

Education Choices and Returns on the Labor and Marriage Markets: Evidence from Data on Subjective Expectations.

Orazio P. Attanasio* and Katja M. Kaufmann†

March 2016

Abstract

In this paper we analyze the role of expected labor and marriage market returns as determinants of the college enrollment decisions of Mexican high school graduates. In particular, we investigate whether the (relative) weights of these factors vary by gender. While expectations regarding labor market outcomes are directly elicited from the youths, marriage market returns are proxied by the (net-)supply of potential partners in the youths' local marriage markets. We find that labor market expectations play a more important role in the decisions of boys. Marriage market considerations on the other hand appear to be more important for girls.

JEL-Classification: I21, I22, O15

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

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

†Corresponding author: Katja Maria Kaufmann, Mannheim University, CESifo and IZA, Address: L 7, 3-5, 68131 Mannheim, Germany, kaufmann@vwl.uni-mannheim.de. We would like to thank Ran Abramitzky, Ghazala Azmat, Martha Bailey, Sandra Black, Janet Currie, Taryn Dinkelman, Christina Gathmann, Pamela Giustinelli, Dan Hammermesh, Caroline Hoxby, Rob Jensen, Victor Lavy, Steve Lehrer, Lance Lochner, Shaun McRae, Aprajit Mahajan, Matthias Messner, Karthik Muralidharan, Aloysius Siow, Jeff Smith, Kevin Stange, Alessandro Tarozzi, Michele Tertilt and Sarah Turner for valuable comments. We are thankful also to conference participants at the NBER Education meeting, at the CEPR labour economics program on the economics of family policy and the CEPR development meeting, 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.

1 Introduction

It is a well-known fact that the gender gap in educational attainment has not only decreased over the past decades, but that in many OECD countries women have overtaken men in terms of college attendance rates. This observation is puzzling in light of the fact that gender gaps in labor force outcomes continue to exist (female labor force participation rates are lower, women are more likely to work part-time and women experience lower earnings than men). This suggests that the determinants of educational decisions might differ for men and women and that –at least for women– labor market returns might not be the only driving force. A natural candidate for an alternative source of returns is the marriage market (as suggested by Goldin (1992), Goldin, Katz, and Kuziemko (2006) and Bailey and Dynarski (2011) who document and discuss the "college enrolment puzzle"). Nevertheless, to this date there is little empirical evidence on whether marriage market considerations play a role in educational decisions.

This paper aims to contribute to closing this gap by investigating whether marriage market returns matter for educational decisions, what are the relative weights of labor and marriage market returns and whether they differ by gender. These questions are not only essential for understanding the driving forces behind educational choices, but also for the effective design of programs intended to increase schooling and to reduce existing gender gaps (e.g. in developing countries).

More specifically, the goals of this paper are twofold. First, we want to improve our understanding of how labor market considerations determine education decisions. In particular, we use subjective expectation data that directly reflect how Mexican youths perceive their labor market prospects (in terms of earnings levels and earnings and employment risks) for different schooling alternatives. Second, we combine this labor market perspective with an analysis of the role of the marriage market in educational decisions. Finding a good partner is generally seen as one of the most important problems that individuals face during early adulthood. Since youths' educational decisions are likely to influence the set of available partners and their attractiveness in the marriage market, it is therefore natural to think that marriage market considerations might play a crucial role in these decisions.¹ Including the marriage market into our analysis allows us to shed some light on whether there are differences in the (relative) weights that females and males assign to labor and marriage market considerations in their educational decisions.

¹For example, by enrolling in college individuals signal their ability and social status and may thereby become more attractive for their potential partners (see, e.g., Lee (2015) for the role of educational attainment in dating and marriage decisions). Moreover, if college enrollment requires the individual to move from a village to a city, this may imply a far larger set of potential partners to choose from (see, e.g. Botticini and Siow (2008)).

The main challenge in the empirical analysis of educational decisions is the fact that we cannot directly observe people's beliefs about their potential returns. Even ex-post we can only observe the realization of the chosen alternative. 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. For example, such an approach has to make assumptions about whether large shocks to an individual's earnings (that take place once the individual has completed her schooling and works) were predicted by the individual ex-ante, i.e. at the time of the schooling decision. Furthermore, the researcher has to make assumptions about whether and how an individual takes into account that other individuals, whose earnings she observes (and uses in forming her own expectations), are typically self-selected into that education level.

These assumptions are of particular concern in the analysis of gender differences in schooling decisions, since there might be important differences in terms of how well boys and girls are informed about future labor market earnings. For example, in Mexico female college attendance and labor force participation rates have been increasing in recent years, but starting from a very low level compared to men. Thus girls might be less well informed than boys, since they only observe few women with college degrees and even fewer women's earnings given their low labor force participation rates. This might affect their expectations about returns as well as their perceptions about the uncertainty associated with different education choices, both of which we elicit directly. Moreover, beliefs about future labor market outcomes depend on an individual's perception of her own skills and on how these skills (and other characteristics) affect her future earnings. Thus if, for example, girls are more likely to underestimate their skills, they are also more likely to underestimate their future earnings than boys are. For these reasons, we elicit boys' and girls' expectations directly, since what matters for the college attendance decision is each individual's perception of her potential returns to schooling (and, more generally, of her labor market outcomes for different schooling levels).

