

# Education Choices and Returns to Schooling: Mothers' and Youths' Subjective Expectations and their Role by Gender.

Orazio P. Attanasio\* and Katja M. Kaufmann†

September 2013, first draft: June 2008

## 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.

JEL-Classification: I21, I22, O15

KEYWORDS: Schooling Decisions, Subjective Expectations, Perceived Earnings and Employment Risk, Gender Differences, Mexico.

---

\*University College London, Institute for Fiscal Studies, NBER, CEPR and BREAD. o.attanasio@ucl.ac.uk

†Department of Economics and IGIER, Università Bocconi, Address: Via Roentgen 1, 20136 Milan, Italy, +39.02.5836.3370, katja.kaufmann@unibocconi.it.

# 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 are important determinants of schooling decisions and whether they differ for male and female youths. In particular, 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 scenarios. Second, we 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 data set on Mexican junior and senior high school graduates that elicits youths' or mothers' beliefs about future labor market outcomes for different scenarios of highest schooling degree. In addition, we have information about the actual schooling choice and an extensive set of controls.

We know surprisingly little about the importance of gender differences in the determinants of schooling choices. To what extent do expected returns in the labour market (and other variables) play a different role in the schooling decision for boys and girls? Are there differences between boys and girls in terms of whether their expectations about returns or their perceptions about employment and earnings risk matter for their educational 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 understanding gender differences in educational attainment as well as for the effective design of programs intended to increase schooling among poor households.

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 individuals with different levels of education whose earnings realizations are observed are typically selected into that education level. We use data on subjective expectations of earnings under different scenarios, and by doing so we avoid these problems.

We address our questions in the context of two schooling decisions: that of youths who have just finished junior high school to enrol (or not) in senior high school, and that of youths who have completed senior high school to enrol (or not) in college. We construct measures of expected earnings, perceived earnings risk and perceived unemployment risk held by mothers and youths for different schooling scenarios, to study the extent to which these expectations affect education choices of male and female youths.

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. In particular, using realized earnings to infer expectations would not allow to accommodate this feature without making assumptions in what direction and by how much the expectations of youths and parents differ. Moreover, it provides us with the opportunity of testing whether mothers' and/or youths' expectations are

predictive of schooling decisions. We therefore proceed to estimate a model of schooling decisions to measure the importance of expected returns to education for boys and girls.

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' 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.

Lastly, we analyze whether our results might be driven by omitting another potentially important return to schooling, that is returns in the marriage market. Therefore, we include as a control a proxy for marriage market returns that captures the (net-) supply of (suitable) potential partners in the youths' locality of residence. The idea for using this proxy is as follows: If the likelihood of finding a (suitable) partner in the locality of residence is low, leaving the locality to go to college or high school has the additional benefit of increasing the chances of finding a (suitable) partner. We show that our results on the relevance of expected labor market outcomes remain unchanged. At the same time, our results are consistent with marriage market returns playing an important role for girls' schooling decision.

**Related literature** Our paper is related to a growing literature using data on people's subjective expectations to understand educational decisions. The seminal paper eliciting subjective expectations of earnings for different schooling degrees is by Dominitz and Manski (1996). They illustrate for a small sample of Wisconsin high school and college students that people are willing and able to answer subjective expectations questions in a meaningful way, but do not analyze the link between earnings expectations and investment into schooling.<sup>1</sup> Three closely related papers investigating the link between subjective expectations of earnings and schooling choices are Jensen (2010), Kaufmann (2009) and Nguyen (2008).<sup>2</sup> Jensen (2010) investigates how perceived returns to education affect schooling investments using data from the Dominican Republic. He finds that the students in his sample of 8th graders significantly underestimate returns to schooling. Informing a random subset of the youths about higher measured returns leads to a significant increase in perceived returns and in attained years of schooling among these students. Nguyen (2008) finds that informing a random subset of students in Madagascar about high returns to schooling increases their attendance rates and their test scores. Kaufmann (2009) uses the same Mexican data set as this paper to analyze the causes and consequences of the large income gradient in college attendance in Mexico. In particular, data on subjective individual expectations enable her to analyze to what extent the gradient can be explained by the poor expecting lower returns to college than the rich,

---

<sup>1</sup>See also the survey paper on the use of subjective expectations data by Manski (2004). Delavande, Giné, and McKenzie (2011a) 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). There is, of course, a large literature on the (ex-post) effects of schooling on earnings and other outcomes, see for example Card (2001).

<sup>2</sup>Another related paper is by Stinebrickner and Stinebrickner (2012) who analyze how college students from low income families form expectations about academic ability. Their results show that learning about ability plays a very prominent role in the college drop-out decision. Arcidiacono, Hotz, and Kang (2012) use data on subjective expectations of students at Duke University to explain college major choices.

for example due to ability and/or information differences. She finds that while differences in expected returns only explain part of the gap, poor individuals require significantly higher returns to be induced to attend college. Testing implications of a school choice model in the presence of credit constraints, she finds that expected returns of students at the margin of attending are significantly higher than average returns of people who already attend college, consistent with credit constraints for poor students. Relative to these papers, we address several complementary questions making use of unique features of the Mexican data.

One novel aspect of this paper is our focus on gender differences in the determinants of schooling. 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 Northwestern University. Reuben, Wiswall, and Zafar (August 2013) just came out with the first version of a paper in which they investigate whether behavioral biases and preferences explain gender differences in college major choices and expected future earnings.

The focus of this paper is instead on the decision to enrol in senior high school and the decision to enrol in college in a very different context (urban Mexico). More importantly, we analyze whether mothers' or youths' own expectations matter for these decisions and shed light on the role of perceptions of earnings and employment risk. Lastly, we show how the role of these determinants depends on age and gender of the youth.

Another two recent papers also show the link between schooling choices and returns to schooling –with some interest in gender differences– by making use of (quasi) experimental changes in returns to schooling. Abramitzky and Lavy (2011) make use of a policy change in Israeli kibbutzim, which increased returns to schooling. They observe that students in early-reforming kibbutzim increased their investment in education in terms of matriculation rates and scores and male youths responded particularly strongly. Jensen (2012) examines whether educational investments in girls in India respond to changes in employment opportunities. He provides recruitment services to young women in randomly selected rural villages helping them get jobs in the business process outsourcing industry and finds that this intervention increases schooling of girls in treatment villages. While the latter two papers do not use data on subjective expectations, these data enable us to also analyze if perceived earnings and unemployment risk play a role in educational decisions (and if their role differs for boys and girls).

Secondly, we provide evidence on whose expectations predict schooling decisions, the ones of the mothers and/or the youths and how this depends on gender and age. In general we know relatively little about the role of children in household decisions. Instead children have been modeled as household public goods (see, e.g., Blundell, Chiappori, and Meghir (2005)) or incorporated through the “caring preferences” of their parents (see, e.g., Bourguignon (1999)). Notable exceptions are the studies by Dauphin, Lahga, Fortin, and Lacroix (2008), Berry (2011) and Giustinelli (2011). Dauphin, Lahga, Fortin, and Lacroix (2008) tests restrictions of a theoretical model on consumption decisions and infer the minimum number of decision-makers in three-person households from parametric constraints. Using UK data, the authors of this paper provide evidence that children start playing a role in consumption decisions at age 16. Berry (2011) uses a very different approach

and analyzes whether the identity of the recipient of cash incentives –either the parent or the child– can influence the effectiveness of conditional cash transfer programs. Giustinelli (2011) analyzes high school curriculum choices of Italian youths modeling the decision based on a direct question about who is the decision-maker and data on subjective expectations.

Our paper, which is most closely related to the last paper in terms of the use of data on subjective expectations, aims to shed light on one very specific aspect of decision processes in households, that is it provides some evidence on whether youths’ own and/or mothers’ expectations can predict schooling decisions. Related to this question, Dinkelman and Martinez (2011) investigate the role of information in schooling decisions and whether the role depends on whether parents or youths receive the information.

Lastly, our robustness check on whether the role of expected labor market outcomes might be driven by omitting a (potentially) very important alternative return, the return in the marriage market, is related to a small, but growing literature on marriage market returns (see, e.g. Chiappori, Iyigun, and Weiss (2009) and Lafortune (2012) on the ex-ante role of marriage market considerations and Kaufmann, Messner, and Solis (2012) who provide evidence on important ex-post marriage market returns to college quality in Chile).

In a recent paper, Wiswall and Zafar (2011) conduct an experiment informing a random subset of NYU students about average earnings in the population and analyze how the resulting changes in expected earnings affect their major choice. In their analysis they include expectation questions on spousal earnings to explain college major choices.<sup>3</sup>

This paper instead is focusing on very different schooling decisions in a very different context (high school and college enrolment in Mexico versus college major choice at NYU) and thus the role of marriage market considerations is likely to be very different for the following reason. In the case of the college major choice, college students are likely to meet potential partners in dorms and at parties independently of their major choice. In our context instead, girls who face unfavorable marriage market conditions might find very different conditions when going to high school or college (and moving to a bigger city). Furthermore, also anecdotal evidence suggest that marriage market considerations are important in the Mexican context (see the acronym ”mmc” –mientras me caso– for girls going to college in Mexico).

**Outline** The rest of the paper is organized as follows: In Section 2 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 3 describes in detail the data, in particular the module on subjective expectations, and presents summary statistics on attendance rates and a variety of background variables. In Section 4, we briefly discuss the literature that shows that people are able to give meaningful answers to the subjective expectation questions and provide evidence on the validity of our expectations data. We also analyze whether there are differences in expectations between mothers and youths about future earnings of the youth. Section 5 discusses the main results we obtain. In this section we also address the concerns

---

<sup>3</sup>The first version of that paper came out about one year after the first version of our paper, see ?).

related to the use of subjective expectation data. Section 6 concludes.

## 2 Schooling Decisions and Returns to Schooling.

In this Section, we discuss the empirical model used to tackle the questions of this paper, that is whether youths' and/or mothers' expectations matter, whether expected returns and/or risk perceptions matter and how the answer to these questions depends on gender and age of the youth.

### 2.1 Determinants of Enrolment.

One possible approach to analyze the relationship between enrolment and expected returns would be the construction of a full dynamic optimization model where individuals choose current activities taking into account current and future benefits and 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, Todd and Wolpin (2006) and Attanasio, Meghir, and Santiago (2011)). Kaufmann (2009) shows how data on people's subjective expectations can be used in a simple model of college enrolment choice. In this paper, rather than following this route, we present probit regressions that relate the probability of enrolment 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 labour market conditions through state dummies and for family background and ability through several variables we observe in our data set.

Specifically, to model the decision to enrol 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 enrolment decision ( $S = 1$  if the individual decides to attend and  $S = 0$  otherwise) we have:

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

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 expectation 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.  $Var(\log Y_z)$  denotes the variances of future earnings under different schooling scenarios and  $P_z^W$  the subjective probability of employment under different scenarios, as discussed further below.

We obviously 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. In Equation (1), we also include the expected return to college to capture the fact that attending (and completing) high school provides the option to attend college later on.

In addition to the expected return, we also want our empirical model to take into account the possibility that the riskiness of a given investment might affect schooling decisions. For this reason, we enter, as determinants of the schooling decision, the variances of the future earnings under different schooling scenarios  $Var(\log Y_z)$ . Moreover, as the questions on future expected earnings are conditional on working, we enter the subjective probability of employment under different scenarios,  $P_z^W$ .

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 Equation (1) is flexible enough to be able to capture these aspects.

Usually, proxies for ability, such as GPA and parental education, are supposed to capture differences in the ability to benefit from high school or college through higher expected returns (see, e.g. Cameron and Heckman (1998) and Cameron and Heckman (2001)) as well as differences in psychological costs/benefits of attending college. However, such variables are only imperfect proxies of the returns that an individual can obtain from her education. Skills are likely to be multi-dimensional and can hardly be captured even with good data on test scores. In addition, individuals have idiosyncratic knowledge about these skills. Since what matters for the individual’s decision is her perception of her skills and her beliefs about how they affect future earnings (conditional on her information set at the time of the decision), this provides a strong rationale for using “perceived” returns and “perceived” risk. We nevertheless also control for GPA and parental education to proxy for preferences for education and for the probability of completing senior high school or college.

To model the decision to enrol in college –taken by youths who have just completed senior high school–, an equation analogous to (1) can be used. 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:

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

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

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.

## 2.2 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 enrol 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 it is processed differ, subjective expectations of the different actors might differ. One goal of this paper is to test whether expectations differ systematically between parents and youths and whose expectations are relevant for the decisions, while we allow for gender differences.

If one had data on both parents' and youths' expectations one could address this question as follows. Neglecting the variance terms, the model we would want to estimate for both school attendance decisions is as follows:

$$S = 1 \Leftrightarrow S^* = \alpha + X'\beta + \gamma_P * Parents'Expectations + \gamma_Y * Youths'Expectations + W > 0. \quad (3)$$

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

In what follows we will show that this is the case in our data. As we discuss below, our data does contain information on the subjective expectations held by mothers as well as by the youth, but unfortunately this information is not as complete as we would like. Therefore the estimation of Equation (3) involves the solution of a number of econometric problems we discuss below.

## 3 Data

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 in 2002/3 a new component known as *Jóvenes con Oportunidades*, an evaluation aimed at measuring its impact was started. The data we use was collected in 2005 as part of that evaluation. As we discuss below, in addition to standard variables, the survey contained a detailed subjective expectation 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.

### 3.1 The Survey

The evaluation survey of *Jóvenes con Oportunidades* was conducted in the fall 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: 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.

The survey consists of a family questionnaire and a youth questionnaire administered to 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 about information on about 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 the following analysis we discuss in detail how we address this issue.

It is important to keep in mind that 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 23000 youths, 5700 individuals had just completed grade 9 and 4397 had completed senior high school ("preparatoria") and decided about enrolling in college.

Dropping those individuals for whom we miss information on any of the individual and family background variables which are used in the 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 Appendix B (made available online)). This leaves us with the final sample of 2364 junior high school graduates and 2829 senior high school graduates.

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.

## 3.2 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 practised 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 stops 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 maximum and minimum amount 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 father's 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 enrolment decisions had been made.

While the timing of the survey (which occurred just after the school enrolment decision had been taken) allows to use the expectations data to estimate an enrolment model, it might also raise a potential concern that individuals might try to rationalize their choices. For example, individuals who decided to enrol 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 enrol, state lower expected returns to college. This would lead to a more dispersed cross-section of returns after the decision compared to before the enrolment decision.<sup>4</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 enrolment 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 enrol 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 5.2 we discuss this issue at length and present evidence that our main results cannot be explained by ex-post rationalization.

### 3.3 Calculation of 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,  $y_{mid}^z = (y_{min}^z + y_{max}^z)/2$ , of the support,  $p = Pr(Y^z > y_{mid}^z)$ . Thus we need an additional distributional assumption,  $f(\cdot)$ , in order to be able

---

<sup>4</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 enrol 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 enrol in college are the ones with particularly high returns and they later state low expected returns.

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 different scenarios 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 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>5</sup>

In Attanasio and Kaufmann (2008), we use three different distributional assumptions, step-wise uniform, bi-triangular and triangular (compare Guiso, Jappelli, and Pistaferri (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.

## 4 Subjective Expectations.

Before using the data on subjective expectations to model schooling choices, we describe the general patterns and provide some evidence on their quality. 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.

### 4.1 Descriptive Statistics: Characteristics of the Youth.

Tables 1 and 2 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.

---

<sup>5</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.

One can summarize the evidence from Tables 1 and 2 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 we observe are conform to our expectations and confirm findings of other empirical studies, for example the fact that girls outperform boys at school.

There are some (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 towards 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.

## 4.2 Youth vs Mother Respondents: A Selection Equation.

To determine whose expectations matter for enrolment decisions, one would ideally want to have data on both parents' and youth' expectations. Unfortunately, we have to deal with two important issues. 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 the 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 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 the case where having access to data on fathers is not necessary anyways, since no father is present. Second, the questions on the subjective distributions of earnings were not asked to both mother and children at the same time.

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 to ask 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, 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 probability held by the mother and the subjective probability 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:

$$R = 1 \Leftrightarrow R^* = \delta + X'\kappa + Z'\lambda + \epsilon \geq 0, \quad (4)$$

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 set of variables. We find that, in our sample, the time of day of the interview does not affect much 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 is the respondent with F-statistics of 32 and 18 for young and old cohort respectively (see Tables 3 and 4). In particular, both for the young 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).

So far we have shown that the condition of instrument relevance is satisfied. In the following we will discuss the second criterium for instrument validity, that is whether the instrument is likely to satisfy the exclusion restriction. 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 the data section). 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 Tables 11 and

12 in Appendix B, made available online). Very importantly, the timing of the interview is not correlated with individuals educational decisions and only in few cases with individual and family background characteristics. The main assumption underlying our analysis 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) to test both for differences between youths’ and mothers’ expectations and to estimate the school decision model, as we discuss below.

### 4.3 Descriptive Statistics: Subjective Expectations

In Tables 5 and 6, we present summary statistics for the variables derived from the subjective expectations questions for the young and old cohort. The top two panels of the Table report expected log earnings and expected (gross) returns. The bottom two panels, instead, focus on perceived earnings and employment risk for three different 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 equation (4) 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.

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 (e.g. 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

---

<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 Table 17 in Appendix B.

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 column of Tables 5 and 6 present results of the test of the differences between youths' and mothers' expectations, for each of the subjective expectation measures, i.e. for expected earnings, expected returns, perceived probability of working, and earnings risk. Table 5 shows 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 the first two differences are significant. For boys, mothers' expectations of junior high school earnings are around 20% higher than those of the male youths, though the difference is not significant. Differences for the older cohort are smaller (15-25% 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 expect lower high school and college returns than youth (difference is not significant for college for boys).

Mothers perceive a significantly lower earnings risk than boys and girls with respect to junior high school earnings and in addition with respect to senior high school earnings for boys. The perceived probability of working is similar for mother and youth respondents, except for boys perceiving a higher probability of working with a senior high school.

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.

#### **4.4 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, Giné, and McKenzie (2011b) 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

---

<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.



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 Appendix B (made available online).

First we present evidence on the fraction of missing and inconsistent answers, separately for the four subsamples of interest, that is male and female youths respondents and their mothers. An additional interesting exercise consists in relating the expected value of future earnings to a number of observable variables, such as those listed in Tables 1 and 2. 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 detailed discussion in Appendix B).

Table 13 shows that individuals were indeed willing to answer the expectation questions. The fraction of missing responses is below 1.5% 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 (elicited probabilities were never larger than one given the elicitation method, see section 3.2) 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 analogous 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. Table 14 shows 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 on 10%).

