefficients and standard errors in brackets. Included controls are all variables discussed in Section 3.1 and state fixed effects (fully interacted with gender dummies).

\*  $p < 0.1$ .

\*\*  $p < 0.05$ .

\*\*\*  $p < 0.01$ .

families with a father present, youths' own expectations are not predictive of their educational decisions consistent with what we found above. Mothers' expectations predict both boys' and girls' decision to enroll in high school.

In families without a father present, boys' own expectations are significant for their schooling decision, while mothers' are not. The likelihood of enrolling in high school increases with boys' expected returns to college (significant at 5%), with the probability of working with a

college degree and with a higher perceived variance of college earnings (which is consistent with a higher variance increasing the option value). For girls on the other hand, mothers' expectations are once again significant (girls' own expectations are significant in two instances, once with the expected sign on the perceived earnings risk with a junior high school degree and once with a – surprisingly – positive coefficient on the probability of working with a junior high school degree).

Thus in the case of junior high school graduates (as for the older cohort), in families without a father, boys' own expectations predict their educational choices, while mothers' expectations do not. In families with a father on the other hand, the expectations of the 15-year-old youths appear not to matter at all for the decision to enroll in high school.

## 5.2. Robustness: Subjective expectations and endogeneity

In this section we discuss in detail concerns that are sometimes raised with respect to data on subjective expectations. In particular, we discuss potential problems of endogeneity due to omitted variables and due to reversed causality (ex-post rationalization).

### 5.2.1. Ex-post rationalization (or reversed causality)

The term “ex-post rationalization” describes the behavior of people who state beliefs to justify their choices (ex-post), that is, the decision affects the beliefs instead of the beliefs affecting the decision. In the context of this paper, one might be worried that people decide to go to college for reasons other than expected monetary returns, and that they justify this decision by stating high expected returns.

To test the possibility that our respondents answer the expected return questions to justify their choices, as the survey on expectations was conducted two–three months after the schooling decision, we conduct the test that we discussed in Section 2. In particular, we compare the cross-section of expected earnings for our cohorts of interest (i.e. students who had just graduated from junior and senior high school before the survey) and cohorts that are one year younger (and thus just starting grade 9 or 12). Let us first concentrate on the college enrollment decision of senior high school graduates. If individuals rationalize their choices, we would expect that individuals who decided to enroll in college would state higher expected college earnings or lower high school earnings. On the other hand, those individuals who decided not to enroll, would state lower college earnings (or higher high school earnings). This would lead to a cross-sectional distribution of expected earnings and returns that is more spread out for the cohort of senior high school graduates who have decided already, compared to the distribution of the cohort that is one year younger and just starting grade 12.<sup>14</sup> The same reasoning holds for perceptions of the probability of working.

To test for differences between the distribution of expectations of those two adjacent cohorts, we use a Kolmogorov–Smirnov test, which is a nonparametric test for the equality of continuous, one-dimensional probability distributions. The null hypothesis is that the cross-sectional distribution of – for example – expected returns is the same for the sample of junior (senior) high school graduates and the sample of a cohort that is one year younger and just starting grade 9 (12). To correct for potential differences in the composition between the two cohorts, we compare the residuals from a regression of the expectations on individual and family background variables (see variables in Tables 1 and 2 and state fixed effects).

<sup>14</sup> This is true, unless people switch positions in the distribution in such a way that the resulting cross-section looks exactly the same as before. This could only happen if people with particularly low expected college earnings are the ones who decide to attend college and later state high college earnings to rationalize their attendance decision, and those with high expected college earnings are the ones who decide not to attend and later state low expected college earnings.

In the Online Appendix we report p-values of the Kolmogorov–Smirnov tests for senior high school graduates and junior high school graduates, respectively. We conduct the test for youth and mother respondents separately and separately by gender for the different measures of subjective expectations, i.e. for expected earnings for different schooling scenarios, expected returns (which is the relevant variable that we use in the main regression analysis instead of expected earnings) and perceived risks.