We address our questions in the context of the college enrollment decision. In particular, we analyze the decision of youths who have completed senior high school to enroll (or not) in college. We construct measures of expected earnings and perceived earnings and unemployment risk for different schooling scenarios, to study the extent to which these expectations affect education choices of male and female youths. In particular, we estimate a model of schooling decisions to measure the importance of expected labor market returns to education for boys and girls. In addition, we provide evidence on the role of returns in the marriage market in education choices. We use two different proxies for marriage market returns: (i) the ratio of unmarried men to women in the locality of residence to capture the availability

of partners; and (ii) the ratio of unmarried men to women with a certain level of schooling, if the suitability of the partner depends on having a similar (or higher) education level.

The idea for using these proxies is simple: if the likelihood of finding a (suitable) partner in the locality of residence is low, leaving the locality to go to college has the benefit of increasing the chances of finding a (suitable) partner and of improving the outside option in the case that no partner is found. Due to social norms in Mexico, the link between schooling and marriage market considerations is likely to be particularly strong for girls, because parents would be very reluctant to let their unmarried daughter move to a different city to work and live by herself, while more willing to let her leave for continuing school (where she can live with classmates). Anecdotal evidence strongly points to the importance of marriage market considerations in schooling decisions: in Mexico there even exists the acronym “mmc”s (*‘mientras me caso’*) for girls going to college to find a husband.

Of course one might be worried that our proxies (sex ratios conditional on education and unconditionally) capture not only marriage market conditions, but also labor market conditions. In this context, our data on each individual’s expectations about her potential labor market outcomes proves very useful to address this concern, since we can analyze jointly the importance of labor market and marriage market considerations as determinants of schooling decisions.²

Our results indicate that both boys and girls expect high labor market returns to college education and perceive lower unemployment and earnings risks for higher education levels. At the same time we find that labor market expectations, in particular expected returns to college, are an important determinant for boys, while they play less of a role for girls. Marriage market considerations, on the other hand, appear to be more important for girls.

Our subjective expectations data allow us to address in a rather direct and straightforward way questions which are complicated to address (without making strong additional assumptions) when beliefs are not observed. On the other hand, the use of expectations data is sometimes criticized on the basis of concerns such as endogeneity or ex-post rationalization. We will discuss these concerns in detail and argue why we believe that these issues are not driving our results.

Related literature Our paper is closely related to the following three strands of the literature, the literature on the role of people’s subjective expectations in their educational decisions, the one on gender differences in education and expectations, and the one on education and marriage market returns.

²For robustness, we show that results continue to hold if we control for state and local labor market conditions and we show that individuals’ own beliefs are more important for their educational decisions than these general labor market conditions.

In recent years there has been a surge in the 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.³

Papers that investigate the link between subjective expectations of earnings and education choices have analyzed school enrolment decisions (see, e.g., Nguyen (2008), Jensen (2010), Attanasio and Kaufmann (2014) and Kaufmann (2014)), high school track choices (see, e.g., Giustinelli (forthcoming) and college major decisions (see, e.g., Arcidiacono, Hotz, and Kang (2012), Stinebrickner and Stinebrickner (2014) and Zafar (2009)). Another literature has analyzed the role of beliefs about the costs of education by providing information about loans and fellowships in a randomized fashion (see, e.g., Dinkelman and Martinez (2014)).

In contrast to the latter three groups, we are interested in the decision to enrol in college or not, and we analyze the importance of gender differences with respect to the relative weight of labor market and marriage market considerations. Our paper is more closely related to the first group, which investigates how perceived returns affect schooling investments. For example, Jensen (2010) finds that 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 them. Nguyen (2008) analyses whether it matters how such information is transmitted and finds bigger effects on schooling outcomes from direct quantitative information about higher returns compared to the effects of role models telling their success stories. Attanasio and Kaufmann (2014) analyze whether youths’ own or parents’ beliefs about future labor market outcomes are relevant for schooling decisions. Kaufmann (2014) analyzes the causes and consequences of the large income gradient in college attendance in Mexico. In particular, data on subjective expectations in combination with a structural model of schooling choices enable her to analyze to what extent the gradient can be explained by the poor expecting lower returns to college than the rich, for example due to ability and/or information differences.

With respect to those papers the main distinguishing features of our work are the focus on differences in beliefs and behavior between males and females and the inclusion of mar-

³See also the survey paper on the use of subjective expectations data by Manski (2004). Since then the literature has started using data on people’s beliefs in a wide variety of areas, such as, for example, in health decisions (see, e.g., Mahajan and Tarozzi (2011), Mahajan, Tarozzi, Yoong, and Blackburn (2011) and Alvarez and Vera-Hernandez (2013)), in financial decisions (see, e.g., Drerup, Enke, and von Gaudecker (2014)), in the decision to engage in crime (see, e.g. Lochner (2007)) and so forth. Attanasio (2009) and Delavande, Giné, and McKenzie (2011) survey the literature that uses data on subjective expectations in developing countries and find that also in this context individuals are willing to answer the expectations questions and understand them reasonably well (sometimes with the help of visual aids).

riage market considerations. Moreover, we also elicit and use information about the risk that individuals perceive in association with each alternative (such as wage risk and unemployment risk), which might be particularly relevant in the comparison of determinants between boys and girls (for example because of gender differences in risk aversion).

Gender differences have been documented not only in terms of preferences (such as in terms of the degree of risk aversion, see, e.g. Charness and Gneezy (2012) or in terms of the willingness to compete, see, e.g. Booth and Nolen (2012) and Gupta, Poulsen, and Villeval (2013)), but also in terms of subjective expectations about future outcomes (see, e.g., Filippin and Ichino (2005) who show that college students' beliefs about the gender gap in earnings are relatively close to the actual observed gap at the beginning of the career, but underestimated compared to the actual gap at later stages). Differences in expectations can arise, for example, due to actual differences in the labor market (see, e.g., the survey articles by Azmat, Guell, and Manning (2006) and Azmat and Petrongolo (2014) and, for example, on reasons for earnings differences, Castillo, Petrie, Torero, and Vesterlund (2013) on differences in bargaining and Mengel (2015) on differences in networking) as well as due to gender differences in self-assessment (see, e.g. Dohmen and Falk. (2011)) or differences in probability judgements (see, e.g., Dohmen, Falk, Huffman, Marklein, and Sunde (2009) for evidence related to financial decisions).