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

---

<sup>8</sup>Our findings in terms of fraction of non-responses for expectation questions and fraction of inconsistent responses are line with, for example findings of Dominitz and Manski (1996), Delavande, Giné, and McKenzie (2011b) and Mahajan, Tarozzi, Yoong, and Blackburn (2011)

validity. At the same time, we need to stress 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.

Table 15 presents results for the young cohort. 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 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 Table 16 shows that a higher GPA is associated with significantly higher earnings, in particular for college as the highest degree. For boys expected earnings decrease in age (consistent with lower expected earnings for those who had to repeat a grade), increase in 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.

In Table 17 we compare expected earnings to observed (Census) earnings for different schooling degrees and show that both are in the same ballpark. In Appendix B we discuss the relevant data sources, 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 (for when they are 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 whose expectations matter for schooling choices.

## 5 Schooling Decisions and Returns to Schooling.

In this section, we address our main questions of interest: 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 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 we obtain modeling the schooling decisions of the two cohorts we consider: youths who have completed junior high school and decide whether to enrol in senior high school, and youths who decide whether to enrol 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 2.2 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 enrolment decision.

Unfortunately this simple strategy is not feasible because, while we do have information on both mothers' and youths' expectations, we do not have them for the same individuals. As we discussed above, we have two different samples, one with mother and one with youth respondents. This feature of the data poses two problems. First, we cannot consider both expectations in the same regression. Second, if we run separately two regressions, one with the youth expectations and one with the mother expectations on the two different samples, we have to take into account the fact that the two subsamples were not randomly selected (compare Section 4.3).

Dealing with the second problem is relatively straightforward. We can use the same sample selection model that we use to correct the test of the differences between mother and youth expectations. When estimating a regression on one of the two subsamples, we can use once again the exclusion restriction that the timing of the interview does not affect education choices and is not correlated with individuals' expectations, while it determines significantly which of the two samples an individual observation belongs to, and correct our estimates using a Heckman-type selection equation (i.e. we estimate jointly the selection equation and the main school choice equation using a Maximum Likelihood procedure).

Addressing the first problem with the data we have is much harder. Suppose that only 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 enrolment. Instead, the results from the sample where only youth expectations are available would provide inconsistent estimates, that is the coefficient on youth expectations might be zero or, to the extent that youth expectations are somewhat correlated with mother expectations, could attract a significant coefficient, albeit probably smaller in size, due to attenuation bias.

On the other extreme, if only the youth's expectations matter, we would obtain consistent estimates only from the sample with youth respondents and inconsistent estimates from the sample

with mothers respondents.

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

In Table 2, we have seen that 29% of male high school graduates enrol in college compared to 31% of girls. We model the college attendance decision using Equation (2), which relates the probability of enrolment 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 7, we present estimates for boys and girls in separate columns, although the coefficients are estimated jointly in a model where all regressors are interacted with gender dummies to test for gender differences. 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). In the following tables, we report coefficient estimates, while we present the complete set of tables with marginal effects in Appendix B, where regressors are evaluated at their median values.

For the sake of brevity, Table 7 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 Appendix B. As controls we use all individual and family background characteristics displayed in Table 2 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 enrol 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 enrol in college, while the likelihood is increasing in having a mother in the labor force. Cost variables, such as distance from university and the level of tuition fees in nearby universities are also important determinants of the decision, even though the level of tuition fees in a nearby university seems to affect only girls.

The key results of Table 18 (upper panel) are three: First, boys' own expectations about returns

to college are a strongly significant predictor for their decision to enrol 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>9</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 enrol 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 to attend college. For boys on the other hand, only their own expectations about returns to college predict their schooling decisions.<sup>10</sup>

Third, our data enable us to look at two potential determinants of schooling decisions which have often been neglected in this analysis, 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 5 and 6), 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 changes.

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, Golsteyn, Heckman, and Meijers (2009) and Schubert, Gysler, Brown, and Brachinger (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, while girls' expectations are not, it would be interesting to analyze whether boys' expectations are significant because they are correlated with the ones of father. Unfortunately, as discussed before,

---

<sup>9</sup>Also both boys and girls expect high returns to schooling.

<sup>10</sup>While the 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.

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 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, see Table 2). We report the coefficient estimates on the main variables of interest in the lower panel of Table 18.

Also for the subsample of families without father, we find that boys' own expectations are strongly significant predictors of their college attendance decision (significant on 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 (on 2%).

In Section 5.2 we discuss in detail the robustness of our results and present supportive evidence against the concern that these results might be driven by people rationalizing their choices *ex-post*. Although we control for potential self-selection problems because of non-randomness of who responds to the expectation 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>11</sup>

To summarize, our evidence shows that in the context of college enrolment 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 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 those types of families.

### 5.1.2 High School Attendance Choice

We now turn our attention to the decision to enrol or not in senior high school for youths who have just finished junior high school. As shown in Table 1, around 78% of boys and 81% of girls graduating from junior high school decide to enrol in senior high school. In Table 8, 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.

---

<sup>11</sup>Results remain basically the same when running the regressions without selection correction (see Appendix B).

As in the case of the college attendance decision, Table 8 only display the coefficients on the main variables of interest, that is expected labor market outcomes, while we present a table including all significant controls in Appendix B. As controls we use all individual and family background characteristics displayed in Table 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 is increasing in GPA and in parental education. A larger number of siblings decreases the likelihood to continue 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 to enrol 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 8 (upper panel) are as follows: First, the expectations of the 15-year-old youths do not predict the decision to enrol in high school. Mothers' expectations on the other hand are a predictor of girls' decision to enrol 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 enrol 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 enrolment in senior high school, as one would expect. The latter result suggests that having the option to enrol in college with a high school degree can be important for the high school enrolment 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>12</sup> Second, our results show that risk perceptions (of unemployment and earnings) can be very important in schooling decisions. Again, as in the case for the college attendance decision, they matter mostly from the perspective of the mother.

In a second step, we move our attention to the subsample of families without father, allowing the coefficients on the expectations to be different in this subsample, and report the results in the lower panel of Table 8. In 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 enrol in high school.

In families without father present, boys' own expectations are significant for their schooling decision, while mothers' are not. The likelihood to enrol in high school increases in boys' expected returns to college (significant on 5%), in 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

---

<sup>12</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 (on 5%) and close to significant in the case of the perceived earnings risk with a junior high school degree (p-value 0.11).

significant (girls' own expectations are significant in two instances, once with the expected sign on the probability of working with a college degree and once with a –surprisingly– negative coefficient on the probability of working with a senior high school degree, which is significant on 10%).

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

### 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 beliefs affecting the decision. For example, an individual might eat a whole cake and justify this decision by stating the belief that otherwise the cake would have gone bad. 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 ex-post 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 3.2. 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 enrolment decision of senior high school graduates. If individuals rationalize their choices, we would expect the following pattern: Individuals who decided to enrol in college rationalize their choice by stating higher expected college earnings or lower high school earnings. Those individuals who decided not to enrol 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>13</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-

---

<sup>13</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.



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 Tables 1 and 2 and state fixed effects).

In Tables 28 and 29 we report p-values of 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 we use in the main regression analysis instead of expected earnings) and perceived risks.

Table 28 displays 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's 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 1 and 2). 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 Figure 3). 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 are not significant in our main regressions, nor is the perception of the probability of working with a senior high school degree.

Table 29 illustrates the results of Kolmogorov-Smirnov tests for the cohort of junior high school graduates. For youth respondents, we can reject equality of distributions in two instances: for boys in terms of expected log earnings with a high school degree and 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 Table 29). 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 Figure 4). This cannot explain why mothers' beliefs about the probability of work (with a junior high school and a college degree) is 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).

### **Classical Endogeneity**

Another concern with using subjective expectations as determinants of schooling choices is another forms 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 or) more important than expected monetary returns and correlated with the latter. One possibility are 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. While it seems hard to come up with a story about omitted variables that can explain these results, it cannot be ruled out with absolute certainty. For that reason we control in the following section for one –potentially very important– additional return that we have so far omitted in our analysis, that is we control for proxies of marriage market returns.

### 5.3 Robustness: Returns in the Labor and Marriage Market

In this section we provide some suggestive evidence that there might be not only differences in the role of (expected) labor market outcomes related to schooling decisions between boys and girls, but also differences in the role of marriage market returns. Furthermore, we show that our conclusion of important gender differences in the role of expected labor market outcomes are not altered when controlling for proxies of marriage market returns.

In particular, in the following analysis, marriage market returns are proxied for by the (net-) supply of potential partners in the youths' local marriage markets. Using census data, we compute the ratio of unmarried men to women in the locality of residence. Limited availability of suitable partners in the locality of residence could constitute an important push-factor to leave the locality.<sup>14</sup> In Mexico social norms are such that Mexican parents are generally reluctant to let their unmarried daughter move to a bigger city to work and live by herself, while they are more willing to let her leave home to attend school and live with classmates. This could create a link between schooling decisions and marriage market considerations, which is likely to be less strong for boys, as they could migrate to a bigger city to work and find a spouse. But even for boys, in the presence of strong patterns of assortative mating, it might be less costly to find an educated partner at school or college. For this reason we interpret “suitable partner” in terms of a partner who has a similar or higher education level (and is in a similar age range) and thus use the ratio of unmarried men to women with an education level as high or higher than the youth deciding about schooling.

We construct our proxies for marriage market returns using locality level data, because we believe that this is the appropriate reference area (i.e. the relevant marriage “market”) for most youths. This proxy is more likely to capture people's actual knowledge about availability of partners than when using a larger reference area.

Our proxy for marriage market returns could, potentially, hide different effects. One key concern is that the sex ratio (conditional on education or unconditionally) captures labor market conditions

---

<sup>14</sup>Edlund (2005) uses a similar argument to explain migration patterns that lead to young women outnumbering young men in urban areas. To proxy for marriage market considerations she uses the ratio of men to women in the municipality of residence and finds supportive evidence for this explanation using Swedish municipality data.

that affect individuals decisions to attend college. In this context, our data on expectations about labor market outcomes is very useful, since what matters for an individual’s decision is her beliefs about her individual future labor market outcomes and we have direct measures of these beliefs. For that reason we can analyze jointly the role of labor market and marriage market considerations in schooling decisions.

A second important concern are potential ‘peer’ effects. We aim to address this concern in two ways. First, we control directly for the fraction of men and women with a senior high school or college degree in the locality (for the young and old cohort respectively), to see if this affects the coefficient on the sex ratio.<sup>15</sup> Second, we also consider an alternative proxy for the marriage market return: the ratio of unmarried men to women without constraining the measure to a specific education level. In that case, this ratio should not capture any peer effects on the schooling decision. Also this measure might be more appropriate, if the youth (for example, the male youth) cares less about the education level of the spouse.

In terms of the implementation, we are facing the following two issues: First, to be able to use proxies for marriage market returns at the local level, we use Census data of the year 2000 (see Appendix B for a detailed data description and discussion of the variable construction). Unfortunately, Census information is only available for part of the localities in which the youths of our sample live. For the sample of junior high school graduates we lose around 30% of our observations. For the sample of senior high school graduates, who live in smaller localities than the junior high school graduates, we lose close to 60% of observations. For this reason we only use this significantly smaller sample for our analysis of marriage market returns, while we use the full sample for the main set of results on labor market returns.

Secondly, since we do not have data on people’s subjective expectations of marriage market returns, but only a proxy which cannot capture differences in expectations of mothers and youths, we only present results for our youth respondents to see how the role of this proxy differs for boys versus girls.

### 5.3.1 College Attendance Choice

In the following, we first relate the college decision to our proxies of marriage market returns and to expected labor market outcomes and controls (see Table 9). As discussed above, we lose almost 60% of the observations when constructing the proxy with locality level Census data. Given the small sample size (around 300 observations for male and female youth respondents), we can only control for the most important variables to achieve convergence in our bivariate probit estimation.

The key results of this table are two (see upper panel of Table 9):<sup>16</sup> First, as in the previous analysis, expectations about returns to college are a strongly significant predictor for boys’ decision to enrol in college. Also in the case of this smaller sample, the coefficient on expected returns

---

<sup>15</sup>When constructing our control variable for potential peer effects, we consider both married and unmarried women (men) who went to college (in contrast to the sex ratio that takes into account only unmarried men and women with at least a high school degree), since all youth in our cohort already have a high school degree and we want to control for potential peer effects arising from observing others going to college.

<sup>16</sup>The full set of coefficients including the ones on the controls is displayed in Appendix B.

is again (significantly) larger for boys than girls and nearly three times as large in magnitude. But now girls' expectations are marginally significant as well (at the 10% level) given the more parsimonious specification we use on this smaller sample.

Secondly, Table 9 shows that –for girls– the coefficient on the ratio of unmarried women to men *conditional* on education is significant and has the expected sign: fewer available suitable partners in the locality of residence (i.e. many women per man) increase the likelihood of girls to attend college (see Column 2). For boys, the coefficient on the analogous ratio of unmarried men to women is negative and not significantly different from zero (Column 1). The difference between boys and girls of the coefficient on the marriage market proxy is significant at the 5% level.

The results so far suggest that marriage market considerations are indeed important for girls' college attendance decision, while this determinant appears to play a smaller (or no) role for boys. Of course, we might be concerned that our proxies for marriage market returns could, potentially, hide different effects. We have already addressed one key concern that the sex ratio in the locality of residence captures not only marriage market, but also labor market conditions that affect the individual's decision to attend college. In this context, our data on expectations about labor market outcomes proves useful, since we can directly control for these beliefs.

A second important concern is related to the fact that our proxy for marriage market returns (the sex ratio of unmarried men to women conditional on having at least a high school degree) could also capture 'peer' effects. We try to address this issue in two ways. First, we consider not only the sex ratio conditional on education, but also an alternative proxy: the ratio of unmarried men to women without constraining the measure to a specific education level. Thus, this ratio should not capture any peer effects on the schooling decision. As discussed, this measure might also be more appropriate, if the youth (for example, the male youth) cares less about the education level of the spouse. Second, we control directly for the fraction of men and women with a college degree in the locality, to see if this affects the coefficient on the sex ratio.<sup>17</sup>

Columns 3 and 4 in Table 9 show the same pattern for the proxy without conditioning on education. Also for this alternative proxy, the coefficient for girls is positive and significant (at the 5% level), that is fewer available partners (i.e. more women per man) increase the likelihood of a girl to decide to enrol in college. The coefficient on the ratio for boys is not significant and significantly different from the coefficient for girls (on 3%).

Controlling for potential peer effects directly, that is controlling for the fraction of women (men) in the locality who have a college degree, these results remain unchanged. The lower panel of Table 9 shows that the marriage market proxy conditional on education actually becomes larger (and is now significant on 5%) when controlling for potential peer effects. At the same time our proxy for potential peer effects is strongly significant for girls, i.e. girls are more likely to go to college if they live in a locality of highly educated women. These two robustness checks provide suggestive evidence that the correlation we find between schooling decisions and our marriage market proxies

---

<sup>17</sup>When constructing our control variable for potential peer effects, we consider both married and unmarried women (men) and only take into account women (men) who went to college, since all youth in our cohort already have a high school degree and we want to control for potential peer effects in the college decision arising from observing others going to college.

is not driven by peer effects in education.

To summarize, our results are suggestive of important gender differences not only in terms of labor market returns as determinants of college attendance decisions, but also of gender differences in terms of the role of marriage market returns. While for boys expected returns to college seem to be particularly important, expected labor market outcomes appear less important for girls. For girls on the other hand, marriage market considerations seem to play a role in the college enrolment decision, consistent with anecdotal evidence (compare the Mexican acronym “mmc”s (*‘mientras me caso’*) for girls going to college to find a husband).

### 5.3.2 High School Attendance Choice

In this section we relate the high school attendance decision to our proxies of marriage market returns and to expected labor market outcomes and controls (see Table 10). As discussed above, we lose about 30% of the observations when constructing the proxy with locality level Census data.

Table 10 shows results for the two different proxies for marriage market returns, the one conditional on education (Columns 1 and 2) and unconditionally (Columns 3 and 4).<sup>18</sup> As in the previous analysis, labor market expectations of the 15-year-old youths do not predict their decision to enrol in senior high school.

Column 2 (upper panel of Table 10) shows that –for girls– the coefficient on the ratio of unmarried women to men *conditional* on education is significant and has the expected sign: fewer available suitable partners in the locality of residence (i.e. many women per man) increase the likelihood of girls to attend college. For boys, the coefficient on the analogous ratio of unmarried men to women is negative and not significantly different from zero (Column 1). The difference between boys and girls of the coefficient on the marriage market proxy is significant at the 5% level. Results remain extremely similar when controlling for the fraction of at least high school educated individuals in the same locality (see lower panel of Table 10).

Columns 3 and 4 in Table 10 show results for the proxy without conditioning on education. In this case the coefficient on the proxy is not significant, neither for girls nor boys. This result is consistent with girls caring about finding a husband who is at least as highly educated as themselves, but this is of course speculation given that the proxy for marriage market returns is at best a very noisy measure of girls’ expected marriage market returns.

## 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 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.

---

<sup>18</sup>The full set of coefficients including the ones on the controls is displayed in Appendix B.

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.

To analyze whether our results might be driven by omitting another potentially important return to schooling, that is returns in the marriage market, we include as a control a proxy for marriage market returns that captures the (net-) supply of suitable potential partners in the youths' locality of residence. We show that our results on the relevance of expected labor market outcomes remain unchanged. At the same time, our results are consistent with marriage market returns playing an important role for girls' schooling decision.

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 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.

## 7 Acknowledgements

We would like to thank Sandra Black, Janet Currie, Pascaline Dupas, Lena Edlund, Claudio Ferraz, Caroline Hoxby, Eliana La Ferrara, Lance Lochner, Shaun McRae, Matthias Messner, Karthik Muralidharan, Bernard Salanie, Aloysius Siow, Kevin Stange, Alessandro Tarozzi, Wilbert van der Klaauw and Basit Zafar for valuable comments. We are thankful also to conference participants at the NBER Education meeting in Spring 2010, at the CEPR labour 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.