First we discuss the results for senior high school graduates. In the case of youth respondents, we do not reject equality of distributions apart from the following instances, that is, girls' perceptions about the probability of working with a senior high school and a college degree, girls' expectations about log college earnings and boys' perceptions about earnings risk with a college degree. In neither of the cases is the distribution of the older cohort more spread out (as one would expect in the case of ex-post rationalization), but some of the older individuals seem to have updated their beliefs about probability of working upward compared to the one-year younger cohort (see figures in the Online Appendix). At the same time, this is not driving our results because in any case girls' own expectations are not significant in the college attendance choice regression and also boys' perceptions about earnings risk are not significant.

For mother respondents we reject equality of distributions of expected return for both boys and girls on 1%. Again the distribution is not more spread out, but some mothers seem to have updated their beliefs about expected returns upward (see figures in the Online Appendix). In principle this could mean that some mothers rationalize their children's choice to go to college by stating higher expected returns to college. But at the same time, mothers' expectations about returns are never significant in our regressions of college attendance, neither for girls nor for boys. In addition we find differences in the case of girls' perceived probability of work with a senior high school degree and in the case of perceptions of earnings risk. Also this is not driving our results, since the coefficients on mothers' risk perceptions of earnings with a senior high school degree are not significant in our main regressions, nor is the perception of the probability of working with a senior high school degree.

In the following we discuss the results of the Kolmogorov–Smirnov tests for the cohort of junior high school graduates. For youth respondents, we can reject equality of distributions in two instances: for girls in terms of expected log earnings with a junior high school degree and for boys in terms of the probability of working with a college degree. In the former case we are not directly interested in the effect of log earnings, but in the effect of return expectations, which are not significantly different. Thus the finding on expected returns having an effect is not threatened by the concern of rationalization. Boys' perceptions about the probability of working do not have a significant effect in our main regressions.

For mother respondents, equality of distributions can be rejected for expected high school and college returns for girls (see the Online Appendix). Again neither of the two measures of expected returns is significant in any of the regressions.

Also the distributions of perceived probability of working shifted for both boys and girls to a very similar extent (not in terms of larger spread, but some mothers have revised their beliefs upward, see figures in the Online Appendix). This cannot explain why mothers' beliefs about the probability of work (with a junior high school degree and a college degree) are significant in explaining high school choices of girls, but not for boys.

In addition to the above analysis, we should also mention the following point: If the results we have shown were driven by ex-post rationalization, one would have to explain why the relationship is significant for boys but not for girl when youths' respond, and vice versa when mothers respond. This would imply that boys engage in ex-post rationalization but not girls, and mothers engage in ex-post rationalization but only in the case of their daughters. If what we observe is due to

ex-post rationalization, it must be the case that there exists the idea (or norm/standard) in the population that the (main acceptable) reason for going to school is to get higher earnings. If this norm existed for boys, for example, then it is hard to explain why boys ex-post justify their decision evoking this norm, but mothers do not justify their sons' decisions in this way (and vice versa for girls).

### 5.2.2. Classical endogeneity

Another concern with using subjective expectations as determinants of schooling choices is another form of endogeneity: the one that would arise from unobserved taste heterogeneity (or omitted variables). Individuals who have high expected returns and go to college more often might also be the ones who attach weight to other outcomes, which are not measured, to which college is conducive.

We have to ask ourselves what could be other outcomes that are as important as or more important than expected monetary returns and correlated with the latter. One possibility is returns in the marriage market.

Can this phenomenon explain our main results? One result is that expected returns to college significantly increase the likelihood to go to college for boys, but not for girls. To explain a positive coefficient on expected labor market returns for boys with an “omitted variable bias” story, one would have to argue that there is a positive correlation between marriage market and labor market returns for boys and boys care mostly about returns in the marriage market instead of labor market returns. This seems hard to believe and runs contrary to the vast majority of papers on educational choices.