There is a large literature documenting the existence of gender differences in educational attainment (in terms of years of schooling, but also, for example, in terms of choice of subjects, see, e.g., Joensen and Nielsen (2015)), in particular in developing countries. Two recent papers that discuss gender issues related to the role of returns to education and school choices are Abramitzky and Lavy (2014) and Jensen (2012). Both of them exploit (quasi-)experimental changes to returns to school. More specifically, Abramitzky and Lavy (2014) exploit a policy change in Israeli kibbutzim which led to a substantial increase in the returns to education. They find 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 that he induces in randomly selected villages through the provision of recruitment services. While these two papers share with our work the interest in gender differences with respect to labor market returns, we aim to complement this literature by adding another potential determinant of education choices, namely marriage market returns, which might be an additional source for gender differences in schooling decisions.

Very few papers consider the importance of marriage market considerations. The hypothesis that marriage market considerations are an important determinant for individuals' educational decisions is supported by the descriptive evidence provided in Goldin (1992).

She shows that American women who went to college and graduated between 1945 and 1960 had substantially higher earning husbands. Two recent papers provide causal evidence on the link between education and marriage market outcomes. Oreopoulos and Salvanes (2011) estimate a number of different non-monetary returns to schooling and show, among other things, that higher education leads to an increased probability to be married. Kaufmann, Messner, and Solis (2012) make use of administrative data from Chile and apply a regression discontinuity design to analyze the longrun effects of being admitted to a more elite university. They find substantial returns in terms of women finding a husband who is smarter, from a better university and from better family background, thus supporting not only the idea that marriage market returns are important, but also providing evidence on the asymmetry in the importance of marriage market returns by gender.⁴ Kaufmann, Messner, and Solis (2015) show that these marriage market returns can be of first-order importance for the next generation, i.e. the students' children.

In contrast to these papers, our goal is to understand the role of marriage market considerations ex-ante, an aim which is shared by Lafortune (2012). She uses a different approach to analyze how marriage market conditions affect pre-marital investment decisions. In particular, she explores whether second generation Americans modify their human capital acquisition decision when faced with a shift in the sex ratio of their state-level marriage market induced by immigration of their own ethnicity. Our paper complements her findings and provides evidence on gender differences and on the (relative) weight of labor market and marriage market considerations in educational decisions.

Outline The rest of the paper is organized as follows: Section 2 presents a basic model of education choices. Section 3 describes in detail the data, in particular the module on subjective expectations, and presents summary statistics. Section 4 provides supporting evidence that individuals are able to give meaningful answers to the subjective expectation questions and investigates how well male and female youths are informed about potential returns to schooling. Section 5 discusses the main results on the role of labor and marriage market returns for male and female youths (including a robustness section on concerns related to the use of subjective expectation data). Section 6 concludes.

⁴This asymmetry can be explained by men and women trading off characteristics in potential partners in different ways, in particular men putting a greater weight on factors not related to cognitive ability, education, earnings, such as beauty or health. This is supported for example by Hamermesh and Biddle (1994) and Lee (2015).

2 Schooling Decisions and Returns to Schooling

The goal of this paper is to improve our understanding of the main determinants of education choices and analyze whether they differ between boys and girls. For this purpose, we show how schooling decisions of young poor Mexicans relate to their expectations on the return and risk to that investment as well as to returns in the marriage market.

2.1 Determinants of Enrolment

To analyze the relationship between enrolment and expected returns, one possible approach 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 decisions (see, for instance, Keane and Wolpin (1997), Todd and Wolpin (2006) and Attanasio, Meghir, and Santiago (2011)). Kaufmann (2014) shows how data on people’s subjective expectations can be used in a simple model of college enrolment choice. Instead of following this route, we present probit regressions that relate the probability of enrolment to the commonly used control variables in the context of schooling decisions (such as ability, parental education, income and so forth), and in addition to subjective expectations of earnings and proxies for returns in the marriage market. 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. Moreover, we control for current labor market conditions through state dummies and –in further robustness checks– we also control for local labor market conditions such as local earnings for different schooling levels

Specifically, to model the decision to enrol in college, having completed senior high school, we use a latent index model we estimate on the sample of senior high school graduates. Denoting with S the college enrolment decision ($S = 1$ if the individual decides to attend and $S = 0$ otherwise), the latent index model for attending college can be written as:

$$S = \begin{cases} 1 & \text{if } S^* \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

where,

$$S^* = \alpha + \beta\rho + \sum_{z=1}^2 \gamma_z Var(\log Y_z) + \sum_{z=1}^2 \delta_z P_z^W + \mu R^M + X'\theta + U. \tag{1}$$

where $z = 1, 2$ denotes 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 characteristics and θ denotes the vector of corresponding parameters. In terms of subjective expectation measures ρ denotes the expected (gross) returns to college defined as $\rho = E(\log Y_2) - E(\log Y_1)$, where $E(\log Y_z)$ is the expected value of the distribution of (log) earnings at age 25 for the scenario that degree z ($z = 1, 2$) 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. R^M is a proxy for marriage market returns.