## References

- ABRAMITZKY, R., AND V. LAVY (2011): “How Responsive is Investment in Schooling to Changes in Redistribution Policies and in Returns?,” *mimeo*.
- ARCIDIACONO, P., V. J. HOTZ, AND S. KANG (2012): “Modelling College Major Choices using Elicited Measures of Expectations and Counterfactuals,” *Journal of Econometrics*, 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., AND K. M. KAUFMANN (2008): “Subjective Returns to Schooling and Risk Perceptions of Future Earnings - Elicitation and Validation of Subjective Distributions of Future Earnings,” *mimeo*.
- ATTANASIO, O. P., C. MEGHIR, AND A. SANTIAGO (2011): “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progreso,” *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., P.-A. CHIAPPORI, AND C. MEGHIR (2005): “Collective Labor Supply With Children,” *Journal of Political Economy*, 113(6), 1277–1306.
- BORGHANS, L., B. H. H. GOLSTEYN, J. J. HECKMAN, AND H. MEIJERS (2009): “Gender Differences in Risk Aversion and Ambiguity Aversion,” *Journal of the European Economic Association, MIT Press*, 8(2-3), 649–658.
- BOURGUIGNON, F. (1999): “The Cost of Children: May the Collective Approach to Household Behavior Help?,” *Journal of Population Economics*, 12(4), 503–522.
- CAMERON, S. V., AND J. J. HECKMAN (1998): “Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males,” *Journal of Political Economy*, 106(2), 262–333.
- (2001): “The Dynamics of Educational Attainment for Black, Hispanic and White Males,” *Journal of Political Economy*, 109(3), 455–499.
- CARD, D. (2001): “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,” *Econometrica*, 69(5), 1127–60.
- CHIAPPORI, P.-A., M. IYIGUN, AND Y. WEISS (2009): “Investment in Schooling and the Marriage Market,” *American Economic Review*, 99(5), 1689–1713.
- DAUPHIN, A., A.-R. E. LAHGA, B. FORTIN, AND G. LACROIX (2008): “Are Children Decision-Makers Within the Household?,” *IZA Discussion Paper No. 3728*.

- DELAVANDE, A., X. GINÉ, AND D. MCKENZIE (2011a): “Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence,” *Journal of Development Economics*, 94(2), 151–163.
- (2011b): “Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence,” *Journal of Development Economics*, 94(2), 151–163.
- DINKELMAN, T., AND C. MARTINEZ (2011): “Investing in Schooling in Chile: The Role of Information about Financial Aid For Higher Education,” *mimeo*.
- DOMINITZ, J., AND C. F. MANSKI (1996): “Eliciting Student Expectations of the Returns to Schooling,” *Journal of Human Resources*, 31 (1), 1–26.
- EDLUND, L. (2005): “Sex and the City,” *Scandinavian Journal of Economics*, 107(1), 25–44.
- GIUSTINELLI, P. (2011): “Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choices of High School Tracks,” *Mimeo*.
- GOLDIN, C., L. KATZ, AND I. KUZIEMKO (2006): “The Homecoming of American College Women: The Reversal of the College Gender Gap,” *Journal of Economic Perspectives*, 20(4), 133–156.
- GUIO, L., T. JAPPELLI, AND L. PISTAFERRI (2002): “An Empirical Analysis of Earnings and Employment Risk,” *Journal of Business and Economic Statistics*, 20 (2), 241–253.
- HECKMAN, J. J. (1979): “Sample Selection as a Specification Error,” *Econometrica*, 47 (1), 153–161.
- JENSEN, R. (2010): “The (Perceived) Returns to Education and the Demand for Schooling,” *Quarterly Journal of Economics*, 125(2), 515–548.
- (2012): “Do Labor Market Opportunities Affect Young Women’s Work and Family Decisions? Experimental Evidence from India,” *Quarterly Journal of Economics*, 127(2), 753–792.
- KAUFMANN, K. M. (2009): “Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns to College,” *mimeo*.
- KAUFMANN, K. M., M. MESSNER, AND A. SOLIS (2012): “Returns to Elite Higher Education in the Marriage Market: Evidence from Chile,” *CESifo Working Paper*.
- KEANE, M. P., AND K. I. WOLPIN (1997): “Career Decisions of Young Men,” *Journal of Political Economy*, 105 (June), 473–522.
- LAFORTUNE, J. (2012): “Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market,” *American Economic Journal: Applied Economics*.
- MAHAJAN, A., A. TAROZZI, J. YOONG, AND B. BLACKBURN (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., M. WISWALL, AND B. ZAFAR (August 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 627*.
- SCHUBERT, R., M. GYSLER, M. BROWN, AND H.-W. BRACHINGER (1999): “Financial Decision-Making: Are Women Really More Risk-Averse?,” *American Economic Review*, 89, 381–385.
- STINEBRICKNER, T., AND R. STINEBRICKNER (2012): “Learning about Academic Ability and the College Drop-Out Decision,” *Journal of Labor Economics*, 30(4), 707–748.
- TODD, P. E., AND K. I. WOLPIN (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,” *American Economic Review*, 96(5), 1384–1417.
- WISWALL, M., AND B. ZAFAR (2011): “Determinants of College Major Choice: Identification using an Information Experiment,” *Federal Reserve Bank of New York Staff Report No. 500*, June 2011.
- ZAFAR, B. (2009): “College Major Choice and the Gender Gap,” *Federal Reserve Bank of New York Staff Reports*.

Table 1: Individual and Family-Background Characteristics: Junior High School Graduates

Junior HS Graduates: Respondent:	Total			Boys			Girls		
	Boys Mean/(SE)	Girls Mean/(SE)	Diff (P-Val)	Youth Mean/(SE)	Mother Mean/(SE)	Diff (P-Val)	Youth Mean/(SE)	Mother Mean/(SE)	Diff (P-Val)
Enrollment Rate: Senior HS	0.793 (0.405)	0.835 (0.371)	(0.009)	0.809 (0.393)	0.771 (0.420)	(0.135)	0.841 (0.366)	0.826 (0.379)	(0.498)
Age	15.641 (0.032)	15.488 (0.027)	(0.000)	15.661 (0.045)	15.613 (0.045)	(0.447)	15.486 (0.035)	15.490 (0.042)	(0.946)
GPA of Jr HS (Scale 0-100)	79.323 (8.216)	82.358 (7.702)	(0.000)	79.087 (8.035)	79.642 (8.452)	(0.270)	82.059 (7.173)	82.818 (8.438)	(0.099)
Obese	0.022 (0.148)	0.025 (0.155)	(0.072)	0.027 (0.161)	0.016 (0.129)	(0.264)	0.019 (0.139)	0.032 (0.177)	(0.181)
Mother's Educ - Primary	0.692 (0.462)	0.715 (0.451)	(0.281)	0.684 (0.465)	0.702 (0.457)	(0.561)	0.699 (0.458)	0.738 (0.439)	(0.182)
Mother's Educ - Jr HS	0.269 (0.444)	0.244 (0.430)	(0.217)	0.272 (0.445)	0.265 (0.442)	(0.814)	0.252 (0.434)	0.233 (0.423)	(0.503)
Mother's Educ - Sr HS	0.032 (0.176)	0.030 (0.171)	(0.798)	0.041 (0.198)	0.021 (0.143)	(0.089)	0.038 (0.191)	0.017 (0.132)	(0.047)
Mother's Educ - Univ	0.005 (0.175)	0.010 (0.099)	(0.291)	0.002 (0.045)	0.010 (0.102)	(0.132)	0.009 (0.099)	0.010 (0.100)	(0.975)
Father's Educ - Primary	0.686 (0.464)	0.695 (0.461)	(0.727)	0.686 (0.464)	0.685 (0.465)	(0.972)	0.695 (0.461)	0.693 (0.461)	(0.963)
Father's Educ - Jr HS	0.244 (0.429)	0.234 (0.423)	(0.648)	0.251 (0.434)	0.234 (0.424)	(0.609)	0.239 (0.427)	0.225 (0.418)	(0.664)
Father's Educ - Sr HS	0.062 (0.242)	0.057 (0.233)	(0.708)	0.054 (0.227)	0.073 (0.261)	(0.329)	0.056 (0.231)	0.059 (0.237)	(0.848)
Father's Educ - Univ	0.007 (0.085)	0.014 (0.116)	(0.230)	0.007 (0.086)	0.007 (0.083)	(0.943)	0.009 (0.094)	0.021 (0.144)	(0.211)
Per Cap Income - 5 to 10k	0.350 (0.477)	0.334 (0.471)	(0.421)	0.363 (0.481)	0.331 (0.471)	(0.272)	0.325 (0.468)	0.348 (0.476)	(0.406)
Per Cap Income - above 10k	0.222 (0.416)	0.25 (0.433)	(0.106)	0.219 (0.414)	0.226 (0.418)	(0.789)	0.267 (0.442)	0.224 (0.417)	(0.084)
Father's Occup - Unsk. Work	0.301 (0.459)	0.273 (0.446)	(0.227)	0.277 (0.447)	0.334 (0.472)	(0.086)	0.288 (0.453)	0.251 (0.434)	(0.237)
Father's Occup - Employee	0.489 (0.500)	0.520 (0.499)	(0.209)	0.542 (0.498)	0.416 (0.493)	(0.000)	0.508 (0.500)	0.539 (0.499)	(0.373)
Father's Occup - Employer	0.009 (0.094)	0.008 (0.091)	(0.890)	0.009 (0.094)	0.009 (0.095)	(0.978)	0.012 (0.108)	0.003 (0.055)	(0.121)
Father's Occup - Self-Empl.	0.184 (0.387)	0.181 (0.385)	(0.908)	0.161 (0.367)	0.215 (0.412)	(0.054)	0.182 (0.386)	0.182 (0.386)	(1.000)
Father's Occup - Fam. Work	0.016 (0.128)	0.015 (0.123)	(0.851)	0.011 (0.105)	0.024 (0.154)	(0.182)	0.009 (0.099)	0.024 (0.154)	(0.133)
Mother in Labor Force	0.295 (0.014)	0.323 (0.013)	(0.141)	0.301 (0.018)	0.288 (0.021)	(0.628)	0.350 (0.017)	0.281 (0.020)	(0.110)
No Father in Household (Mother Single/Sep/Div)	0.208 (0.406)	0.254 (0.435)	(0.008)	0.215 (0.411)	0.199 (0.399)	(0.514)	0.262 (0.440)	0.241 (0.428)	(0.396)
Number of Male Siblings	1.43 (1.163)	1.354 (1.179)	(0.116)	1.38 (1.137)	1.496 (1.195)	(0.103)	1.35 (1.198)	1.36 (1.149)	(0.884)
Number of Female Siblings	1.411 (1.206)	1.446 (1.221)	(0.477)	1.381 (1.166)	1.452 (1.259)	(0.334)	1.403 (1.208)	1.514 (1.239)	(0.118)
Locality Size - 15 to 50k	0.185 (0.389)	0.189 (0.392)	(0.817)	0.177 (0.382)	0.196 (0.397)	(0.411)	0.167 (0.373)	0.223 (0.416)	(0.017)
Locality Size - above 50k	0.318 (0.466)	0.339 (0.473)	(0.288)	0.349 (0.477)	0.276 (0.447)	(0.009)	0.375 (0.484)	0.283 (0.451)	(0.001)
Observations	1111	1253		638	473		759	494	

Table 2: Individual and Family-Background Characteristics: Senior High School Graduates

Senior HS Graduates: Respondent:	Total			Boys			Girls		
	Boys	Girls	Diff	Youth	Mother	Diff	Youth	Mother	Diff
	Mean/(SE)	Mean/(SE)	(P-Val)	Mean/(SE)	Mean/(SE)	(P-Val)	Mean/(SE)	Mean/(SE)	(P-Val)
Enrollment Rate: College	0.289 (0.453)	0.316 (0.645)	(0.114)	0.243 (0.429)	0.325 (0.468)	(0.001)	0.237 (0.426)	0.401 (0.490)	(0.000)
Age	18.632 (0.032)	18.487 (0.029)	(0.000)	18.607 (0.045)	18.653 (0.044)	(0.472)	18.472 (0.039)	18.503 (0.043)	(0.586)
GPA of Jr HS (Scale 0-100)	81.469 (7.819)	82.922 (9.623)	(0.000)	81.032 (6.722)	81.816 (8.578)	(0.063)	83.021 (7.419)	82.818 (11.519)	(0.686)
Obese	0.021 (0.144)	0.031 (0.173)	(0.096)	0.018 (0.136)	0.023 (0.150)	(0.595)	0.032 (0.176)	0.030 (0.171)	(0.819)
Mother's Educ - Primary	0.800 (0.400)	0.782 (0.413)	(0.286)	0.794 (0.405)	0.805 (0.397)	(0.674)	0.791 (0.407)	0.772 (0.419)	(0.455)
Mother's Educ - Jr HS	0.173 (0.378)	0.179 (0.383)	(0.724)	0.181 (0.385)	0.168 (0.374)	(0.590)	0.171 (0.376)	0.188 (0.391)	(0.442)
Mother's Educ - Sr HS	0.020 (0.141)	0.036 (0.187)	(0.022)	0.018 (0.133)	0.022 (0.147)	(0.651)	0.033 (0.180)	0.039 (0.194)	(0.594)
Mother's Educ - Univ	0.005 (0.076)	0.002 (0.050)	(0.24)	0.007 (0.082)	0.005 (0.071)	(0.729)	0.005 (0.071)	0.000 (0.000)	(0.083)
Father's Educ - Primary	0.760 (0.427)	0.754 (0.43)	(0.772)	0.747 (0.435)	0.769 (0.421)	(0.445)	0.763 (0.425)	0.746 (0.436)	(0.529)
Father's Educ - Jr HS	0.183 (0.386)	0.187 (0.389)	(0.826)	0.191 (0.394)	0.176 (0.381)	(0.574)	0.181 (0.386)	0.192 (0.394)	(0.674)
Father's Educ - Sr HS	0.045 (0.207)	0.047 (0.212)	(0.831)	0.053 (0.224)	0.039 (0.195)	(0.372)	0.051 (0.220)	0.043 (0.204)	(0.575)
Father's Educ - Univ	0.011 (0.107)	0.011 (0.105)	(0.955)	0.008 (0.091)	0.014 (0.117)	(0.433)	0.004 (0.064)	0.018 (0.135)	(0.033)
Per Cap Income - 5 to 10k	0.262 (0.440)	0.283 (0.451)	(0.216)	0.252 (0.434)	0.271 (0.445)	(0.429)	0.269 (0.444)	0.298 (0.457)	(0.223)
Per Cap Income - above 10k	0.184 (0.388)	0.183 (0.386)	(0.935)	0.187 (0.390)	0.181 (0.385)	(0.811)	0.179 (0.383)	0.187 (0.390)	(0.661)
Father's Occup - Unsk. Work	0.443 (0.497)	0.430 (0.495)	(0.559)	0.408 (0.492)	0.469 (0.499)	(0.057)	0.413 (0.493)	0.448 (0.497)	(0.236)
Father's Occup - Employee	0.278 (0.448)	0.284 (0.451)	(0.727)	0.312 (0.464)	0.252 (0.434)	(0.038)	0.312 (0.464)	0.256 (0.437)	(0.035)
Father's Occup - Employer	0.006 (0.077)	0.009 (0.098)	(0.349)	0.005 (0.069)	0.007 (0.084)	(0.634)	0.014 (0.116)	0.005 (0.073)	(0.150)
Father's Occup - Self-Empl.	0.234 (0.423)	0.241 (0.427)	(0.725)	0.229 (0.421)	0.238 (0.426)	(0.745)	0.226 (0.419)	0.256 (0.437)	(0.247)
Father's Occup - Fam. Work	0.038 (0.192)	0.034 (0.182)	(0.599)	0.045 (0.208)	0.034 (0.181)	(0.362)	0.034 (0.182)	0.034 (0.182)	(0.987)
Mother in Labor Force	0.186 (0.011)	0.208 (0.010)	(0.1331)	0.192 (0.016)	0.180 (0.014)	(0.5902)	0.211 (0.015)	0.205 (0.015)	(0.7890)
No Father in Household (Mother Single/Sep/Div)	0.153 (0.360)	0.160 (0.367)	(0.592)	0.182 (0.386)	0.130 (0.337)	(0.011)	0.166 (0.372)	0.155 (0.362)	(0.553)
Number of Male Siblings	1.444 (1.27)	1.361 (1.171)	(0.07)	1.379 (1.274)	1.496 (1.265)	(0.095)	1.312 (1.122)	1.413 (1.22)	(0.096)
Number of Female Siblings	1.263 (1.164)	1.457 (1.314)	(0.000)	1.237 (1.187)	1.285 (1.145)	(0.457)	1.417 (0.301)	1.499 (1.327)	(0.228)
Distance to Univ - 20 to 40km	0.243 (0.429)	0.263 (0.44)	(0.242)	0.25 (0.433)	0.239 (0.426)	(0.626)	0.267 (0.443)	0.258 (0.438)	(0.698)
Distance to Univ - above 40km	0.237 (0.425)	0.218 (0.414)	(0.244)	0.235 (0.424)	0.239 (0.426)	(0.871)	0.204 (0.403)	0.234 (0.423)	(0.166)
Municip with Univ (Tuit above median)	0.391 (0.488)	0.366 (0.482)	(0.169)	0.449 (0.498)	0.344 (0.475)	(0.001)	0.416 (0.493)	0.312 (0.463)	(0.000)
Locality Size - 15 to 50k	0.058 (0.234)	0.065 (0.246)	(0.465)	0.081 (0.272)	0.041 (0.198)	(0.003)	0.056 (0.231)	0.074 (0.262)	(0.174)
Locality Size - above 50k	0.09 (0.286)	0.084 (0.278)	(0.537)	0.103 (0.304)	0.08 (0.271)	(0.155)	0.105 (0.307)	0.063 (0.243)	(0.003)
Observations	1320	1509		583	737		778	731	

Table 3: First-stage Regression for Whether the Youth Responds Herself: Young Cohort

Dep Var:	Youth Respondent: Yes/No		
	Instruments Coeff/(SE)	Instr and Interact Coeff/(SE)	Instr and Controls Coeff/(SE)
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)
GPA * Male			-0.003** (0.002)
GPA * Female			-0.003 (0.002)
No Father in HH * Male			0.038 (0.043)
No Father in HH * Female			-0.079* (0.041)
Mother Sr HS * Male			0.106 (0.091)
Mother Sr HS * Female			0.143* (0.087)
Father Employee * Male			0.059* (0.036)
Father Employee * Female			-0.043 (0.035)
Father Self-Empl * Male			-0.077* (0.047)
Father Self-Empl * Female			-0.063 (0.047)
Mother in LF * Male			-0.028 (0.034)
Mother in LF * Female			0.057* (0.033)
Loc Size above 50k * Male			0.046 (0.035)
Loc Size above 50k * Female			0.062* (0.033)
Age * Gender	No	No	Not Sign
Obesity * Gender	No	No	Not Sign
N of Siblings * Gender	No	No	Not Sign
Father's Educ * Gender	No	No	Not Sign
Parental Income * Gender	No	No	Not Sign
Observations	2364	2364	2364
F-Statistic	32.452	14.595	7.566
R-Squared	0.040	0.042	0.143
Adjusted R-Squared	0.038	0.039	0.124

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects (fully interacted with gender dummies).