Even if one believes that the coefficient on labor market returns could pick up effects of marriage market returns for boys, we are less interested in the magnitude of the coefficient and more in the differences between boys and girls and the differences between (male/female) youth respondents and mother respondents. Thus, if it is true that labor and marriage market returns are positively correlated and therefore the coefficient on monetary returns picks up marriage market effects, then it is hard to explain, why the coefficient on expected monetary returns is positive for boys but zero for girls. One explanation could be that girls neither care about their marriage market outcomes nor about their labor market outcomes, but again it seems hard to believe that girls (in particular from poor families) would decide to go to college based on something entirely different than labor or marriage market returns.

Can the differential effect between mother and youth respondents be explained by endogeneity? In particular, we find a positive and significant effect of boys' expectations on their likelihood to go to college, but no effect of mothers' expectations for their sons. Also we find positive coefficients on mothers' expectations for girls' decision to go to college, but no effects for girls' own expectations. It seems difficult to come up with a story about omitted variables that can explain these results.

## 6. Conclusion

The results of this paper speak to several important questions about the determinants of investment in human capital among poor households in Mexico. In particular, we have analyzed how expectations of future labor market outcomes affect schooling decisions of the two cohorts of Mexican children, one cohort that has just completed junior high school and one that has just completed senior high school. In doing so we consider not only the expected monetary returns (as expressed by future earnings under different schooling alternatives) but also the risks involved with such choices.

Our data on subjective expectations also show that there are significant and important differences in the expectations held by youths and their mothers on the return to education for the same youth. This implies that it is interesting to understand whose expectations matter and how this depends on age and gender of the youth. Direct data on people's beliefs enable us to address this question, while being agnostic

about how exactly expectations differ (depending on age and gender etc.).

Addressing these questions allows us to get a more in depth understanding of educational decisions and suggests a way to shed further light on intrahousehold decision processes by making use of data on mothers', fathers' and youths' own expectations about future schooling-related outcomes of the youth (data, which – to the best of our knowledge – are currently not available).

Our main result is that the schooling decision is quite different for boys and girls in several dimensions. Our results indicate that labor market expectations, such as expected monetary returns, are important determinants of educational choices, but appear particularly important for boys.

We provide evidence that not only expected (monetary) returns matter for educational decisions, but also risk perceptions, which are important from the perspective of mothers. This result is important as these considerations have been neglected in the literature, partly for the lack of appropriate data. We show how data on subjective expectations allow us to address these issues directly.

Lastly, our paper adds to the literature on subjective expectations in illustrating that – also in developing countries, at least conditional on a certain level of education – people seem able and willing to respond meaningfully to questions about their perceptions of future earnings and employment and that these data can improve our understanding of important economic decisions, such as investment into human capital.

### Acknowledgments

We would like to thank the editor and the two anonymous referees for their valuable comments. Orazio Attanasio acknowledges his European Research Council Advanced Grant 249612 on "Exiting Long Run Poverty: The determinants of asset accumulation in developing countries". We are grateful to Sandra Black, Janet Currie, Pascaline Dupas, Caroline Hoxby, Lance Lochner, Shaun McRae, Matthias Messner, Karthik Muralidharan, Kevin Stange and Alessandro Tarozzi for their insightful discussions. We are thankful also to the conference participants at the NBER Education meeting in Spring 2010, at the CEPR Labor Economics Program on the Economics Of Family Policy and the CEPR development meeting in Spring 2010, at the NYU/Caltech La Pietra conference, at the BREAD summer workshop, at the SITE meeting on Women and the Economy and at the CESifo conference on the Economics of Education and to seminar participants at Boston Fed, New York Fed, Rochester and UVA. All remaining errors are of course our own.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.jdeveco.2014.04.003>.

### References

Abramitzky, R., Lavy, V., 2011. How responsive is investment in schooling to changes in redistribution policies and in returns?, mimeo.