We obviously expect the return to college relative to senior high school, as perceived by the individual, to have a positive effect on the decision to attend college. 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 include in our regressions 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 senior high school degree to have a positive effect on the probability of continuing to college, and a high variance of log earnings with a college degree to have a negative effect.

In conventional approaches without data on subjective expectations, 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 as well as differences in psychological costs of attending college (see, e.g., Cameron and Heckman (1998) and Cameron and Heckman (2001)). 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. Most importantly, 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 schooling decision). This provides a strong rationale for using "perceived" returns and "perceived" risk. We nevertheless also control for GPA and parental education both of which turn out to be very important determinants of the college enrolment decision we model. These variables might be proxying for the probability of completing college and for preferences for education.

To investigate whether marriage market considerations are important for education choices, we include in the regressions proxies for returns in the marriage market, R^M . Using census data, we compute the ratio of unmarried men to women with at least a senior high school degree 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.⁵ 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 creates 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 education level (or higher) 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 (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 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 college attendance decisions. Moreover, we show that results remain unchanged if we control, in addition, for local labor market conditions, such as average local earnings for different schooling degrees.

A second important concern are potential ‘peer’ effects. We address this concern in two ways. First, 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.⁶ Second, we also consider an alternative proxy for the marriage market return, which should not suffer from the concern that it can be confounded with educational peer effects: the ratio of unmarried men to women without constraining the measure to a specific education level. Also this measure might be more appropriate, if the youth (for example, the male youth) cares less about the

⁵Edlund (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.

⁶When constructing our control variable for potential peer effects, we consider both married and unmarried women (men) who went to college as a fraction of all women (men) in the locality of residence (in contrast to the sex ratio that takes into account only unmarried men and women with at least a high school degree), since we want to control for potential peer effects in the college enrolment decision which could arise from observing others who went to college.

education level of the spouse.

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.

3 Data

The conditional cash transfer program *Oportunidades*, previously known as PROGRESA, has introduced a new component known as *Jóvenes con Oportunidades* in 2002/3. The data we use was collected in 2005 as part of the evaluation of this new program component. 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 beneficiaries of *Jóvenes con Oportunidades*. To enter this program, individuals have to be in their last year of junior high school (9th grade) or attend senior high school (10 to 12th grade), they have to be younger than 22 years of age, and they have to be 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.

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 concentrate only on students who

just finished grade 12 (senior high school graduates) and decide about enrolling in college or not (around 2500 observations).⁷ We use the data on the siblings sample for a number of robustness exercises on the expectations data.

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. For our cohort of senior high school graduates the following 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 two 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 second 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?*

⁷Unfortunately, we lose nearly 60% of observations when merging our data with Census data to construct marriage market proxies on locality level, since Census information is only available for part of the localities in which the youths of our sample live (for a more detailed description, see section 3.5).

- (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. These questions were included in the youth questionnaire and therefore are available for every youth in the household aged 15 to 25. When a specific youth was not present, these questions were answered by the mother (referring in the expectation questions to the specific child not present). This poses potential selection issues that we discuss at length below.

This type of subjective expectations questions has been used extensively in a variety of contexts (see, e.g., the survey papers by Manski (2004), Attanasio (2009) and Delavande, Giné, and McKenzie (2011)). The latter two survey the literature that uses data on subjective expectations in developing countries and find that individuals are willing to answer the expectations questions and understand them reasonably well (sometimes with the help of visual aids). In section 4 we show that – also in our context – respondents seem to have understood the questions and that the data pass a number of internal and external validity tests.

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 raises the potential concern that individuals might try to rationalize their choices. 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.⁸

To check whether ex-post rationalization introduces biases in expectations, we look at the expectations data of the siblings of our main subjects, that is 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

⁸This 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 people with *low* expected returns to college decide to enrol in college and now state high returns to college and vice versa.

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 5.3 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 two possible scenarios of final education attainment: senior high school ($z = 1$) and college ($z = 2$). 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 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 ($z = 1, 2$) 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. We compute these as the difference between expected log college earnings and expected log senior high school earnings, $\rho = E(\ln(Y^2)) - E(\ln(Y^1))$.⁹

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 dis-

⁹Notice 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.

tribution, but we perform robustness checks using the other two distributional assumptions and point out differences if they occur.

3.4 Potential Sample Selection Issues

As discussed above, 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.

Attanasio and Kaufmann (2009) use these data to analyze the intra-household decision process, that is they investigate whose expectations matter for schooling decisions, the ones of the mother or the ones of the youth. They show that in terms of the college attendance decision youths’ own expectations play the major role. For that reason, we will be using the expectations held by the youths’ in the following analysis, that is we use the subsample of senior high school graduates for which the youths answered the expectation questions themselves. Since the respondent was not randomly determined, we address the concern of sample selection bias as follows: we use a standard Heckman selection correction approach (see Heckman (1979)) adjusted to a non-linear context, that is we estimate jointly a latent index model for college attendance and a sample selection equation.

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 schooling decision 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). When considering the timing variables we experimented with various set of variables. We find that, in our sample, youths are significantly more likely to be at home (and thus able to answer the expectation questions themselves) during evenings on weekends. In addition, the indicator for holiday weeks is a strongly significant determinant of who is the respondent (see Table 1).

3.5 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. We

construct the ratio of unmarried men to women who are in a similar age range as the youth, that is age 18 to 30.

To construct those proxies, we make use of Census data of the year 2000, for which information is available at the locality level.¹⁰ 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, this strategy has the drawback that Census information is only available for part of the localities in which the youths of our sample live. Therefore we lose close to 60% of observations of our original sample of senior high school graduates.