Table 4: First-stage Regression for Whether the Youth Responds Herself: Old Cohort

Dep Var:	Youth Respondent: Yes/No		
	Instruments Coeff/(SE)	Instr and Interact Coeff/(SE)	Instr and Controls Coeff/(SE)
Interview on a Weekday	-0.069*** (0.026)	-0.087** (0.034)	-0.085** (0.035)
Interview in Week 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)
GPA * Male			-0.005*** (0.002)
GPA * Female			-0.000 (0.001)
N of Siblings * Male			-0.009 (0.008)
N of Siblings * Female			-0.019** (0.008)
No Father in HH * Male			0.132*** (0.046)
No Father in HH * Female			-0.008 (0.041)
Father Employee * Male			0.088** (0.038)
Father Employee * Female			0.053 (0.035)
Loc Size 15 to 50k * Male			0.191*** (0.059)
Loc Size 15 to 50k * Female			0.032 (0.053)
Loc Size above 50k * Male			-0.010 (0.053)
Loc Size above 50k * Female			0.113** (0.050)
Age * Gender	No	No	Not Sign
Obesity * Gender	No	No	Not Sign
Mother's Education * Gender	No	No	Not Sign
Father's Education * Gender	No	No	Not Sign
Parental Income * Gender	No	No	Not Sign
Mother in Labor Force * Gender	No	No	Not Sign
Distance to Univ * Gender	No	No	Not Sign
Tuition Costs * Gender	No	No	Not Sign
Observations	2829	2829	2829
F-Statistic	18.284	10.376	4.638
R-Squared	0.019	0.025	0.087
Adjusted R-Squared	0.018	0.023	0.068

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 2 and state fixed effects (fully interacted with gender dummies).

Table 5: Subjective Expectations of Future Earnings: Junior High School Graduates

Junior HS Graduates Resp:	Boys			Girls		
	Youth	Mother	Corr Diff (Y-M) (P-Val)	Youth	Mother	Corr Diff (Y-M) (P-Val)
	Mean/(SE)			Mean/(SE)		
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 First-stage Tables). In addition we correct for potential differences in the composition of the mother and youth samples.



Table 6: Subjective Expectations of Future Earnings: Senior High School Graduates

Senior HS Graduates: Resp:	Boys			Girls		
	Youth	Mother	Corr Diff (Y-M) (P-Val)	Youth	Mother	Corr Diff (Y-M) (P-Val)
	Mean/(SE)			Mean/(SE)		
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.

Table 7: College Attendance Choice: Youth and Mother Expectations

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
<b>Pooled Data</b>				
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)
<b>By Presence of Father</b>				
Expected Return College * No Father in Household	1.318*** (0.409)	0.217 (0.318)	-0.389 (0.425)	0.444 (0.372)
Prob of Work Sr HS * No Father in Household	-0.000 (0.918)	0.945 (0.951)	-1.219 (1.047)	0.848 (0.842)
Prob of Work College * No Father in Household	-0.502 (1.100)	-0.683 (1.095)	0.401 (1.271)	2.280** (1.066)
Var of Log Earn Sr HS * No Father in Household	0.404 (24.533)	-36.804 (27.107)	-2.860 (22.830)	14.687 (20.441)
Var of Log Earn College * No Father in Household	-17.356 (35.725)	-1.807 (17.793)	19.396 (36.115)	8.706 (24.441)
Expected Return College * Father in Household	0.274 (0.196)	0.071 (0.157)	-0.022 (0.157)	0.015 (0.158)
Prob of Work Sr HS * Father in Household	-0.819 (0.551)	-0.000 (0.398)	-0.543 (0.397)	0.028 (0.405)
Prob of Work College * Father in Household	0.075 (0.507)	-0.203 (0.389)	0.450 (0.393)	0.280 (0.396)
Var of Log Earn Sr HS * Father in Household	-8.182 (10.536)	-8.916 (8.183)	14.041 (10.153)	-0.453 (8.862)
Var of Log Earn College * Father in Household	-4.958 (13.816)	8.184 (10.720)	-22.514* (12.521)	10.707 (13.294)
Observations	2829		2829	
Censored Obs	1468		1361	

Notes: Table displays coefficients and standard errors in brackets. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ . Included controls are all variables in Table 2 and state fixed effects (fully interacted with gender dummies).

Table 8: High School Attendance Choice: Youth and Mother Expectations

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
<b>Pooled Data</b>				
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)
<b>By Presence of Father</b>				
Expected Return Sr HS * No Father in Household	0.500 (0.496)	-0.111 (0.461)	0.131 (0.643)	-0.384 (0.600)
Expected Return College * No Father in Household	0.907** (0.444)	0.280 (0.399)	0.371 (0.697)	-0.011 (0.511)
Prob of Work Jr HS * No Father in Household	-0.827 (0.796)	-0.534 (0.762)	-1.272 (1.279)	-0.026 (1.064)
Prob of Work Sr HS * No Father in Household	1.641* (0.930)	1.311 (1.036)	-2.095 (2.137)	-2.624* (1.418)
Prob of Work College * No Father in Household	0.312 (0.910)	-2.320* (1.348)	0.985 (1.870)	2.422** (1.174)
Var of Log Earn Jr HS * No Father in Household	-0.791 (21.235)	14.693 (21.792)	15.255 (20.122)	-19.453 (19.893)
Var of Log Earn Sr HS * No Father in Household	-1.201 (27.531)	15.146 (31.630)	36.600 (47.002)	-5.700 (31.236)
Var of Log Earn College * No Father in Household	92.069* (48.860)	-4.134 (24.867)	24.736 (40.142)	84.584 (61.050)
Expected Return Sr HS * Father in Household	-0.348 (0.241)	0.092 (0.216)	-0.009 (0.298)	-0.165 (0.334)
Expected Return College * Father in Household	0.023 (0.192)	0.197 (0.210)	-0.200 (0.284)	-0.461 (0.287)
Prob of Work Jr HS * Father in Household	0.499 (0.481)	0.070 (0.404)	1.138** (0.540)	-1.351** (0.617)
Prob of Work Sr HS * Father in Household	-1.038 (0.670)	0.002 (0.555)	-1.235 (0.758)	0.484 (0.873)
Prob of Work College * Father in Household	0.749 (0.481)	-0.244 (0.475)	0.970 (0.609)	0.194 (0.714)
Var of Log Earn Jr HS * Father in Household	3.794 (8.004)	12.040 (9.889)	20.103* (11.552)	-3.968 (10.676)
Var of Log Earn Sr HS * Father in Household	11.935 (11.128)	-9.817 (9.492)	-11.611 (12.769)	-22.208** (11.103)
Var of Log Earn College * Father in Household	-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. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects (fully interacted with gender dummies).

Table 9: College Attendance Choice: Labor and Marriage Market Returns

Dependent Variable	College Attendance Decision			
	Marriage Boys	Market Proxy 1 Girls	Marriage Boys	Market Proxy 2 Girls
<b>Without Direct Controls for Peer Effects</b>				
Ratio of Unmarried Men to Women (Sr HS)	-0.147 (0.136)			
Ratio of Unmarried Women to Men (Sr HS)		0.186* (0.100)		
Ratio of Unmarried Men to Women			-0.589 (0.449)	
Ratio of Unmarried Women to Men				0.642** (0.282)
Expected Return - College	0.748*** (0.247)	0.279* (0.158)	0.721*** (0.247)	0.277* (0.155)
Prob of Work - Sr HS	0.056 (0.569)	0.44 (0.460)	0.085 (0.562)	0.427 (0.456)
Prob of Work - College	-0.514 (0.599)	0.383 (0.553)	-0.414 (0.591)	0.364 (0.543)
Log Var of Earn - Sr HS	7.028 (11.383)	12.938 (10.290)	9.078 (11.368)	15.564 (10.049)
Log Var of Earn - College	-0.144 (18.474)	-6.242 (13.371)	-2.360 (18.217)	-5.960 (13.145)
<b>With Direct Controls for Peer Effects</b>				
Ratio of Unmarried Men to Women (Sr HS)	-0.102 (0.147)			
Ratio of Unmarried Women to Men (Sr HS)		0.223** (0.107)		
Ratio of Unmarried Men to Women			-0.5 (0.474)	
Ratio of Unmarried Women to Men				0.605** (0.294)
Expected Return - College	0.782*** (0.248)	0.343** (0.166)	0.764*** (0.251)	0.331** (0.164)
Prob of Work - Sr HS	0.091 (0.601)	0.400 (0.484)	0.113 (0.599)	0.400 (0.483)
Prob of Work - College	-0.563 (0.632)	0.314 (0.590)	-0.48 (0.630)	0.309 (0.583)
Log Var of Earn - Sr HS	5.047 (12.408)	12.272 (10.633)	6.961 (12.489)	15.556 (10.454)
Log Var of Earn - College	-0.062 (19.671)	-7.841 (14.337)	-1.874 (19.576)	-7.019 (14.238)
Fraction of College Educated Men (Women)	1.440 (1.228)	3.491*** (1.225)	1.309 (1.223)	2.958** (1.222)
Observations		1237		1237
Censored Obs		564		564

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01.

Table 10: High School Attendance Choice: Labor and Marriage Market Returns

Dependent Variable	High School Attendance Decision			
	Marriage Market Proxy 1		Marriage Market Proxy 2	
	Boys	Girls	Boys	Girls
<b>Without Direct Controls for Peer Effects</b>				
Ratio of Unmarried Men to Women (Jr HS)	0.000 (0.103)			
Ratio of Unmarried Women to Men (Jr HS)		0.627** (0.300)		
Ratio of Unmarried Men to Women			-0.064 (0.425)	
Ratio of Unmarried Women to Men				0.551 (0.502)
Expected Return - Sr HS	-0.041 (0.205)	0.098 (0.190)	-0.037 (0.206)	0.147 (0.193)
Expected Return - College	0.277 (0.172)	0.264 (0.179)	0.242 (0.174)	0.297 (0.182)
Prob of Work - Jr HS	0.099 (0.379)	-0.030 (0.332)	0.131 (0.383)	0.004 (0.336)
Prob of Work - Sr HS	0.028 (0.501)	-0.224 (0.444)	-0.054 (0.509)	-0.226 (0.448)
Prob of Work - College	0.656 (0.406)	-0.223 (0.462)	0.584 (0.413)	-0.170 (0.466)
Var of Log Earn - Jr HS	4.516 (6.517)	9.129 (7.926)	3.648 (6.568)	9.552 (8.072)
Var of Log Earn - Sr HS	13.750 (9.634)	-8.669 (8.244)	12.779 (9.654)	-9.381 (8.369)
Var of Log Earn - College HS	5.400 (11.112)	4.862 (10.876)	6.451 (11.333)	5.863 (10.968)
<b>With Direct Controls for Peer Effects</b>				
Ratio of Unmarried Men to Women (Jr HS)	-0.005 (0.107)			
Ratio of Unmarried Women to Men (Jr HS)		0.650** (0.313)		
Ratio of Unmarried Men to Women			-0.055 (0.434)	
Ratio of Unmarried Women to Men				0.553 (0.512)
Expected Return - Sr HS	-0.044 (0.208)	0.100 (0.193)	-0.042 (0.209)	0.150 (0.196)
Expected Return - College	0.281 (0.175)	0.289 (0.182)	0.241 (0.177)	0.300 (0.184)
Prob of Work - Jr HS	0.106 (0.386)	-0.035 (0.338)	0.130 (0.390)	-0.004 (0.342)
Prob of Work - Sr HS	0.021 (0.511)	-0.225 (0.451)	-0.081 (0.518)	-0.229 (0.457)
Prob of Work - College	0.670 (0.415)	-0.222 (0.469)	0.621 (0.422)	-0.161 (0.474)
Var of Log Earn - Jr HS	4.534 (6.640)	9.224 (8.055)	3.425 (6.684)	9.409 (8.185)
Var of Log Earn - Sr HS	13.924 (9.792)	-8.696 (8.385)	13.216 (9.804)	-9.235 (8.511)
Var of Log Earn - College HS	5.521 (11.310)	4.709 (11.067)	6.438 (11.541)	5.077 (11.144)
Fraction of At Least Sen HS Educ Men (Women)	0.697 (1.148)	-0.742 (1.428)	0.274 (1.180)	0.084 (1.453)
Observations	44	2332	2332	2332
Censored Obs		955	955	955

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects (fully interacted with gender dummies).

## APPENDIX: NOT FOR PUBLICATION

### Detailed Discussion of the Descriptive Statistics

Tables 1 and 2 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.

Around 79% of boys with a junior high school degree decide to enrol in senior high school compared to 83% of girls. In terms of college enrolment 29% of male high school graduates enrol in college compared to 32% of girls. Mexico has reached the stage where boys’ and girls’ educational attainment is very similar (or even higher in the case of high school enrolment), while for example in the US girls have overtaken boys in terms of both high school and college attainment (see Goldin, Katz, and Kuziemko (2006)). Comparing the samples of mother and youth respondents, enrolment rates are significantly lower for the youth sample in the case of the old cohort. For example, enrolment rates for girls are 24% for the youth sample compared to 40% of the mother sample (24% versus 33% for boys), which can be explained by the fact that youths who enrol in college are less likely to be at home when the interviewer arrives to conduct the survey (and they are thus less likely to be in the youth sample). In the case of the young cohort, there are no significant differences in enrolment rates between the samples of mother and youth respondents.

The young cohort is around 15.5 years of age, boys are slightly older than girls. There are no significant differences in age between mother and youth samples. The older cohort is around 18.5 years of age. Again boys are slightly older than girls, and youths in the two samples are the same age.

In our analysis we make use of data on GPA (grade point average between 0 and 100) of junior high school as a proxy for academic achievement. Tables 1 and 2 show that girls have a significantly higher GPA than boys for both junior and senior high school graduates. This is consistent with empirical evidence for many countries that girls outperform boys at school (see Goldin, Katz, and Kuziemko (2006)). Comparing Tables 1 and 2, we notice that individuals of the older cohort have a slightly higher GPA than those of the young cohort. This most likely reflects the fact that the senior high school graduates are a self-selected sample compared to the junior high school graduates who might or might not attend and finish senior high school. In the case of the young cohort, there are significant (on 10%) but small differences between the sample of mother and youth respondents for girls, with higher GPA for the mother sample (82.8 versus 82.1). For the old cohort, there are significant but small differences between mother and youth sample, this time only for boys, again with higher GPA for the mother sample (81.8 versus 81).

To control for parental education as one of the most important determinants of children’s schooling choices, we use information on parents’ years of completed schooling in the form of four education dummies for both mother and father (unless the household is single headed): for primary education, junior and senior high school and university. For the young cohort, about 70% of mothers

and fathers have only some primary education, while around 25% have attended junior high school. Fathers' education is slightly higher than that of mothers: about 6% have attended senior high school (3% for mothers) and 1% have some university education (0.7% for mothers). For the old cohort, parents are slightly less educated (75-80% primary and 17-18% secondary education). There are few and only small differences between mother and youth sample: for the young cohort, mothers in the mother sample are slightly less educated in the case of boys. For the old cohort, mothers in the mother sample are slightly less educated in the case of girls.

We create three per capita income categories, where the thresholds are equal to twice and four times the minimum wage.<sup>19</sup> For the young cohort, about 40% of the sample is in the lowest income category (that is yearly per capita income is below 5000 pesos) and thus relatively poor, reflecting the fact that our sample only consists of Oportunidades families. 35% are in the second highest category (5000 to 10000 pesos) and the remaining 25% in the top income category. The old cohort is slightly poorer with about 55% in the lowest income category. Differences between mother and youth sample are only significant in one instance for the young cohort, that is for girls the mother sample is poorer –consistent with lower parental education (see above).

The same pattern for young and old cohort can be found in terms of father's occupation: For the young cohort, 30% of fathers are unskilled workers, another 50% employees, around 18% are self-employed, 1.5% are family workers and less than 0.9% are employers. For the old cohort the numbers are 43%, 28%, 24%, 3.5% and 0.8%, respectively. Thus fathers in the youth sample are significantly less likely to be unskilled workers or self-employed and more likely to be employees. Comparing mother and youth samples, fathers in the youth sample are significantly more likely to be employees (and less likely to be self-employed or unskilled worker) than in the mother sample.

Only a minority of mothers are in the labor force, 30% for the young cohort and 20% for the old cohort. There are no significant differences between boys and girls, nor between the different samples of mother and youth respondents.

Part of the analysis that follows is conducted on the subsample of households where the father is not present because the mother is single, separated, divorced or widowed. In our sample about 22% (15%) of students of the young (old) cohort live in a household without father. There are no significant differences between boys and girls. Differences between mother and youth sample are only significant for the old cohort in the case of boys, where boys in the youth sample are more likely to live without a father. This can be explained by boys being more likely to stay at home, if there is no father in the household.

Another potentially important factor in schooling decisions is the number of siblings (in addition to the resource effect which we aim to capture by controlling for parental income divided by family size). For the young cohort, male and female youths have around 1.4 brothers and 1.4 sisters. In the case of the old cohort, male youths have around 1.44 brothers, female youths have significantly

---

<sup>19</sup>Per capita parental income is constructed using parental income –such as parents' labour earnings, other income sources such as rent, profits from a business, pension income etc and remittances– divided by family size. Median yearly per capita income is 6066 pesos (approximately 606 US\$). We add the income measures in the form of dummies to allow –in a flexible way– for nonlinear effects of income. The reason for the chosen income thresholds is their approximate correspondence with eligibility requirements for receiving fellowships (even though they are quantitatively not very important, see Kaufmann (2009)).

less brothers (1.36), while male youths have less sisters than female youths (1.26 versus 1.46). This pattern is similar for the young cohort but not significant. Differences between the youth sample and mother sample are only significant for the old cohort: Boys and girls in the mother sample have significantly more brothers (on 10%).

Public schools in Mexico do not charge tuition fees, while public universities charge for enrolment and tuition. In addition, in the case of university education, youths are much less likely to have a university close to their locality of residence compared to senior high schools. Therefore direct costs will be particularly important for the decision to attend college. We proxy for costs of living using the distance from the locality of residence of the youth to the closest university and for tuition costs in nearby universities (for the data sources and the exact definition of the variables, see Kaufmann (2009)). Around 50% of the youths live less than 20 km from the closest university (a distance which might allow daily commute), while around 25% live either between 20 and 40 km or more than 40 km from the closest university. There are no significant differences between boys and girls or mother and youth sample. Concerning tuition costs, about 40% of youths face tuition costs of more than 750 pesos, which corresponds to more than 15% of per capita median income and thus implies a substantial burden in particular for poor families. Tuition costs are significantly higher in the case of the youth sample compared to the mother sample, because youths in the youth sample live in larger cities with universities that charge higher fees.