- Arcidiacono, P., Hotz, V.J., Kang, S., 2012. Modelling college major choices using elicited measures of expectations and counterfactuals. *J. Econom.* 166 (1), 3–16.
- Attanasio, O.P., 2009. Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review, Papers and Proceedings*.
- Attanasio, O.P., Kaufmann, K.M., 2008. Subjective returns to schooling and risk perceptions of future earnings – elicitation and validation of subjective distributions of future earnings, mimeo.
- Attanasio, O.P., Meghir, C., Santiago, A., 2011. Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progres, *Review of Economic Studies*.
- Berry, J., 2011. Child Control in Education Decisions: An Evaluation of Targeted Incentives to Learn in India, mimeo, MIT.
- Blundell, R., Chiappori, P.-A., Meghir, C., 2005. Collective labor supply with children. *J. Polit. Econ.* 113 (6), 1277–1306.
- Borghans, L., Golsteyn, B.H.H., Heckman, J.J., Meijers, H., 2009. Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association*, 8(2–3). MIT Press, pp. 649–658.
- Bourguignon, F., 1999. The cost of children: May the collective approach to household behavior help? *J. Popul. Econ.* 12 (4), 503–522.
- Cameron, S.V., Heckman, J.J., 1998. Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males. *J. Polit. Econ.* 106 (2), 262–333.
- Dauphin, A., Lahga, A.-R.E., Fortin, B., Lacroix, G., 2008. Are children decision-makers within the household? IZA Discussion Paper No. 3728.
- Delavande, A., Gine, X., McKenzie, D., 2011. Measuring subjective expectations in developing countries: a critical review and new evidence. *J. Dev. Econ.* 94 (2), 151–163.
- Dinkelmann, T., Martinez, C., Investing in schooling in Chile: the role of information about financial aid for higher education. *The Review of Economics and Statistics*, (forthcoming).
- Dominitz, J., Manski, C.F., 1996. Eliciting student expectations of the returns to schooling. *J. Hum. Resour.* 31 (1), 1–26.
- Giustinelli, P., 2011. Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choices of High School Tracks, Mimeo.
- Guiso, L., Jappelli, T., Pistaferri, L., 2002. An empirical analysis of earnings and employment risk. *J. Bus. Econ. Stat.* 20 (2), 241–253.
- Heckman, J.J., 1979. Sample selection as a specification error. *Econometrica* 47 (1), 153–161.
- Cameron, S.V., Heckman, J.J., 2001. The dynamics of educational attainment for Black, Hispanic and White males. *J. Polit. Econ.* 109 (3), 455–499.
- Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. *Q. J. Econ.* 125 (2), 515–548.
- Jensen, R., 2012. Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *Q. J. Econ.* 127 (2), 753–792.
- Kaufmann, K.M., 2014. Understanding the income gradient in college attendance in Mexico: the role of heterogeneity in expected returns. *Quantitative Economics*The Econometric Society (forthcoming).
- Keane, M.P., Wolpin, K.I., 1997. Career decisions of young men. *J. Polit. Econ.* 105 (June), 473–522.
- Mahajan, A., Tarozzi, A., Yoong, J., Blackburn, B., 2011. Bednets, Information and Malaria in Orissa, mimeo, Stanford University, Duke University and RAND.
- Manski, C.F., 2004. Measuring expectations. *Econometrica* 72 (5), 1329–1376.
- Nguyen, T., 2008. Information, role models and perceived returns to education: experimental evidence from Madagascar, mimeo.
- Reuben, E., Wiswall, M., Zafar, B., 2013. Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender. *Federal Reserve Bank of New York Staff Reports* p. 627 (August).
- Schubert, R., Gysler, M., Brown, M., Brachinger, H.-W., 1999. Financial decision-making: are women really more risk-averse? *Am. Econ. Rev.* 89, 381–385.
- Stinebrickner, T., Stinebrickner, R., 2012. Learning about academic ability and the college drop-out decision. *J. Labor Econ.* 30 (4), 707–748.
- Todd, P.E., Wolpin, K.I., 2006. Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *Am. Econ. Rev.* 96 (5), 1384–1417.
- Zafar, B., 2009. College Major Choice and the Gender Gap. *Federal Reserve Bank of New York Staff Reports*.