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. We also discuss some of the econometric issues involved with the use of these data.

4.1 Descriptive Statistics: Characteristics of the Youth

Table 2 reports summary statistics of individual and family background characteristics for our sample of senior high school graduates. We present results separately for male and female youths and report whether differences are significant.

In the first three columns of Table 2, we present summary statistics for the full sample of high school graduates, while in columns 4 to 6 we present summary statistics for the subsample of youths respondents, which is the relevant sample for our analysis. As discussed in section 3.4, for nearly half the sample the questions on subjective expectations were not answered by the youths themselves, but by their mothers. For our analysis of college attendance decisions, we will be using the sample of youth respondents, since we have shown in a related paper that for the decision to go to college, youths’ expectations are the ones that are relevant (see Attanasio and Kaufmann (2009)).

We show statistics for both full sample and sample of youth respondents for two reasons: First, we want to show how the subsample of youth respondents compares to the full sample in terms of characteristics. Second, we want to show that differences between boys and girls are very similar in both the full sample and the subsample (once we have controlled

¹⁰The 2000 Census, “XII Censo General de Poblacion y Vivienda”, can be found on the website of the Mexican statistical institute, INEGI.

for potential sample selection problems). Testing for gender differences in the youth sample requires to correct for potential sample selection problems, since the selection process might have been different for boys and girls. We use a standard Heckman selection correction and use variables that capture the timing of the interview as exclusion restriction (see section 3.4).

In the youth sample, around 27% of male and female high school graduates enrol in college (see columns 4 and 5). For the whole sample, attendance rates are slightly higher, around 30% for males and 33% for females (gender differences are not significant). Mexico has reached the stage where boys' and girls' educational attainment is very similar, while for example in the US girls have overtaken boys in terms of college attainment (see, e.g., Goldin, Katz, and Kuziemko (2006) and Bailey and Dynarski (2011)).

In what follows, we make use of data on GPA (grade point average between 0 and 100) of junior high school as a proxy for academic achievement. Table 2 shows that girls have a significantly higher GPA than boys (in both the total sample and the youth sample). This is consistent with empirical evidence for many countries that girls outperform boys at school (see, e.g., Goldin, Katz, and Kuziemko (2006)).

To control for parental education as one of the most important determinants of children's schooling choices, we use information on mothers' years of completed schooling in the form of four education dummies: for primary education, junior and senior high school and university. About 80% of mothers have only some primary education, while around 16% have attended junior high school, 3% senior high school and less than 1% of mothers have some university education. Mothers are slightly more likely to have a senior high school degree in the case of girls. Results are very similar for full sample and youth sample.

We create three per capita income categories, where the thresholds are equal to twice and four times the minimum wage.¹¹ About 45-50% 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. About 30% are in the second highest category (5000 to 10000 pesos) and the remaining 20-25% in the top income category. There are no significant gender differences (neither in the full nor youth sample).

We also control for the presence of the father in the household, that is whether the mother is single, separated, divorced or widowed, to see if this has an effect of children's schooling

¹¹Per capita parental income is constructed using parental income –such as parents' labor 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 (2014)).

decision (on top of a resource effect that we aim to control for directly by including measures of parental income divided by family size). In around 20-25% of households, no father is present. While this fraction is slightly higher in the youth sample, there are no significant gender differences in either sample.

Another potentially important factor in schooling decisions is the number of siblings. Boys have around 2.4 to 2.6 siblings, girls have a slightly larger number of siblings (2.7-2.8), but gender differences are not significant.

Turning to our proxies for marriage market returns, the ratio of unmarried men to women is 1.11 for boys and girls, while the ratio of unmarried men to women with at least a senior high school degree is 1.05 and slightly larger in the full sample and for girls (difference insignificant).

We can summarize the evidence from Table 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. The gender differences we find are very similar for the full sample and for the subsample of youth respondents, that is whenever we find significant differences for the full sample, we also find significant differences for the sample of youth respondents and vice versa. We now turn to analyze the differences in subjective expectations.

4.2 Descriptive Statistics: Subjective Expectations

In Table 3, we present summary statistics for the variables derived from the subjective expectations questions. 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 two different schooling degrees, senior high school and college. We summarize these measures separately for male and female youths and determine if there are significant differences between boys' and girls' expectations for the sample of youth respondents, while correcting for potential sample selection using a standard Heckman approach as discussed in section 3.4 and in the last section.

Not surprisingly, but reassuringly, expected log earnings increase in schooling level. Gross returns to college –measured as the difference between expected log earnings of the two consecutive schooling degrees, high school and college– are large (as they are in terms of realized earnings, see the following section 4.3). Male youths expect slightly higher earnings and lower returns consistent with a gender earnings gap that decreases in schooling level as observed in actual earnings data from Mexico (see section 4.3).

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 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 and similar probabilities of working at age 25.¹²

4.3 Data Validation: Expected Earnings and Individual Characteristics

The goal of this section is to present evidence that people are able to understand questions on their subjective expectations of earnings and give meaningful answers. For that purpose we first compare the data on subjective expectations we are using to earnings data from the Census. Second, we relate the expected value of future earnings to a number of individual and family background characteristics. Since the evidence provided in this section is meant to convince the reader of the validity of our data, but is not essential for the main goals of this paper, we defer the relevant tables to the Online Appendix.