### **Data Source for Marriage Markets Returns.**

To proxy for the returns that education might have in the marriage market, we measure the ratio of unmarried men to women in the locality of residence of the youth in the sample. For the young cohort we construct the ratio of unmarried men to women who are in a similar age range as the youth, that is age 15 to 30 in the locality of residence, for the old cohort we use the ratio of unmarried men to women in the age range of 18 to 35.

To construct those proxies, we make use of Census data of the year 2000, for which information is available at the locality level.<sup>20</sup> Having information at the locality level is important, because we believe that our "marriage market return" proxy works best in terms of capturing people's knowledge about availability of partners in the case of a small reference area (marriage "market") and we believe that the locality of residence is the correct reference area for most of the youths.

Unfortunately, Census information is only available for part of the localities in which the youths of our sample live. For the sample of junior high school graduates we lose around 30% of our observations. For the sample of senior high school graduates, who live in significantly smaller localities than the junior high school graduates (see descriptive statistics in the next section), we lose close to 60% of observations. For this reason we only use this significantly smaller sample for our analysis of marriage market returns, while we use the full sample for the first set of results on intra-household decision-making and labor market returns.

---

<sup>20</sup>The 2000 Census, "XII Censo General de Poblacion y Vivienda", can be found on the website of the Mexican statistical institute, INEGI.



## Validity of the Instrument: Timing of the Interview

Table 11: What predicts the timing of the interview? Old Cohort

	Week 40 to 42		Week 45 to 46		Weekday	
	Boys	Girls	Boys	Girls	Boys	Girls
Locality Size 15 to 50k	0.098* (0.053)	0.239*** (0.045)	0.000 (0.054)	-0.095** (0.046)	0.014 (0.044)	-0.033 (0.037)
Locality Size above 50k	0.223*** (0.048)	0.093** (0.047)	-0.141*** (0.049)	-0.057 (0.048)	-0.006 (0.040)	0.046 (0.039)
Dist to Univ 20 to 40km	-0.105*** (0.032)	-0.080*** (0.029)	0.037 (0.032)	0.028 (0.029)	0.008 (0.026)	0.017 (0.024)
Dist to Univ above 40km	-0.041 (0.036)	-0.069** (0.034)	0.162*** (0.037)	0.136*** (0.035)	0.007 (0.030)	0.051* (0.029)
Municip with Univ (Tuit above median)	0.060 (0.041)	0.061 (0.037)	-0.233*** (0.042)	-0.183*** (0.038)	0.061* (0.034)	0.099*** (0.031)
Attend College	-0.018 (0.027)	0.003 (0.025)	0.025 (0.028)	0.003 (0.026)	0.037 (0.023)	0.019 (0.021)
Age	0.001 (0.008)	-0.013 (0.008)	-0.014 (0.009)	-0.006 (0.009)	0.012* (0.007)	0.007 (0.007)
Obese	0.028 (0.084)	-0.013 (0.065)	0.051 (0.086)	0.027 (0.066)	0.080 (0.070)	0.021 (0.054)
Number of Siblings	0.004 (0.007)	-0.006 (0.007)	-0.002 (0.007)	0.002 (0.007)	-0.002 (0.006)	0.002 (0.006)
GPA of Jr HS (0-100)	-0.001 (0.001)	0.002** (0.001)	-0.000 (0.001)	-0.002** (0.001)	-0.001 (0.001)	0.001 (0.001)
No Father in Household	0.023 (0.041)	0.009 (0.037)	0.008 (0.042)	-0.060 (0.037)	0.012 (0.034)	-0.002 (0.030)
Mother's Educ - Jr HS	0.045 (0.037)	0.010 (0.033)	-0.098*** (0.037)	-0.020 (0.034)	0.017 (0.030)	-0.047* (0.028)
Mother's Educ - Sr HS	0.072 (0.097)	-0.079 (0.069)	0.101 (0.100)	0.035 (0.071)	0.035 (0.081)	0.000 (0.057)
Father's Educ - Jr HS	0.036 (0.039)	-0.002 (0.036)	-0.016 (0.040)	-0.066* (0.037)	-0.013 (0.033)	-0.003 (0.030)
Father's Educ - Sr HS	0.138* (0.073)	-0.025 (0.068)	-0.079 (0.074)	-0.052 (0.069)	0.079 (0.061)	0.017 (0.056)
Per cap Income - below 5k	-0.009 (0.028)	-0.018 (0.028)	0.013 (0.029)	0.015 (0.028)	-0.001 (0.024)	-0.018 (0.023)
Per cap Income - above 10k	-0.009 (0.036)	0.003 (0.034)	0.037 (0.037)	-0.005 (0.034)	-0.018 (0.030)	0.011 (0.028)
Father's Occup - Employee	-0.017 (0.034)	0.042 (0.031)	0.047 (0.035)	-0.055* (0.032)	0.024 (0.028)	0.001 (0.026)
Father's Occup - Self-Empl	-0.016 (0.034)	0.032 (0.032)	-0.006 (0.035)	0.013 (0.033)	-0.037 (0.029)	-0.035 (0.027)
Father's Occup - Fam Worker	-0.028 (0.073)	0.035 (0.072)	-0.035 (0.075)	0.026 (0.074)	0.016 (0.061)	0.094 (0.060)
Mother in Labor Force	-0.068* (0.036)	-0.007 (0.031)	0.012 (0.037)	0.011 (0.032)	-0.012 (0.030)	-0.056** (0.026)
Observations	2829		2829		2829	
F-Statistic	5.009		7.024		2.911	
R-Squared	0.086		0.116		0.052	
Adjusted R-Squared	0.069		0.100		0.034	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 2 and state fixed effects (fully interacted with gender dummies).

Table 12: What predicts the timing of the interview? Young Cohort

	Week 41 to 42		Week 45 to 46		Weekday	
	Boys	Girls	Boys	Girls	Boys	Girls
Locality Size 15 to 50k	0.030 (0.041)	0.143*** (0.034)	-0.031 (0.027)	-0.106*** (0.023)	0.079*** (0.029)	0.024 (0.024)
Locality Size above 50k	0.182*** (0.036)	-0.018 (0.039)	-0.078*** (0.024)	-0.019 (0.026)	0.015 (0.025)	0.080*** (0.027)
Attend HS	0.029 (0.037)	0.062 (0.039)	0.013 (0.025)	-0.024 (0.026)	-0.040 (0.026)	-0.034 (0.027)
Age	-0.010 (0.012)	0.012 (0.012)	0.013* (0.008)	0.001 (0.008)	-0.011 (0.008)	-0.013 (0.008)
Obese	0.031 (0.098)	-0.026 (0.034)	0.047 (0.066)	0.001 (0.023)	-0.023 (0.069)	-0.023 (0.024)
GPA	0.000 (0.002)	-0.032 (0.089)	-0.002* (0.001)	0.039 (0.059)	0.002 (0.001)	0.026 (0.063)
N of Siblings	0.026*** (0.010)	0.015 (0.009)	-0.001 (0.007)	-0.003 (0.006)	-0.010 (0.007)	0.003 (0.006)
No Father in HH	0.004 (0.045)	-0.004*** (0.002)	-0.018 (0.030)	0.000 (0.001)	-0.037 (0.032)	0.001 (0.001)
Mother Jr HS	0.038 (0.038)	0.009 (0.043)	-0.021 (0.025)	0.037 (0.029)	0.000 (0.027)	-0.005 (0.030)
Mother Sr HS	-0.052 (0.095)	-0.005 (0.036)	-0.031 (0.064)	-0.006 (0.024)	-0.095 (0.067)	-0.032 (0.025)
Father Jr HS	-0.000 (0.044)	0.178** (0.091)	-0.014 (0.030)	-0.095 (0.061)	-0.013 (0.031)	-0.038 (0.064)
Father Sr HS	0.021 (0.079)	0.016 (0.043)	-0.058 (0.053)	-0.027 (0.028)	-0.026 (0.056)	-0.032 (0.030)
Per cap Inc below 5k	-0.075* (0.038)	-0.005 (0.078)	0.027 (0.026)	0.003 (0.052)	0.009 (0.027)	0.026 (0.055)
Per cap Inc above 10k	0.048 (0.040)	0.039 (0.037)	0.026 (0.026)	0.034 (0.025)	0.029 (0.028)	-0.026 (0.026)
Father Employee	0.012 (0.037)	-0.043 (0.036)	-0.039 (0.025)	0.021 (0.024)	-0.021 (0.026)	-0.058** (0.026)
Father Self-Empl	-0.093* (0.049)	0.049 (0.036)	0.030 (0.033)	-0.020 (0.024)	-0.014 (0.034)	0.038 (0.026)
Father Fam Worker	0.183 (0.137)	-0.008 (0.049)	-0.047 (0.092)	0.042 (0.032)	-0.130 (0.097)	0.069** (0.034)
Mother in Labor Force	-0.042 (0.036)	-0.128 (0.137)	0.058** (0.024)	-0.024 (0.092)	-0.027 (0.025)	0.007 (0.097)
Observations	2364		2364		2364	
F-Statistic	4.051		5.615		4.267	
R-Squared	0.074		0.100		0.078	
Adjusted R-Squared	0.056		0.082		0.060	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects (fully interacted with gender dummies).

## Validity Checks: Expectations

Table 13: Subjective Expectations: Missing and Inconsistent Answers

	Boys			Girls		
	Youth Mean (Std Dev)	Mother Mean (Std Dev)	Difference Mean (P-Val)	Youth Mean (Std Dev)	Mother Mean (Std Dev)	Difference Mean (P-Val)
<b>Young Cohort</b>						
Missing	0.0062 (0.0788)	0.0132 (0.1144)	-0.0070 (0.1750)	0.0116 (0.1069)	0.0048 (0.0692)	0.0067 (0.1630)
Inconsistent	0.1973 (0.3982)	0.2036 (0.4030)	-0.0064 (0.7670)	0.1912 (0.3934)	0.2035 (0.4029)	-0.0123 (0.5460)
<b>Old Cohort</b>						
Missing	0.0030 (0.0544)	0.0068 (0.0824)	-0.0039 (0.2910)	0.0111 (0.1048)	0.0058 (0.0759)	0.0053 (0.2240)
Inconsistent	0.1346 (0.3416)	0.1547 (0.3618)	-0.0201 (0.2660)	0.1254 (0.3314)	0.1491 (0.3564)	-0.0237 (0.1480)

Table 14: Subjective Expectations: Missing and Inconsistent Answers Conditional on Schooling Decision

	Youth Respondent			Mother Respondent		
	Continues School No Mean (Std Dev)	Yes Mean (Std Dev)	Difference Mean (P-Val)	Continues School No Mean (Std Dev)	Yes Mean (Std Dev)	Difference Mean (P-Val)
	<b>BOYS</b>					
<b>Young Cohort</b>						
Missing	0.0000 (0.0000)	0.0076 (0.0872)	-0.0076 (0.2880)	0.0136 (0.1162)	0.0131 (0.1140)	0.0005 (0.9650)
Inconsistent	0.1701 (0.3770)	0.2034 (0.4028)	-0.0333 (0.3600)	0.2517 (0.4355)	0.1882 (0.3913)	0.0635 (0.1170)
<b>Old Cohort</b>						
Missing	0.0039 (0.0626)	0.0000 (0.0000)	0.0039 (0.4180)	0.0068 (0.0821)	0.0069 (0.0830)	-0.0001 (0.9810)
Inconsistent	0.1297 (0.3363)	0.1497 (0.3579)	-0.0200 (0.5110)	0.1508 (0.3582)	0.1626 (0.3697)	-0.0118 (0.6500)
<b>GIRLS</b>						
<b>Young Cohort</b>						
Missing	0.0261 (0.1601)	0.0088 (0.0932)	0.0174 (0.0650)	0.0000 (0.0000)	0.0058 (0.0761)	-0.0058 (0.4280)
Inconsistent	0.1830 (0.3879)	0.1927 (0.3947)	-0.0097 (0.7790)	0.2037 (0.4046)	0.2035 (0.4030)	0.0002 (0.9960)
<b>Old Cohort</b>						
Missing	0.0115 (0.1069)	0.0096 (0.0978)	0.0019 (0.8160)	0.0039 (0.0623)	0.0085 (0.0922)	-0.0047 (0.3760)
Inconsistent	0.1328 (0.3396)	0.1010 (0.3020)	0.0318 (0.2250)	0.1440 (0.3514)	0.1567 (0.3640)	-0.0127 (0.6060)

Table 15: Differences between Mothers' and Youths' Expectations: Junior High School Graduates

	Expected Earnings					
	Boys			Girls		
	Jun HS	Sen HS	Univ	Jun HS	Sen HS	Univ
Adolescent Respondent	-0.319 (0.198)	-0.115 (0.171)	-0.240 (0.168)	-0.675*** (0.216)	-0.516*** (0.198)	-0.316* (0.185)
Fract At Least Sen HS Educated	0.306 (0.220)	0.306 (0.190)	0.182 (0.187)	-0.254 (0.226)	-0.306 (0.207)	-0.171 (0.193)
Mother in Labor Force	-0.146*** (0.041)	-0.080** (0.036)	-0.024 (0.035)	-0.030 (0.042)	-0.045 (0.038)	-0.024 (0.036)
N of Brothers	-0.034** (0.015)	-0.023* (0.013)	-0.020 (0.013)	0.007 (0.015)	0.008 (0.013)	0.000 (0.013)
Father Missing, Mother Sep/Div/Widow	0.087* (0.051)	0.090** (0.044)	0.058 (0.043)	0.023 (0.053)	0.024 (0.048)	0.043 (0.045)
Mother's Educ - Jr HS	-0.025 (0.042)	-0.036 (0.036)	0.012 (0.035)	-0.074* (0.042)	-0.051 (0.038)	-0.032 (0.036)
Father's Educ - Jr HS	-0.059 (0.048)	0.009 (0.042)	-0.030 (0.041)	0.080 (0.050)	0.063 (0.046)	0.071* (0.043)
Per cap Income - 5 to 10k	0.008 (0.040)	0.030 (0.034)	0.041 (0.034)	0.049 (0.040)	0.089** (0.036)	0.097*** (0.034)
Per cap Income - more than 10k	0.021 (0.048)	0.012 (0.042)	0.036 (0.041)	0.092* (0.048)	0.085* (0.044)	0.113*** (0.041)
Father's Occup - Employer	0.051 (0.203)	0.338* (0.175)	0.391** (0.172)	-0.127 (0.219)	-0.332* (0.200)	-0.465** (0.187)
Father's Occup - Fam. Worker	-0.186 (0.151)	0.016 (0.131)	0.037 (0.129)	-0.150 (0.165)	-0.176 (0.151)	-0.263* (0.141)
Observations	1111	1111	1111	1253	1253	1253
R-Squared	0.143	0.166	0.102	.	.	0.052
Adjusted R-Squared	0.121	0.145	0.079	.	.	0.030

Notes: Table displays coefficients and standard errors in brackets. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ . Control variables are all variables in Table 1 and state fixed effects.

Table 16: Differences between Mothers' and Youths' Expectations: Senior High School Graduates

	Expected Earnings			
	Boys		Girls	
	Sen HS	Univ	Sen HS	Univ
Adolescent Respondent	-0.171 (0.198)	-0.060 (0.194)	-0.263* (0.156)	-0.132 (0.155)
Fract with Some Coll	0.128 (0.278)	-0.096 (0.274)	0.013 (0.300)	-0.369 (0.299)
Age	-0.024** (0.011)	-0.036*** (0.011)	-0.001 (0.011)	-0.003 (0.011)
GPA of Junior HS (0-100)	0.001 (0.002)	0.003* (0.002)	0.003*** (0.001)	0.005*** (0.001)
Number of Female Siblings	0.024** (0.011)	0.023** (0.011)	-0.005 (0.010)	0.001 (0.009)
Father Missing, Mother Sep/Div/Widow	0.033 (0.049)	0.016 (0.048)	-0.066* (0.038)	-0.054 (0.038)
Mother's Educ - Jr HS	-0.003 (0.038)	-0.001 (0.037)	-0.104*** (0.035)	-0.080** (0.035)
Father's Educ - Univ	-0.130 (0.150)	0.019 (0.147)	0.251* (0.148)	0.182 (0.147)
Per cap Income - 5 to 10k	0.064** (0.032)	0.006 (0.031)	0.034 (0.029)	0.011 (0.029)
Per cap Income - more than 10k	0.083** (0.037)	0.032 (0.037)	0.041 (0.036)	0.017 (0.036)
Father's Occup - Employee	-0.055 (0.039)	-0.004 (0.039)	-0.048 (0.034)	-0.078** (0.034)
Father's Occup - Fam. Worker	-0.075 (0.078)	-0.105 (0.076)	-0.120 (0.075)	-0.159** (0.074)
Observations	1320	1320	1509	1509
R-Squared	0.120	0.103	0.098	0.066
Adjusted R-Squared	0.101	0.084	0.080	0.048

Notes: Table displays coefficients and standard errors in brackets. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ . Control variables are all variables in Table 2 and state fixed effects.

## Descriptive Statistics: Earnings Expectations and Census Earnings

Concerning the comparison of expectations to Census earnings, we are particularly interested in the pattern for the different respondents, that is for boys and girls and for youth and mother respondents. We use Census data of the year 2000 and compute average log earnings (by gender and by schooling degree) in the municipality of residence for each youth in the sample. While the expectations questions refer to earnings at age 25, we use earnings of individuals who are between 25 and 30 years old to get a sufficiently large sample size for each municipality.

While the comparison between the subjective expectations data and the Census data is certainly informative, it is not enough to check whether the subjective expectations data are roughly of the same order of magnitude, a direct comparison and a formal test of equality between the two would be misleading. There are many reasons why the Census data and the subjective expectations would be different. First, the former refer to a specific year (2000) and are therefore affected by specific aggregate shocks that might have been relevant in that year. To test some version of Rational expectations, one would need several years of realizations to average out aggregate shocks. Second,

the data refer to individuals who were between 25 and 30 in 2000 and therefore belong to a different cohort from the individuals whose expectations were elicited in our survey (who are around 15 or 18 in 2005, and who were asked about their expected earnings when they are 25). Finally, the Census data report realizations for individuals who self-selected into a specific education level and do not contain “counterfactual” earnings, which are instead elicited in the expectations questions. Similarly, earnings realizations are for individuals who self-selected into the labor force.

From these considerations, it should be clear that to establish whether the elicited expectations are ‘rational’ is probably impossible. And, in a sense, it is not too important: for modeling education choices what matters is whether the elicited expectations reflect the subjective expectations that people base their decisions on.

Table 17 presents mean expected earnings based on the distribution of earnings and Census earnings for different schooling degrees. We correct Census earnings for selection into the labor force (using the standard Heckman selection correction and relying on variables proposed in the literature, that is nonlabor income, marital status and number of children, see, e.g. Heckman and Sedlacek (1985), Mulligan and Rubinstein (2008) and so forth).

Mean earnings are presented separately for boys and girls and mother and youth respondent. To correct for sample selection we present estimates that are based on the Heckman selection correction (again using the timing of the interview as an exclusion restriction). Estimates for the old (young) cohort are in the top (bottom) part of the table.

For the young cohort, all respondents expect lower junior high school earnings than observed in the Census. Youth respondents expect slightly lower senior high school earnings and similar college earnings. Mothers in general expect higher earnings and thus their expectations are closer to actually observed senior high school earnings but higher than currently observed college earnings. The old cohort expects similar high school earnings and higher college earnings. This pattern is consistent with a recent trend of a decrease of junior high school earnings in real terms, stagnating senior high school earnings and a rise in college earnings.

Girls responding themselves expect lower earnings than boys. The expected gender gap is similar to the one observed in Census earnings in 2000 with the exception of junior high school earnings, for which the actual gender gap is larger than expected. Also mothers expect their daughters to earn less than their sons (with one exception for the young cohort for senior high school earnings).

In terms of returns, girls’ and boys’ expectations are similar for returns to high school, while girls expect higher returns to college. According to current Census earnings, returns are indeed larger for girls for both schooling degrees. While girls’ return expectations are slightly higher than observed, boys’ expectations are much larger than observed returns.

Table 17: Expected Earnings of Mothers and Youths and Realized Census Earnings (With Heckman Selection Correction)

Respondent:	Boys		Girls	
	Youth Mean	Mother Mean	Youth Mean	Mother Mean
<b>Old Cohort</b>				
Exp Log Earnings				
- Senior HS	7.829 (0.201)	7.746 (0.151)	7.377 (0.142)	7.371 (0.134)
- College	8.506 (0.200)	8.406 (0.151)	8.254 (0.137)	8.206 (0.129)
Log Census Earnings				
- Senior HS	7.809 (0.135)	7.804 (0.158)	7.695 (0.135)	7.347 (0.158)
- College	8.425 (0.167)	8.428 (0.139)	8.104 (0.104)	8.257 (0.146)
Observations	507	605	585	542
<b>Young Cohort</b>				
Exp Log Earnings				
- Junior HS	6.982 (0.108)	7.089 (0.121)	6.539 (0.138)	7.133 (0.136)
- Senior HS	7.580 (0.095)	7.522 (0.108)	7.242 (0.122)	7.641 (0.127)
- College	8.132 (0.090)	8.200 (0.101)	7.898 (0.117)	7.917 (0.146)
Log Census Earnings				
- Junior HS	7.698 (0.073)	7.461 (0.095)	7.129 (0.083)	7.318 (0.115)
- Senior HS	7.980 (0.074)	7.857 (0.071)	7.683 (0.061)	7.450 (0.096)
- College	8.283 (0.110)	7.991 (0.092)	8.237 (0.059)	7.820 (0.126)
Observations	599	437	631	424

Notes: In this table we use Census data of the year 2000 and compute average log earnings (by gender and by schooling degree) in the municipality of residence for each youth in the sample. For this reason average Census earnings can differ between youth and mother sample, because the individuals in the two samples come from different municipalities. This table differs in the number of observations compared to the main tables, because for some municipalities in which junior and senior high school graduates live, we do not have Census data. Therefore we drop observations with missing information on actual earnings in this table (to compare expectations with actual earnings for the same municipalities), while we keep those observations for the main analysis.

## Main Results: Displaying Controls

Table 18: College Attendance Choice: Youth and Mother Expectations

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
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)
Age	-0.152*** (0.047)	-0.143*** (0.047)	-0.018 (0.035)	-0.057 (0.037)
Obese	-0.439 (0.579)	-0.713* (0.389)	-0.287 (0.331)	-0.338 (0.298)
GPA of Jr HS (0-100)	0.027*** (0.009)	0.023*** (0.007)	0.019*** (0.007)	0.009** (0.004)
Mother's Educ - Jr HS	0.196 (0.171)	-0.056 (0.159)	0.337** (0.149)	0.327** (0.147)
Mother's Educ - Sr HS	0.368 (0.467)	0.508 (0.325)	0.952** (0.398)	0.664** (0.290)
Father's Educ - Jr HS	0.066 (0.194)	0.309* (0.166)	0.231 (0.158)	-0.033 (0.155)
Father's Educ - Sr HS	0.524 (0.321)	0.369 (0.277)	-0.120 (0.316)	0.539* (0.306)
Per cap Income - below 5k	0.079 (0.147)	-0.146 (0.133)	-0.224* (0.121)	-0.147 (0.130)
Per cap Income - above 10k	0.137 (0.179)	0.399** (0.155)	0.265* (0.148)	0.583*** (0.147)
Father's Occup - Self-Empl	-0.338* (0.192)	-0.118 (0.158)	-0.234* (0.142)	0.027 (0.138)
Father's Occup - Fam Worker	0.169 (0.340)	-0.472 (0.417)	-0.773* (0.410)	-0.189 (0.327)
Mother in Labor Force	0.110 (0.190)	0.175 (0.145)	0.259* (0.147)	0.052 (0.138)
Dist to Univ 20 to 40km	-0.312* (0.167)	-0.258* (0.141)	0.033 (0.138)	0.188 (0.127)
Dist to Univ above 40km	-0.117 (0.181)	-0.537*** (0.169)	0.169 (0.155)	0.210 (0.153)
Munip with Univ (Tuit above Median)	-0.027 (0.198)	-0.742*** (0.186)	-0.075 (0.174)	-0.243 (0.174)
Observations	2829		2829	
Censored Obs	1468		1361	
Log Likelihood	-2496.644		-2704.032	
Sample Sel: Corr of Errors	0.256		-0.151	
P-val: LR Test of Indep Eqns	0.473		0.670	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 2 and state fixed effects (fully interacted with gender dummies).



Table 19: College Attendance Choice: No Father in Household (Mother Single/Sep/Div)

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
Expected Return College * Father in Household	0.274 (0.196)	0.071 (0.157)	-0.022 (0.157)	0.015 (0.158)
Prob of Work Sr HS * Father in Household	-0.819 (0.551)	-0.000 (0.398)	-0.543 (0.397)	0.028 (0.405)
Prob of Work College * Father in Household	0.075 (0.507)	-0.203 (0.389)	0.450 (0.393)	0.280 (0.396)
Var of Log Earn Sr HS * Father in Household	-8.182 (10.536)	-8.916 (8.183)	14.041 (10.153)	-0.453 (8.862)
Var of Log Earn College * Father in Household	-4.958 (13.816)	8.184 (10.720)	-22.514* (12.521)	10.707 (13.294)
Expected Return College * No Father in Household	1.318*** (0.409)	0.217 (0.318)	-0.389 (0.425)	0.444 (0.372)
Prob of Work Sr HS * No Father in Household	-0.000 (0.918)	0.945 (0.951)	-1.219 (1.047)	0.848 (0.842)
Prob of Work College * No Father in Household	-0.502 (1.100)	-0.683 (1.095)	0.401 (1.271)	2.280** (1.066)
Var of Log Earn Sr HS * No Father in Household	0.404 (24.533)	-36.804 (27.107)	-2.860 (22.830)	14.687 (20.441)
Var of Log Earn College * No Father in Household	-17.356 (35.725)	-1.807 (17.793)	19.396 (36.115)	8.706 (24.441)
Age	-0.149*** (0.048)	-0.142*** (0.048)	-0.038 (0.035)	-0.034 (0.036)
Obese	-0.384 (0.588)	-0.662* (0.394)	-0.306 (0.331)	-0.350 (0.299)
GPA of Jr HS (0-100)	0.027*** (0.009)	0.025*** (0.007)	0.017** (0.007)	0.010** (0.004)
No Father in Household	-0.303 (0.914)	-0.312 (0.707)	0.700 (0.841)	-2.652*** (0.896)
Mother's Educ - Jr HS	0.224 (0.172)	-0.040 (0.161)	0.320** (0.150)	0.334** (0.147)
Mother's Educ - Sr HS	0.308 (0.470)	0.475 (0.332)	0.941** (0.392)	0.685** (0.288)
Father's Educ - Jr HS	0.099 (0.198)	0.341** (0.169)	0.218 (0.158)	-0.029 (0.155)
Father's Educ - Sr HS	0.560* (0.324)	0.414 (0.282)	-0.093 (0.314)	0.470 (0.304)
Per cap Income - below 5k	0.117 (0.150)	-0.144 (0.135)	-0.212* (0.121)	-0.118 (0.130)
Per cap Income - above 10k	0.162 (0.184)	0.389** (0.152)	0.271* (0.148)	0.559*** (0.147)
Father's Occup - Fam Worker	0.184 (0.346)	-0.472 (0.424)	-0.732* (0.406)	-0.175 (0.321)
Dist to Univ 20 to 40km	-0.393** (0.161)	-0.270** (0.134)	0.040 (0.132)	0.132 (0.122)
Dist to Univ above 40km	-0.161 (0.180)	-0.534*** (0.164)	0.175 (0.151)	0.154 (0.149)
Tuition more than 750 Pesos	-0.026 (0.201)	-0.737*** (0.183)	-0.096 (0.172)	-0.217 (0.173)
Observations	2829		2829	
Censored Obs	1468		1361	
Log Likelihood	-2498.693		-2712.876	
Sample Sel: Corr of Errors	0.124		-0.228	
P-val: LR Test of Indep Eqns	0.719		0.511	

Notes: Table displays coefficients and standard errors in brackets. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ . Excl. categories: not obese, father in household, mother's and father's education primary or less, father's occupation worker, per capita income between 5000 and 10000 pesos, distance to university less than 20km, tuition costs less than 750 pesos. All specifications include state dummies.

Table 20: High School Attendance Choice: Youth and Mother Expectations

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
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)
Number of Brothers	-0.160*** (0.053)	-0.080* (0.047)	0.024 (0.058)	0.031 (0.066)
Number of Sisters	-0.117** (0.051)	-0.047 (0.048)	-0.101* (0.052)	0.053 (0.059)
GPA of Jr HS (0-100)	0.027*** (0.006)	0.033*** (0.006)	0.025*** (0.007)	0.030*** (0.007)
No Father in Household	-0.127 (0.174)	0.526*** (0.184)	0.173 (0.209)	-0.051 (0.217)
Mother's Educ - Jr HS	0.245 (0.167)	0.414** (0.172)	0.374* (0.195)	0.534** (0.224)
Father's Educ - Jr HS	0.071 (0.190)	0.173 (0.179)	0.418* (0.223)	-0.046 (0.225)
Per cap Income - below 5k	0.164 (0.151)	-0.052 (0.146)	-0.208 (0.173)	-0.506*** (0.184)
Father's Occup - Self-Empl	0.250 (0.216)	0.410** (0.201)	-0.161 (0.208)	-0.082 (0.263)
Father's Occup - Fam Worker	5.551 (1.2e+04)	-0.137 (0.684)	0.207 (0.491)	-0.986** (0.498)
Observations	2364	2364		
Censored Obs	967	1397		
Log Likelihood	-1979.944	-1843.957		
Sample Sel: Corr of Errors	-0.586	0.424		
P-val: LR Test of Indep Eqns	0.036	0.185		

Notes: Table displays coefficients and standard errors in brackets. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ . Excl. categories: size of locality less than 15k, not obese, father in household, mother's and father's education primary or less, father's occupation worker, per capita income between 5000 and 10000 pesos. All specifications include state dummies.

Table 21: High School Attendance Choice: No Father in Household (Mother Single/Sep/Div)

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
Expected Return Sr HS * Father in Household	-0.348 (0.241)	0.092 (0.216)	-0.009 (0.298)	-0.165 (0.334)
Expected Return College * Father in Household	0.023 (0.192)	0.197 (0.210)	-0.200 (0.284)	-0.461 (0.287)
Prob of Work Jr HS * Father in Household	0.499 (0.481)	0.070 (0.404)	1.138** (0.540)	-1.351** (0.617)
Prob of Work Sr HS * Father in Household	-1.038 (0.670)	0.002 (0.555)	-1.235 (0.758)	0.484 (0.873)
Prob of Work College * Father in Household	0.749 (0.481)	-0.244 (0.475)	0.970 (0.609)	0.194 (0.714)
Var of Log Earn Jr HS * Father in Household	3.794 (8.004)	12.040 (9.889)	20.103* (11.552)	-3.968 (10.676)
Var of Log Earn Sr HS * Father in Household	11.935 (11.128)	-9.817 (9.492)	-11.611 (12.769)	-22.208** (11.103)
Var of Log Earn College * Father in Household	-2.851 (11.605)	9.979 (13.777)	-24.522 (17.163)	-2.572 (13.357)
Expected Return Sr HS * No Father in Household	0.500 (0.496)	-0.111 (0.461)	0.131 (0.643)	-0.384 (0.600)
Expected Return College * No Father in Household	0.907** (0.444)	0.280 (0.399)	0.371 (0.697)	-0.011 (0.511)
Prob of Work Jr HS * No Father in Household	-0.827 (0.796)	-0.534 (0.762)	-1.272 (1.279)	-0.026 (1.064)
Prob of Work Sr HS * No Father in Household	1.641* (0.930)	1.311 (1.036)	-2.095 (2.137)	-2.624* (1.418)
Prob of Work College * No Father in Household	0.312 (0.910)	-2.320* (1.348)	0.985 (1.870)	2.422** (1.174)
Var of Log Earn Jr HS * No Father in Household	-0.791 (21.235)	14.693 (21.792)	15.255 (20.122)	-19.453 (19.893)
Var of Log Earn Sr HS * No Father in Household	-1.201 (27.531)	15.146 (31.630)	36.600 (47.002)	-5.700 (31.236)
Var of Log Earn College * No Father in Household	92.069* (48.860)	-4.134 (24.867)	24.736 (40.142)	84.584 (61.050)
Number of Brothers	-0.156*** (0.055)	-0.095** (0.049)	0.010 (0.063)	0.037 (0.071)
Number of Sisters	-0.144*** (0.053)	-0.061 (0.049)	-0.122** (0.057)	0.048 (0.063)
GPA of Jr HS (0-100)	0.032*** (0.006)	0.029*** (0.006)	0.021*** (0.007)	0.035*** (0.007)
No Father in Household	-2.143** (0.852)	1.681 (1.054)	1.203 (1.378)	-0.818 (1.016)
Mother's Educ - Jr HS	0.290 (0.177)	0.453*** (0.175)	0.401* (0.208)	0.581** (0.242)
Father's Educ - Jr HS	0.101 (0.196)	0.155 (0.185)	0.435* (0.235)	-0.041 (0.245)
Per cap Income - below 5k	0.144 (0.157)	-0.079 (0.151)	-0.215 (0.181)	-0.559*** (0.194)
Father's Occup - Self-Empl	0.212 (0.231)	0.390* (0.208)	-0.300 (0.231)	-0.101 (0.294)
Father's Occup - Fam Worker	3.978 (101.636)	-0.144 (0.717)	0.292 (0.516)	-1.127** (0.505)
Mother in Labor Force	0.107 (0.149)	-0.244* (0.143)	0.256 (0.198)	-0.261 (0.200)
Locality Size above 50k	-0.098 (0.147)	-0.083 (0.146)	0.231 (0.201)	0.414* (0.234)
	xiv			
Observations	2364	2364		
Censored Obs	967	1397		
Log Likelihood	-1974.291	-1818.847		
Sample Sel: Corr of Errors	-0.448	0.139		
P-val: LR Test of Indep Eqns	0.089	0.719		

Table 22: College Attendance Choice: Labor and Marriage Market Returns

Dependent Variable	College Attendance Decision			
	Marriage Market Proxy 1		Marriage Market Proxy 2	
	Boys	Girls	Boys	Girls
Ratio of Unmarried Men to Women (Sr HS)	-0.102 (0.147)			
Ratio of Unmarried Women to Men (Sr HS)		0.223** (0.107)		
Ratio of Unmarried Men to Women			-0.5 (0.474)	
Ratio of Unmarried Women to Men				0.605** (0.294)
Expected Return - College	0.782*** (0.248)	0.343** (0.166)	0.764*** (0.251)	0.331** (0.164)
Prob of Work - Sr HS	0.091 (0.601)	0.400 (0.484)	0.113 (0.599)	0.400 (0.483)
Prob of Work - College	-0.563 (0.632)	0.314 (0.590)	-0.48 (0.630)	0.309 (0.583)
Log Var of Earn - Sr HS	5.047 (12.408)	12.272 (10.633)	6.961 (12.489)	15.556 (10.454)
Log Var of Earn - College	-0.062 (19.671)	-7.841 (14.337)	-1.874 (19.576)	-7.019 (14.238)
Fraction of College Educated Men (Women)	1.440 (1.228)	3.491*** (1.225)	1.309 (1.223)	2.958** (1.222)
GPA of Jr HS (0-100)	0.050*** (0.010)	0.037*** (0.010)	0.056*** (0.011)	0.035*** (0.010)
Number of Siblings	0.047 (0.065)	-0.006 (0.050)	0.059 (0.064)	0.009 (0.049)
No Father in Household	0.552** (0.228)	-0.282 (0.183)	0.551** (0.231)	-0.282 (0.181)
Mother's Educ - Jr HS	0.143 (0.239)	0.0120 (0.203)	0.129 (0.239)	0.093 (0.197)
Mother's Educ - Sr HS	-4.651 (761.757)	0.883** (0.368)	-4.445 (641.853)	0.906** (0.365)
Per cap Income - below 5k	0.305 (0.228)	-0.261 (0.206)	0.310 (0.228)	-0.306 (0.200)
Per cap Income - above 10k	-0.055 (0.245)	0.424** (0.184)	-0.029 (0.245)	0.441** (0.183)
Observations		1237		1237
Censored Obs		564		564
Log Likelihood		-1151.164		-1152.450
Sample Sel: Corr of Errors (P-Val)		-0.174 (0.736)		-0.240 (0.625)

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Excl. categories: father in household, mother's education primary or less, per capita income between 5000 and 10000 pesos. A dummy for mother's education university is included but not displayed due to space constraints (insignificant). Perceptions of unemployment and earnings risk are included in all specifications but not displayed due to space constraints (insignificant).