As we will discuss in detail below, it is probably impossible to establish whether the elicited expectations are ‘rational’. But, in the following sense, it is not too important for the goals of this paper: for modeling education choices what matters is whether the elicited expectations reflect the subjective expectations that people base their decisions on. In that case data on people’s earnings expectations can help us improve our understanding of educational decisions. We therefore aim to convince the reader that people are indeed able to give meaningful answers to the questions on their earnings expectations and that the elicited expectations are able to capture –at least to some degree– the beliefs that people base their decisions on.

In the following analysis, we are particularly interested in the pattern for male versus

¹²In 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.

female youths. We compare earnings expectations to Census earnings using 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, if nothing else 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 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.

Appendix Table 7 presents mean expected earnings based on the distribution of earnings and Census earnings for different schooling degrees, for boys and girls respectively. It shows that boys’ and girls’ expectations about future earnings are clearly of the same order of magnitude as observed Census earnings. Both boys and girls expect future college earnings (that will be realized in 2012) to be higher than college earnings that are observed at the time of the Census in 2000. Girls also expect their future high school earnings to be slightly higher than female high school earnings that are observed in 2000, while boys believe them to be very similar to contemporaneous high school earnings. Boys and girls thus expect high future returns to college and they expect them to be larger than the ones observed in 2000. These beliefs are consistent with the continuation of previous trends of rising returns to college (driven by rising college earnings and stagnating high school earnings).

Girls expect lower high school earnings than boys and similar college earnings. Interestingly, the gender gap in expected future earnings is smaller than the gender gap that was observed in the Census in 2000, while expected and actual gender gap are decreasing in the level of education.

The second interesting exercise consists of relating the expected value of future earnings to a number of observable variables, such as those listed in the summary statistics table 2. Such an exercise serves the purposes of a validation exercise. In particular, 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. We refer the interested reader to Attanasio and Kaufmann (2008) for a more detailed discussion.¹³

Appendix Table 8 shows that both boys and girls expect higher earnings if they are from richer families. This might be due to the fact that richer families have networks that provide better job opportunities. Parental income might also proxy for skills of the child, such as cognitive and non-cognitive skills (e.g. motivation, perseverance etc) that are favorable to higher earnings. Interestingly, our measure of performance at school, GPA, appears important for girls, in that girls with higher GPA expect higher earnings (in particular for the case of having a college degree), while for boys there does not seem to be a correlation between GPA and expected earnings. Mothers' education is not a significant predictor of earnings expectations, though for boys the coefficient on having a college educated mother is large and positive. For boys we find that a larger number of siblings corresponds to higher expected high school earnings (note that parental income is held constant). One potential explanation could be that boys without siblings need to take up a specific job that is expected by the parents, e.g. taking over the parents' small scale business, while boys with siblings have more freedom.

Our analysis shows that there is a considerable amount of heterogeneity in expectations. This still holds after controlling for individual and family background characteristics, which reflects the importance of unobserved heterogeneity in terms of perceptions about cognitive and social skills and differences in information sets e.g. about skill prices.¹⁴

To conclude, the evidence of this section suggests that our sample of Mexican youths understand the expectation questions and give meaningful answers. Both boys and girls expect high returns to college in the labor market and perceive that unemployment and earnings risk decrease with higher education.

Even though the goal of this section was not to show how rational or well-informed people are (or how people form expectations), but only to give the reader an idea about how expectations compare to observed earnings and how they correlate with individual characteristics, we can tentatively conclude the following: We do not find evidence that girls are less well-informed about potential labor market outcomes. First, girls' expectations are not more noisy than boys' expectations. Second, girls' expectations about earnings are of

¹³It is important to keep in mind that our goal is not to explain how people form expectations. Our data clearly does not allow us to conduct such an ambitious exercise given the fact that we have do not have panel data of expectations and only small sample sizes.

¹⁴An alternative explanation is that the remaining "heterogeneity" reflects noise. But we show that subjective expectations are able to predict schooling choices even after controlling for an extensive set of individual and family background characteristics. This suggests that at least part of the heterogeneity captures factors unobserved to the researcher, such as skills and information about skill prices, which influence earnings expectations.

a similar order of magnitude as observed earnings (as is the case for boys). Third, girls' expectations vary with characteristics in a way one would expect.

At the same time there are some interesting differences in the way expectations vary with characteristics when comparing boys and girls. For example, girls who performed better at school expect higher earnings (in particular with a college degree), i.e. girls believe that performance at school is rewarded in the labor market, while this is not the case for boys. Such findings point towards potential differences in the way that girls and boys form expectations, which is a question we plan to investigate further in future research (with more appropriate data, that is repeated observations of expectations and a larger sample size). This illustrates that direct data on subjective expectations can be useful, since it would be difficult to predict ex-ante how girls and boys differ in their information sets and the way they form expectations (while it would be necessary to make such assumptions when using earnings realizations to infer people's expectations).

5 Schooling Decisions and Returns to Schooling

In this Section, we show how college attendance decisions of young poor Mexicans relate to their expectations about labor market returns and risks as well as to returns in the marriage market. We are particularly interested in whether the determinants of education choices differ between boys and girls.

We proceed as follows: First, we relate the college attendance decision to expected labor market outcomes, that is monetary returns to college and to perceived risk of unemployment and earnings. Second, we show how the decision relates to our two proxies of marriage market returns for boys and girls. In a third step, we regress the college decision jointly on both determinants, labor market returns and marriage market returns. This allows to assess in how far labor market expectations might have captured marriage market considerations or whether the two 'marriage market' proxies have picked up labor market effects.