Table 23: High School Attendance Choice: Labor and Marriage Market Returns

Dependent Variable	High School Attendance Decision			
	Marriage Market Proxy 1		Marriage Market Proxy 2	
	Boys	Girls	Boys	Girls
Ratio of Unmarried Men to Women (Jr HS)	-0.005 (0.107)			
Ratio of Unmarried Women to Men (Jr HS)		0.650** (0.313)		
Ratio of Unmarried Men to Women			-0.055 (0.434)	
Ratio of Unmarried Women to Men				0.553 (0.512)
Expected Return - Sr HS	-0.044 (0.208)	0.100 (0.193)	-0.042 (0.209)	0.150 (0.196)
Expected Return - College	0.281 (0.175)	0.303* (0.182)	0.241 (0.177)	0.300 (0.184)
Prob of Work - Jr HS	0.106 (0.386)	-0.035 (0.338)	0.130 (0.390)	-0.004 (0.342)
Prob of Work - Sr HS	0.021 (0.511)	-0.225 (0.451)	-0.081 (0.518)	-0.229 (0.457)
Prob of Work - College	0.670 (0.415)	-0.222 (0.469)	0.621 (0.422)	-0.161 (0.474)
Var of Log Earn - Jr HS	4.534 (6.640)	9.224 (8.055)	3.425 (6.684)	9.409 (8.185)
Var of Log Earn - Sr HS	13.924 (9.792)	-8.696 (8.385)	13.216 (9.804)	-9.235 (8.511)
Var of Log Earn - College HS	5.521 (11.310)	4.709 (11.067)	6.438 (11.541)	5.077 (11.144)
Fraction of At Least Sen HS Educ Men	0.697 (1.148)	-0.742 (1.428)	0.274 (1.180)	0.084 (1.453)
Number of Brothers	-0.175*** (0.055)	-0.081* (0.049)	-0.190*** (0.056)	-0.080 (0.049)
Number of Sisters	-0.116** (0.052)	-0.038 (0.050)	-0.125** (0.053)	-0.040 (0.051)
GPA of Jr HS (0-100)	0.028*** (0.007)	0.029*** (0.007)	0.023*** (0.007)	0.035*** (0.007)
No Father in Household	-0.129 (0.169)	0.399** (0.175)	-0.159 (0.170)	0.394** (0.176)
Mother's Educ - Jr HS	0.251 (0.171)	0.369** (0.175)	0.251 (0.172)	0.389** (0.176)
Father's Occup - Self-Empl	0.220 (0.224)	0.537** (0.218)	0.182 (0.227)	0.501** (0.219)
Observations	2332		2332	
Censored Obs	955		955	
Log Likelihood	-1930.423		-1932.389	
Sample Sel: Corr of Errors	-0.485		-0.420	
P-val: LR Test of Indep Eqns	0.069		0.117	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects.

## Main Results: Without Selection Correction

Table 24: College Attendance Choice: Youth and Mother Expectations

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
Expected Return - College	0.388** (0.171)	0.185 (0.135)	-0.007 (0.144)	0.199 (0.144)
Prob of Work - Sr HS	-0.525 (0.459)	0.266 (0.362)	-0.432 (0.370)	0.102 (0.355)
Prob of Work - College	0.119 (0.490)	0.072 (0.422)	0.192 (0.434)	1.105*** (0.403)
Var of Log Earn - Sr HS	-5.222 (9.317)	-8.652 (7.282)	13.993 (8.999)	5.509 (8.167)
Var of Log Earn - College	-7.255 (12.768)	7.842 (8.628)	-18.991* (10.867)	13.656 (11.800)
Age	-0.151*** (0.048)	-0.139*** (0.048)	-0.017 (0.035)	-0.054 (0.037)
Obese	-0.419 (0.590)	-0.735* (0.393)	-0.268 (0.330)	-0.346 (0.298)
GPA of Jr HS (0-100)	0.029*** (0.008)	0.024*** (0.007)	0.020*** (0.006)	0.009** (0.004)
Mother's Educ - Jr HS	0.200 (0.173)	-0.036 (0.159)	0.343** (0.149)	0.342** (0.143)
Mother's Educ - Sr HS	0.402 (0.472)	0.545* (0.324)	0.977** (0.392)	0.681** (0.287)
Father's Educ - Jr HS	0.064 (0.197)	0.322* (0.167)	0.228 (0.159)	-0.025 (0.155)
Father's Educ - Sr HS	0.512 (0.325)	0.370 (0.281)	-0.127 (0.317)	0.541* (0.307)
Per cap Income - below 5k	0.098 (0.147)	-0.162 (0.133)	-0.215* (0.120)	-0.161 (0.126)
Per cap Income - above 10k	0.151 (0.180)	0.423*** (0.151)	0.275* (0.146)	0.593*** (0.144)
Father's Occup - Self-Empl	-0.344* (0.194)	-0.102 (0.159)	-0.235* (0.142)	0.034 (0.138)
Father's Occup - Fam Worker	0.133 (0.342)	-0.489 (0.421)	-0.792* (0.408)	-0.206 (0.326)
Mother in Labor Force	0.127 (0.191)	0.183 (0.146)	0.269* (0.145)	0.051 (0.139)
Dist to Univ 20 to 40km	-0.330** (0.166)	-0.275** (0.139)	0.021 (0.136)	0.183 (0.128)
Dist to Univ above 40km	-0.115 (0.184)	-0.541*** (0.170)	0.164 (0.156)	0.215 (0.153)
Munic with Univ (Tuit above Median)	-0.013 (0.200)	-0.766*** (0.180)	-0.067 (0.174)	-0.244 (0.174)
Observations	1361		1468	
Log Likelihood	-667.513		-874.734	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 2 and state fixed effects (fully interacted with gender dummies).

Table 25: College Attendance Choice: No Father in Household (Mother Single/Sep/Div)

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
Expected Return College * Father in Household	0.274 (0.198)	0.070 (0.158)	-0.032 (0.159)	0.017 (0.161)
Prob of Work Sr HS * Father in Household	-0.823 (0.554)	-0.009 (0.399)	-0.530 (0.402)	0.029 (0.411)
Prob of Work College * Father in Household	0.076 (0.509)	-0.196 (0.390)	0.443 (0.399)	0.284 (0.402)
Var of Log Earn Sr HS * Father in Household	-8.326 (10.572)	-9.072 (8.210)	14.519 (10.242)	-0.475 (9.023)
Var of Log Earn College * Father in Household	-4.856 (13.882)	8.317 (10.755)	-23.172* (12.635)	10.774 (13.550)
Expected Return College * No Father in Household	1.326*** (0.409)	0.218 (0.320)	-0.390 (0.432)	0.445 (0.378)
Prob of Work Sr HS * No Father in Household	-0.014 (0.921)	0.942 (0.954)	-1.240 (1.061)	0.867 (0.855)
Prob of Work College * No Father in Household	-0.490 (1.103)	-0.677 (1.097)	0.413 (1.291)	2.343** (1.073)
Var of Log Earn Sr HS * No Father in Household	0.575 (24.629)	-36.987 (27.140)	-3.085 (23.141)	14.330 (20.856)
Var of Log Earn College * No Father in Household	-17.174 (35.848)	-2.065 (17.856)	18.565 (36.682)	9.610 (24.807)
Age	-0.148*** (0.048)	-0.139*** (0.047)	-0.037 (0.035)	-0.030 (0.036)
Obese	-0.376 (0.591)	-0.669* (0.395)	-0.283 (0.333)	-0.365 (0.301)
GPA of Jr HS (0-100)	0.028*** (0.008)	0.025*** (0.007)	0.019*** (0.006)	0.010** (0.004)
No Father in Household	-0.342 (0.911)	-0.317 (0.710)	0.652 (0.852)	-2.709*** (0.899)
Mother's Educ - Jr HS	0.224 (0.173)	-0.031 (0.159)	0.330** (0.150)	0.357** (0.143)
Mother's Educ - Sr HS	0.322 (0.470)	0.490 (0.330)	0.979** (0.387)	0.715** (0.285)
Father's Educ - Jr HS	0.100 (0.199)	0.347** (0.168)	0.215 (0.160)	-0.016 (0.156)
Father's Educ - Sr HS	0.552* (0.324)	0.411 (0.283)	-0.108 (0.317)	0.477 (0.307)
Per cap Income - above 10k	0.168 (0.184)	0.398*** (0.150)	0.286* (0.146)	0.575*** (0.144)
Father's Occup - Fam Worker	0.165 (0.343)	-0.476 (0.425)	-0.763* (0.405)	-0.199 (0.323)
Mother in Labor Force	0.070 (0.194)	0.190 (0.144)	0.247* (0.146)	0.078 (0.138)
Dist to Univ 20 to 40km	-0.398** (0.161)	-0.271** (0.134)	0.027 (0.132)	0.134 (0.124)
Dist to Univ above 40km	-0.157 (0.180)	-0.526*** (0.164)	0.175 (0.152)	0.173 (0.147)
Munic with Univ (Tuit above Median)	-0.022 (0.202)	-0.748*** (0.179)	-0.087 (0.173)	-0.224 (0.174)
Observations	1361		1468	
Log Likelihood	-661.135		-875.469	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 2 and state fixed effects (fully interacted with gender dummies).

Table 26: High School Attendance Choice: Youth and Mother Expectations

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
Expected Return - Sr HS	-0.072 (0.216)	0.158 (0.201)	0.245 (0.252)	-0.086 (0.289)
Expected Return - College	0.233 (0.179)	0.255 (0.186)	-0.017 (0.247)	-0.214 (0.248)
Prob of Work - Jr HS	0.126 (0.402)	0.021 (0.350)	0.544 (0.454)	-0.837* (0.488)
Prob of Work - Sr HS	-0.061 (0.535)	-0.160 (0.469)	-0.890 (0.655)	-0.062 (0.647)
Prob of Work - College	0.684 (0.431)	-0.221 (0.486)	0.903 (0.568)	1.104** (0.556)
Var of Log Earn - Jr HS	5.322 (6.938)	10.944 (8.514)	13.413 (8.355)	-7.887 (8.577)
Var of Log Earn - Sr HS	12.929 (10.060)	-10.983 (8.698)	-9.405 (11.348)	-17.637* (9.194)
Var of Log Earn - College HS	6.219 (11.747)	8.042 (11.468)	-10.653 (13.000)	7.635 (10.653)
Number of Brothers	-0.182*** (0.055)	-0.084* (0.050)	0.030 (0.061)	0.037 (0.069)
Number of Sisters	-0.117** (0.054)	-0.055 (0.051)	-0.103* (0.055)	0.054 (0.063)
GPA of Jr HS (0-100)	0.026*** (0.007)	0.033*** (0.007)	0.024*** (0.007)	0.030*** (0.007)
No Father in Household	-0.095 (0.174)	0.381** (0.180)	0.266 (0.202)	-0.267 (0.220)
Mother's Educ - Jr HS	0.272 (0.177)	0.438** (0.179)	0.371* (0.203)	0.573** (0.230)
Father's Educ - Jr HS	0.096 (0.201)	0.387 (0.388)	0.467** (0.232)	0.216 (0.537)
Father's Educ - Sr HS	0.598 (0.516)	-0.056 (0.155)	0.820 (0.549)	-0.539*** (0.185)
Father's Occup - Employee	0.094 (0.157)	0.380* (0.215)	0.188 (0.186)	-0.169 (0.278)
Father's Occup - Self-Empl	0.211 (0.233)	-0.235 (0.765)	-0.273 (0.209)	-1.107** (0.477)
Locality Size 15 to 50k	0.251 (0.186)	-0.126 (0.170)	0.397** (0.201)	0.115 (0.195)
Locality Size above 50k	-0.040 (0.149)	-0.047 (0.148)	0.316* (0.180)	0.487** (0.205)
Observations	1369		959	
Log Likelihood	-545.134		-400.017	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects (fully interacted with gender dummies).



Table 27: High School Attendance Choice: No Father in Household (Mother Single/Sep/Div)

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
Expected Return Sr HS * Father in Household	-0.352 (0.251)	0.105 (0.225)	0.056 (0.295)	-0.158 (0.339)
Expected Return College * Father in Household	0.027 (0.201)	0.210 (0.219)	-0.181 (0.283)	-0.443 (0.287)
Prob of Work Jr HS * Father in Household	0.527 (0.501)	0.069 (0.421)	1.103** (0.541)	-1.332** (0.617)
Prob of Work Sr HS * Father in Household	-1.096 (0.696)	0.034 (0.580)	-1.175 (0.759)	0.508 (0.876)
Prob of Work College * Father in Household	0.763 (0.500)	-0.290 (0.496)	0.878 (0.608)	0.190 (0.718)
Var of Log Earn Jr HS * Father in Household	4.521 (8.351)	12.217 (10.376)	18.805* (11.000)	-3.903 (10.709)
Var of Log Earn Sr HS * Father in Household	12.236 (11.596)	-10.520 (9.911)	-10.951 (12.739)	-21.756** (10.940)
Var of Log Earn College * Father in Household	-3.191 (12.021)	9.184 (14.196)	-25.349 (17.250)	-2.960 (13.377)
Expected Return Sr HS * No Father in Household	0.507 (0.521)	-0.096 (0.478)	0.134 (0.641)	-0.356 (0.604)
Expected Return College * No Father in Household	0.993** (0.462)	0.273 (0.412)	0.325 (0.686)	-0.008 (0.507)
Prob of Work Jr HS * No Father in Household	-0.837 (0.840)	-0.486 (0.789)	-1.258 (1.296)	-0.052 (1.071)
Prob of Work Sr HS * No Father in Household	1.807* (0.970)	1.407 (1.085)	-2.110 (2.179)	-2.490* (1.422)
Prob of Work College * No Father in Household	0.272 (0.957)	-2.459* (1.408)	1.094 (1.858)	2.189* (1.170)
Var of Log Earn Jr HS * No Father in Household	1.587 (22.776)	18.352 (22.839)	18.536 (20.071)	-20.750 (19.959)
Var of Log Earn Sr HS * No Father in Household	0.218 (28.860)	14.394 (33.035)	34.070 (47.007)	-3.239 (31.410)
Var of Log Earn College * No Father in Household	93.789* (50.343)	-4.568 (26.137)	28.031 (40.436)	85.065 (61.500)
Number of Brothers	-0.169*** (0.056)	-0.098* (0.050)	0.008 (0.063)	0.036 (0.070)
Number of Sisters	-0.141** (0.055)	-0.068 (0.051)	-0.115** (0.057)	0.049 (0.064)
GPA of Jr HS (0-100)	0.031*** (0.006)	0.029*** (0.006)	0.021*** (0.007)	0.033*** (0.007)
No Father in Household	-2.255** (0.888)	1.538 (1.090)	1.207 (1.369)	-0.810 (1.025)
Mother's Educ - Jr HS	0.311* (0.183)	0.465*** (0.179)	0.407* (0.209)	0.589** (0.241)
Father's Educ - Jr HS	0.122 (0.202)	0.143 (0.190)	0.438* (0.235)	-0.039 (0.244)
Per cap Income - below 5k	0.115 (0.163)	-0.079 (0.156)	-0.241 (0.175)	-0.564*** (0.192)
Father's Occup - Self-Empl	0.175 (0.242)	0.366* (0.217)	-0.356* (0.215)	-0.076 (0.286)
Father's Occup - Fam Worker		-0.209 (0.766)		-1.128** (0.479)
Locality Size 15 to 50k	0.310 (0.191)	-0.135 (0.171)	0.345* (0.209)	0.052 (0.201)
Locality Size above 50k	-0.045 (0.151)	-0.047 (0.149)	0.302 (0.185)	0.419* (0.216)
	XX			
Observations		1369		959
Log Likelihood		-537.046		-381.843

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Included controls are all variables in Table 1 and state fixed effects (fully interacted with gender dummies).

## Tests Ex-Post Rationalization

Table 28: College Attendance Choice: Rationalization of Choices

Distrib of Resid of	Youth Resp		Mother Resp	
	Boys	Girls	Boys	Girls
	P-Val of KS-Test		P-Val of KS-Test	
Exp Log Earnings				
- Senior HS	0.611	0.777	0.459	0.057
- College	0.339	0.062	0.007	0.097
Exp Return				
- College	0.797	0.458	0.002	0.010
Prob of Work				
- Senior HS	0.140	0.005	0.374	0.093
- College	0.889	0.033	0.483	0.458
Var of Log Earnings				
- Senior HS	0.331	0.532	0.002	0.012
- College	0.028	0.824	0.007	0.224
Observations				
(Sen HS Grads/Grade 12)	583/212	778/257	737/298	731/375

Notes: Table displays the p-values of Kolmogorov-Smirnov tests of equality of distributions. The null hypothesis is that the cross-sectional distribution of -for example- expected returns is the same for the sample of senior high school graduates (whose schooling decision we are analyzing) and the sample of a cohort that is one year younger and just starting grade 12 (who have thus not decided yet about whether to enrol in college or not). To control for potential differences in the composition of the two cohorts, we regress expected log earnings, expected returns, the probability of working and the variance of log earnings on individual and family characteristics (see list of variables in Table 2) and compare the distributions of the residuals.

Figure 1: Comparing Expectations of Senior High School Graduates with a One-Year Younger Cohort: Youth Respondents' Earnings Expectations

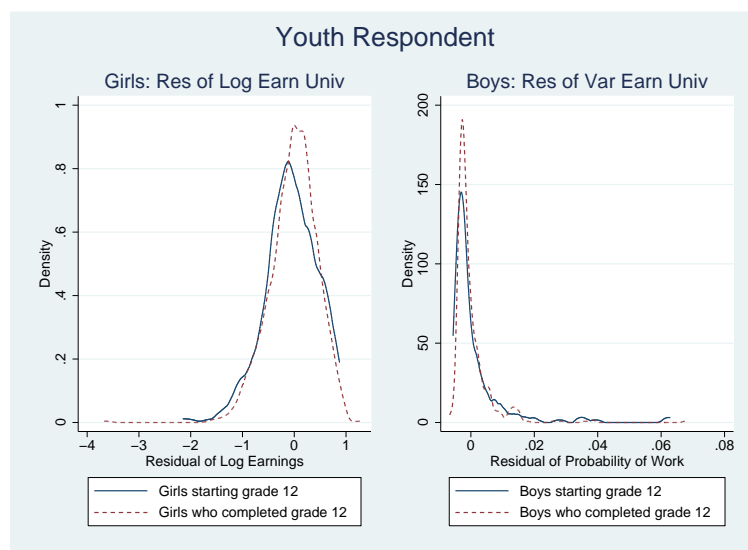


Figure 2: Comparing Expectations of Senior High School Graduates with a One-Year Younger Cohort: Youth Respondents' Work Perceptions

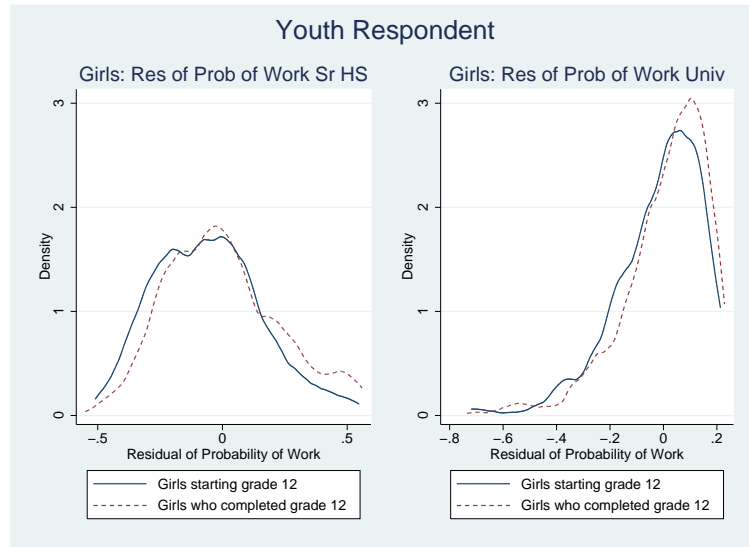


Figure 3: Comparing Expectations of Senior High School Graduates with a One-Year Younger Cohort: Mother Respondents

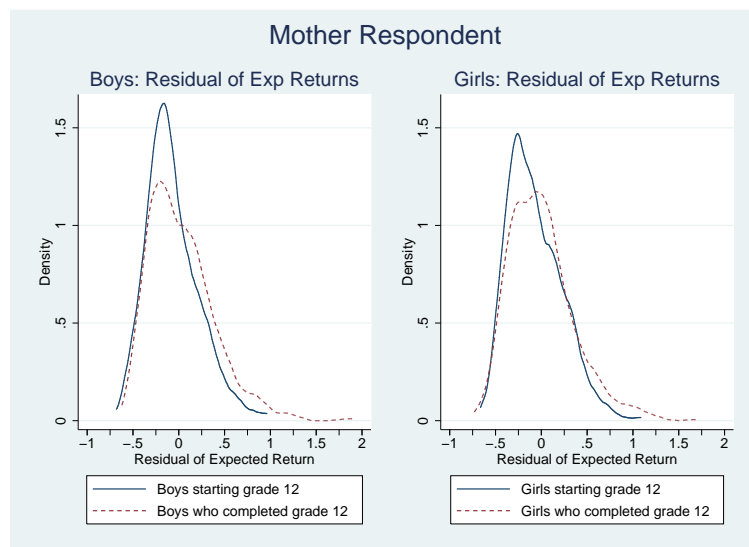
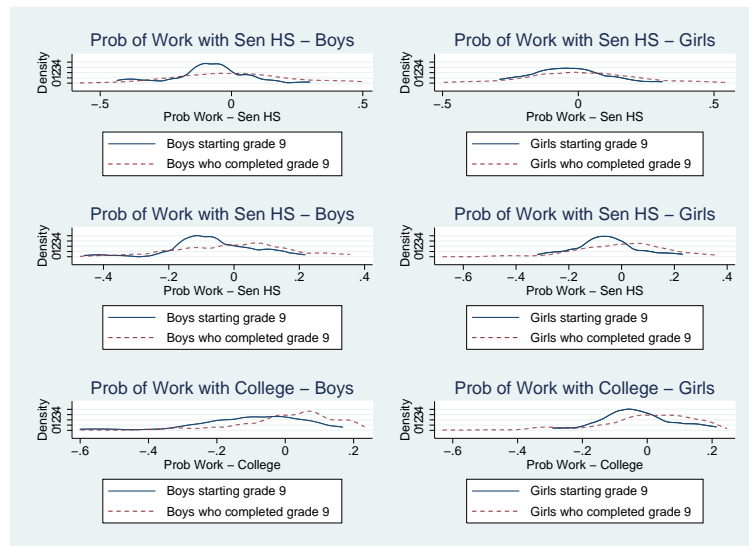


Table 29: High School Attendance Choice: Rationalization of Choices

Distrib of Resid of	Youth Resp		Mother Resp	
	Boys	Girls	Boys	Girls
	P-Val of KS-Test		P-Val of KS-Test	
Exp Log Earnings				
- Junior HS	0.937	0.039	0.000	0.000
- Senior HS	0.760	0.511	0.000	0.000
- College	0.963	0.228	0.000	0.000
Exp Return				
- Senior HS	0.173	0.139	0.276	0.010
- College	0.251	0.158	0.231	0.062
Prob of Work				
- Junior HS	0.182	0.182	0.001	0.015
- Senior HS	0.139	0.466	0.000	0.000
- College	0.074	0.513	0.000	0.020
Var of Log Earnings				
- Junior HS	0.447	0.723	0.159	0.022
- Senior HS	0.984	0.187	0.146	0.152
- College	0.484	0.460	0.069	0.002
Observations				
(Jun HS Grads/Grade 9)	683/119	759/141	473/267	494/197

Notes: Table displays the p-values of Kolmogorov-Smirnov tests of equality of distributions. The null hypothesis is that the cross-sectional distribution of -for example- expected returns is the same for the sample of junior high school graduates (whose schooling decision we are analyzing) and the sample of a cohort that is one year younger and just starting grade 9 (who have thus not decided yet about whether to enrol in senior high school or not). To control for potential differences in the composition of the two cohorts, we regress expected log earnings, expected returns, the probability of working and the variance of log earnings on individual and family characteristics (see list of variables in Table 1) and compare the distributions of the residuals.

Figure 4: Comparing Expectations of Junior High School Graduates with a One-Year Younger Cohort: Mother Respondents



## Returns Based on Minimum and Maximum Earnings

Table 30: College Enrollment Decision: Return Based on Minimum Earnings with Different Degrees

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
"Min" Expected Return - College	0.329* (0.169)	0.077 (0.132)	0.016 (0.141)	0.186 (0.143)
Prob of Work - Sr HS	-0.578 (0.450)	0.239 (0.356)	-0.433 (0.364)	0.064 (0.348)
Prob of Work - College	0.198 (0.478)	0.117 (0.415)	0.187 (0.427)	1.133*** (0.398)
Var of Log Earn - Sr HS	-10.059 (9.774)	-8.924 (7.496)	13.571 (9.275)	2.932 (8.658)
Var of Log Earn - College	-0.145 (12.684)	9.207 (8.851)	-18.397* (11.061)	17.549 (12.017)
Age	-0.153*** (0.047)	-0.144*** (0.047)	-0.018 (0.035)	-0.057 (0.037)
Obese	-0.429 (0.582)	-0.720* (0.391)	-0.286 (0.331)	-0.340 (0.298)
GPA of Jr HS (0-100)	0.027*** (0.009)	0.024*** (0.007)	0.019*** (0.007)	0.009** (0.004)
Mother's Educ - Jr HS	0.190 (0.171)	-0.055 (0.159)	0.337** (0.149)	0.324** (0.147)
Mother's Educ - Sr HS	0.356 (0.472)	0.522 (0.325)	0.951** (0.398)	0.664** (0.290)
Father's Educ - Jr HS	0.072 (0.194)	0.306* (0.166)	0.231 (0.158)	-0.030 (0.155)
Father's Educ - Sr HS	0.516 (0.322)	0.372 (0.278)	-0.122 (0.315)	0.541* (0.306)
Per cap Income - below 5k	0.079 (0.147)	-0.149 (0.134)	-0.225* (0.121)	-0.144 (0.130)
Per cap Income - above 10k	0.146 (0.179)	0.395** (0.154)	0.264* (0.148)	0.584*** (0.147)
Father's Occup - Self-Empl	-0.336* (0.192)	-0.113 (0.158)	-0.235* (0.142)	0.021 (0.138)
Father's Occup - Fam Worker	0.153 (0.341)	-0.474 (0.416)	-0.772* (0.410)	-0.187 (0.327)
Mother in Labor Force	0.112 (0.190)	0.170 (0.145)	0.258* (0.147)	0.053 (0.138)
Dist to Univ 20 to 40km	-0.307* (0.166)	-0.262* (0.141)	0.033 (0.138)	0.187 (0.127)
Dist to Univ above 40km	-0.115 (0.181)	-0.533*** (0.169)	0.166 (0.156)	0.208 (0.153)
Municip with Univ (Tuition above Med)	-0.026 (0.198)	-0.732*** (0.185)	-0.077 (0.174)	-0.244 (0.174)
Observations	2829		2829	
Censored Obs	1468		1361	
Log Likelihood	-2498.153		-2704.131	
Sample Sel: Corr of Errors	0.237		-0.154	
P-val: LR Test of Indep Eqns	0.503		0.662	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Excl. categories: not obese, father in household, mother's and father's education primary or less, father's occupation worker, per capita income between 5000 and 10000 pesos, distance to university less than 20km, tuition costs less than 750 pesos. All specifications include state dummies.

Table 31: College Enrollment Decision: Return Based on Maximum Earnings with Different Degrees

Dependent Variable	College Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
"Max" Expected Return - College	0.370** (0.167)	0.187 (0.133)	-0.012 (0.144)	0.177 (0.144)
Prob of Work - Sr HS	-0.541 (0.450)	0.269 (0.355)	-0.440 (0.366)	0.079 (0.350)
Prob of Work - College	0.137 (0.479)	0.055 (0.415)	0.198 (0.431)	1.115*** (0.401)
Var of Log Earn - Sr HS	-1.430 (9.153)	-6.847 (7.066)	13.709 (9.084)	7.573 (8.029)
Var of Log Earn - College	-10.730 (12.854)	6.363 (8.559)	-18.514* (10.963)	11.614 (11.898)
Age	-0.151*** (0.047)	-0.143*** (0.047)	-0.019 (0.035)	-0.057 (0.037)
Obese	-0.422 (0.579)	-0.713* (0.389)	-0.286 (0.331)	-0.341 (0.298)
GPA of Jr HS (0-100)	0.027*** (0.009)	0.023*** (0.007)	0.019*** (0.007)	0.009** (0.004)
Mother's Educ - Jr HS	0.194 (0.171)	-0.057 (0.159)	0.337** (0.149)	0.328** (0.147)
Mother's Educ - Sr HS	0.362 (0.468)	0.508 (0.325)	0.952** (0.398)	0.668** (0.290)
Father's Educ - Jr HS	0.063 (0.194)	0.309* (0.166)	0.232 (0.158)	-0.033 (0.155)
Father's Educ - Sr HS	0.525 (0.321)	0.370 (0.277)	-0.118 (0.316)	0.537* (0.306)
Per cap Income - below 5k	0.079 (0.147)	-0.146 (0.133)	-0.224* (0.121)	-0.147 (0.130)
Per cap Income - above 10k	0.138 (0.179)	0.397** (0.155)	0.265* (0.148)	0.583*** (0.147)
Father's Occup - Self-Empl	-0.338* (0.192)	-0.118 (0.158)	-0.233* (0.142)	0.026 (0.138)
Father's Occup - Fam Worker	0.168 (0.340)	-0.473 (0.417)	-0.773* (0.411)	-0.189 (0.327)
Mother in Labor Force	0.110 (0.190)	0.176 (0.145)	0.260* (0.147)	0.054 (0.138)
Dist to Univ 20 to 40km	-0.311* (0.167)	-0.258* (0.141)	0.033 (0.138)	0.189 (0.127)
Dist to Univ above 40km	-0.116 (0.181)	-0.537*** (0.169)	0.170 (0.155)	0.212 (0.153)
Tuition more than 750 Pesos	-0.022 (0.198)	-0.741*** (0.186)	-0.074 (0.174)	-0.242 (0.173)
Observations	2829		2829	
Censored Obs	1468		1361	
Log Likelihood	-2496.807		-2704.224	
Sample Sel: Corr of Errors	0.257		-0.149	
P-val: LR Test of Indep Eqns	0.472		0.673	

Notes: Table displays coefficients and standard errors in brackets. \*  $p < 0.1$  \*\*  $p < 0.05$  \*\*\*  $p < 0.01$ . Excl. categories: not obese, father in household, mother's and father's education primary or less, father's occupation worker, per capita income between 5000 and 10000 pesos, distance to university less than 20km, tuition costs less than 750 pesos. All specifications include state dummies.



Table 32: High School Enrollment Decision: Return Based on Minimum Earnings with Different Degrees

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
"Min" Expected return - Sr HS	-0.063 (0.196)	0.169 (0.187)	0.099 (0.245)	-0.106 (0.261)
"Min" Expected return - College	0.186 (0.164)	0.279 (0.172)	0.105 (0.230)	-0.272 (0.240)
Prob of Work - Jr HS	0.089 (0.375)	0.020 (0.330)	0.515 (0.439)	-0.806* (0.468)
Prob of Work - Sr HS	-0.065 (0.498)	-0.178 (0.440)	-0.925 (0.638)	-0.120 (0.624)
Prob of Work - College	0.686* (0.402)	-0.175 (0.453)	0.936* (0.551)	1.101** (0.537)
Var of Log Earn - Jr HS	5.429 (7.529)	7.573 (8.518)	12.638 (9.269)	-6.179 (9.386)
Var of Log Earn - Sr HS	8.201 (10.765)	-12.030 (8.764)	-11.057 (13.296)	-15.215 (10.193)
Var of Log Earn - College HS	9.610 (11.973)	12.650 (11.419)	-8.350 (13.238)	3.705 (11.031)
Number of Brothers	-0.165*** (0.053)	-0.084* (0.048)	0.037 (0.060)	0.031 (0.068)
Number of Sisters	-0.117** (0.051)	-0.050 (0.048)	-0.110** (0.054)	0.054 (0.061)
GPA of Jr HS (0-100)	0.027*** (0.006)	0.033*** (0.007)	0.024*** (0.007)	0.031*** (0.007)
No Father in Household	-0.107 (0.166)	0.400** (0.172)	0.184 (0.213)	-0.165 (0.230)
Mother's Educ - Jr HS	0.254 (0.169)	0.404** (0.173)	0.351* (0.199)	0.540** (0.228)
Father's Educ - Jr HS	0.076 (0.191)	0.146 (0.180)	0.443* (0.229)	-0.021 (0.234)
Per cap Income - below 5k	0.166 (0.153)	-0.049 (0.147)	-0.191 (0.176)	-0.510*** (0.187)
Father's Occup - Self-Empl	0.246 (0.220)	0.407** (0.204)	-0.191 (0.219)	-0.159 (0.277)
Father's Occup - Fam Worker	4.249 (233.662)	-0.079 (0.709)	0.326 (0.499)	-1.004** (0.508)
Locality Size 15 to 50k	0.228 (0.178)	-0.083 (0.163)	0.374* (0.198)	0.174 (0.196)
Locality Size above 50k	-0.087 (0.143)	-0.107 (0.143)	0.257 (0.183)	0.459** (0.209)
Observations	2364		2364	
Censored Obs	967		1397	
Log Likelihood	-1983.138		-1838.881	
Sample Sel: Corr of Errors	-0.511		0.326	
P-val: LR Test of Indep Eqns	0.051		0.350	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Excl. categories: size of locality less than 15k, not obese, father in household, mother's and father's education primary or less, father's occupation worker, per capita income between 5000 and 10000 pesos. All specifications include state dummies.

Table 33: High School Enrollment Decision: Return Based on Maximum Earnings with Different Degrees

Dependent Variable	High School Attendance Decision			
	Youth Respondent		Mother Respondent	
	Boys	Girls	Boys	Girls
"Max" Expected Return - Sr HS	-0.012 (0.203)	0.154 (0.194)	0.194 (0.250)	0.016 (0.282)
"Max" Expected Return - College	0.234 (0.170)	0.208 (0.176)	-0.011 (0.240)	-0.278 (0.241)
Prob of Work - Jr HS	0.104 (0.375)	0.028 (0.329)	0.534 (0.439)	-0.789* (0.469)
Prob of Work - Sr HS	-0.062 (0.501)	-0.196 (0.442)	-0.930 (0.638)	-0.144 (0.624)
Prob of Work - College	0.637 (0.404)	-0.187 (0.456)	0.934* (0.554)	1.138** (0.545)
Var of Log Earn - Jr HS	4.054 (6.289)	11.486 (8.172)	14.701* (8.409)	-8.285 (8.269)
Var of Log Earn - Sr HS	14.325 (9.599)	-10.037 (8.542)	-11.473 (10.895)	-19.339** (9.534)
Var of Log Earn - College HS	3.838 (11.038)	5.329 (10.940)	-10.518 (12.885)	10.722 (10.675)
Number of Brothers	-0.167*** (0.053)	-0.085* (0.047)	0.034 (0.060)	0.029 (0.068)
Number of Sisters	-0.120** (0.051)	-0.052 (0.048)	-0.107** (0.054)	0.054 (0.061)
GPA of Jr HS (0-100)	0.027*** (0.006)	0.034*** (0.007)	0.025*** (0.007)	0.031*** (0.007)
No Father in Household	-0.106 (0.166)	0.404** (0.172)	0.176 (0.213)	-0.167 (0.231)
Mother's Educ - Jr HS	0.255 (0.169)	0.407** (0.172)	0.356* (0.199)	0.541** (0.228)
Father's Educ - Jr HS	0.071 (0.192)	0.153 (0.180)	0.443* (0.230)	-0.014 (0.233)
Per cap Income - below 5k	0.167 (0.153)	-0.050 (0.147)	-0.191 (0.176)	-0.504*** (0.187)
Father's Occup - Self-Empl	0.247 (0.220)	0.415** (0.204)	-0.189 (0.219)	-0.161 (0.277)
Father's Occup - Fam Worker	4.621 (687.279)	-0.095 (0.705)	0.311 (0.497)	-0.991* (0.508)
Locality Size 15 to 50k	0.222 (0.178)	-0.089 (0.163)	0.378* (0.198)	0.172 (0.196)
Locality Size above 50k	-0.091 (0.143)	-0.110 (0.143)	0.255 (0.183)	0.453** (0.209)
Observations	2364		2364	
Censored Obs	967		1397	
Log Likelihood	-1983.551		-1838.826	
Sample Sel: Corr of Errors	-0.511		0.334	
P-val: LR Test of Indep Eqns	0.050		0.343	

Notes: Table displays coefficients and standard errors in brackets. \* p<0.1 \*\* p<0.05 \*\*\* p<0.01. Excl. categories: size of locality less than 15k, not obese, father in household, mother's and father's education primary or less, father's occupation worker, per capita income between 5000 and 10000 pesos. All specifications include state dummies.