We estimate the schooling decisions jointly for boys and girls and interact all regressors with gender dummies to study gender differences. In the tables, we present the estimation results in two separate columns. As discussed in section 4.2, we address the potential sample selection problem by using a Heckman selection correction in a non-linear context, that is we estimate jointly a selection equation and the main college attendance equation. The model is identified by variables that capture the date and time of the interview, which are strong predictors for whether the youth is present at the interview to answer the expectations questions (see Table 1 discussed in section 4.2).

5.1 Labor Market Returns

While the goal is to model the college attendance decision using equation (1) in the following section, we first relate the probability of college enrolment to labor market outcomes, that is expected returns to college and perceived unemployment and earnings risk (under the two different schooling scenarios) and a set of controls. In the following tables we present estimates for boys and girls in separate columns, even though they are estimated together in one model where all regressors are interacted with gender dummies to test for gender differences in coefficients.

Table 4 shows that –in terms of individual characteristics– academic performance, as measured by the GPA, is an important determinant of the decision to attend college for both boys and girls. Past academic performance is a measure of the psychological costs or benefits of getting further education and also captures the likelihood of being able to complete the following schooling levels (compare Stinebrickner and Stinebrickner (2012)). Interestingly, having a highly educated mother is more important for girls’ decision to attend college than for boys, while boys are more likely to go to college if there is no father in the household. Also parental income appears to be more important for girls than boys in the decision to attend college.

The key results of Table 4 are as follows: Boys’ expectations about returns to college are a strongly significant predictor for their decision to enrol in college (significant on 1%). Girls’ expectations about returns are significant as well, but only on 10%. 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 (nor is it the case that girls’ expectations are measured in a more noisy way, as shown in Section 4.3). Instead the coefficient on girls’ expected returns is only about a third of the size of boys’ coefficient and significantly smaller.

Perceptions about unemployment or earnings risk are not significant in the college enrolment decision. Attanasio and Kaufmann (2009) show that risk perceptions are important in the high school decision, suggesting that this result is not driven by risk perceptions being measured in a too noisy way, but that risk might be less important in college decisions.

As we will discuss in detail in Section 5.3, it is unlikely that these results are 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 (we never reject that the correlation between the error terms of selection equation and main equation is zero).

To summarize, our results suggest that in the context of college enrolment decisions, there are important gender differences in the importance of labor market returns as determinants of school choices.

5.2 Marriage Market Returns

In the last section, we found differences in the role of expected labor market returns between boys versus girls. In what follows, we want to explore further how determinants of schooling decisions differ for boys and girls, and consider a second potentially very important determinant of schooling: returns in the marriage market.

As discussed in the introduction, there is surprisingly little evidence on the importance of this determinant of schooling, despite the fact that finding an appropriate partner is certainly one of the big challenges individuals face during their lifetime (for example because, for many people, the relationship with their romantic partner is more long-term than job relationships and the quality of the partner affects the quality of the offspring etc). Thus, one would expect that individuals assign a high value to being in an environment where it is more likely to find a high quality partner. In this section, we aim to shed light on the importance of marriage market returns in the decision to enrol in college.

In the following, we first relate the college decision to our two proxies of marriage market returns and controls (see the upper panel of Table 5). Then we investigate jointly the importance of labor market and marriage market considerations for boys and girls, as modeled in equation (1) (see the lower panel of Table 5). We report estimates using two different proxies for marriage market returns, the sex ratio conditional on education in columns 1 (for boys) and 2 (for girls) and not conditional on education in columns 3 and 4 (for boys and girls, respectively).¹⁵

The upper panel of Table 5 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 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.

Columns 3 and 4 show the same pattern for the proxy without conditioning on education (see the upper panel of Table 5). 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 enrol in college. The coefficient on the ratio for boys on the other hand is only marginally significant (many men per women decrease the likelihood to go to college) and significantly different from the coefficient for girls (on 3%).

The results of the upper panel of Table 5 suggest that marriage market considerations

¹⁵While we include the same control variables as in Table 4 in all the following regression tables, we do not display the coefficients on those controls in the main paper due to space constraints, but defer the tables with the full set of coefficients to the Online Appendix.

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. One key concern is that—in addition to marriage market effects—the sex ratio in the locality of residence (conditional on education or unconditionally) captures labor market conditions which in turn affect individuals' decisions to attend college. In this context, our data on expectations about labor market outcomes proves very useful, since what matters for an individual's school choice is her beliefs about her own potential labor market outcomes (conditional on different schooling levels), and we can directly control for these beliefs.

The lower panel of Table 5 therefore combines the analysis of the role of labor market and marriage market outcomes in college attendance decisions and confirms our previous two key findings (see Table 4 and the upper panel of Table 5). In terms of labor market returns, the coefficient on expected returns to college is again highly significant for boys and the magnitude of the coefficient is basically unchanged. Also the coefficient on girls' expected return is unchanged and still significantly smaller in magnitude than for boys. In terms of marriage market considerations, both proxies are still significant predictors of girls' college attendance decision also after controlling for labor market expectations. For boys, both proxies are now insignificant and still significantly different from the ones' for girls.

These key findings remain unchanged if we control not only for individuals' beliefs about their idiosyncratic labor market returns and risks, but also for local labor market conditions in general (i.e. average local earnings for high school and for college educated individuals). The lower panel of Table 6 also shows that—after controlling for individuals' beliefs about their own potential labor market outcomes—coefficients on local labor market conditions are as expected in terms of sign (negative for high school earnings, positive for college earnings for boys), but not significant.

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 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 (so that this ratio should not capture any peer effects with respect to 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 our marriage market proxy.¹⁶

¹⁶When constructing our control variable for potential peer effects, we consider both married and unmar-

Tables 5 and 6 show that results point in the same direction for both proxies conditional on education and unconditionally: for girls, the coefficients on our marriage market proxies are always significant and significantly different from the ones of boys.

Controlling for potential peer effects directly, that is controlling for the fraction of women (men) in the locality who have a college degree, our results remain unchanged. The upper panel of Table 6 shows that the marriage market proxy conditional on education actually becomes larger and more strongly significant 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 suggest that the correlation we find between schooling decisions and our marriage market proxies is not driven peer effects in education.

It is also important to note that –in contrast to some other papers in the literature–, we construct our proxies for marriage market returns using a relatively small geographic area, that is the locality level, because we believe that this is the appropriate reference area (marriage “market”) for most youths. This had the advantage that our proxy is more likely to capture people’s actual knowledge about availability of partners than when using a larger reference area.

To summarize, our results suggest that there are important gender differences in the importance of labor market and marriage market returns as determinants of college attendance decisions. 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 an important 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 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

ried women (men) and only take into account women (men) who went to college, since we want to control for potential peer effects in the college decision arising from observing others going to college.

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.

We think that ex-post rationalization of college enrolment, which would imply that the main reason to attend college is its consumption value (or similar reason, see discussion on classical endogeneity below) rather than expected future returns, is unlikely for poor Mexican youths. One could think that such an explanation might ring true for rich or even middle class individuals, but not for individuals from extremely poor families for whom college enrolment is often extremely difficult from a financial point of view. However, 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 cohort of interest (i.e. students who had just graduated from senior high school before the survey) with a cohort that is one year younger (and thus just starting grade 12). 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 12.¹⁷ The same reasoning holds for perceptions of the probability of working and perceptions of earnings risk.

To test for differences between the distribution of expectations of those two adjacent cohorts, we use a Kolmogorov-Smirnov test, which is a nonparametric test for the equality of continuous, one-dimensional probability distributions. The null hypothesis is that the cross-sectional distribution of -for example- expected returns is the same for the sample of senior high school graduates and the sample of a cohort that is one year younger and just starting grade 12.

In Appendix Table 13 we report p-values of Kolmogorov-Smirnov tests 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.

Appendix Table 13 illustrates that the distribution of expectations is the same for the

¹⁷As discussed in section 3.2, 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 people with *low* expected returns to college decide to enrol in college and now state high returns to college and vice versa.

senior high school graduates themselves compared to the cohort that is one year younger. We never reject equality of distributions apart from one instance, that is girls' perceptions about the probability of working with a senior high school degree. In that case the distribution of the older cohort is not more spread out, but some of the older girls seem to have updated their beliefs about probability of working upward compared to the one-year younger cohort (see Appendix Figure 1). At the same time, this is not driving any of our results because in any case girls' perceptions about the probability of working are never significant in the college attendance choice regression. The other instance where we come close to rejecting equality of distributions is for girls' expectations about college earnings (p-val 0.112). In that case the distribution of the older cohort of senior high school graduates is actually more compressed (instead of more spread out), contrary to what one should expect in the case ex-post justification (see Appendix Figure 2).¹⁸

Classical Endogeneity

A second concern with using subjective expectations as determinants of schooling choices arises from another form of endogeneity: the one that is due to omitted variables, such as, for example, unobserved taste heterogeneity. Individuals who have high expected returns and go to college more often might also be the ones who have high preferences for being in college or attach weight to other outcomes, which are not measured, to which college is conducive. While we have discussed above the concern that individuals go to college because of the consumption value of college, in the following we discuss the concern that individuals might go to college because they expect returns other than labor market returns. 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 of course are returns in the marriage market.

Can this phenomenon explain our main results? We have shown in the last section that are two main results are unchanged when we include both determinants, labor and marriage market concerns, jointly in our model of college choice. Both measures we use, data on people's subjective expectations and two different marriage market proxies, appear to be capturing the labor market and marriage market concerns that we want to measure (where, in the case of boys, labor market returns appear most important, while in the case of girls marriage market considerations appear important as well).

To argue that instead of labor market and marriage market considerations, a third omitted determinant is driving college attendance decisions of boys and girls, one would have to not only come up with such a determinant (which is as or more important than labor

¹⁸We illustrate graphically only the two instances, in which distributions are (close to) significantly different. The other graphs can be obtained from the authors upon request.

and marriage market considerations), but would also need to explain why this third factor is correlated with expected returns and marriage market proxy for girls (while our two variables 'expected returns' and 'proxy' are themselves basically uncorrelated), but only with expected returns for boys. Or else, there have to be two other determinants affecting boys and girls differentially (in which case we are back to gender differences in determinants). We find it hard to think of determinants other than labor market returns, marriage market returns and preferences, that are more important than these three determinants that are discussed most extensively in the literature.¹⁹

To conclude, in this section we provided evidence and discussed in detail why we believe that endogeneity concerns do not affect our main conclusions and cannot be driving these results: there are significant differences in determinants of the college attendance decisions between boys and girls.

6 Conclusion

The results of this paper speak to several important questions about the determinants of investment in human capital. In particular, we have analyzed how expectations of future returns affect schooling decisions of Mexican youths. 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 and the returns on the marriage market.

Our main result is that determinants of schooling decisions differ between boys and girls. While expected labor market outcomes, in particular returns to college, are very important for boys, they seem to be less important for girls. Girls on the other hand appear to care about marriage market considerations in their decision to attend college, consistent with anecdotal evidence.

Direct data on people's beliefs enable us to be agnostic about differences in information sets and ways of forming expectations between boys and girls. Avoiding strong assumptions about how people form expectations and what information they possess is crucial for our analysis, as results on gender differences in determinants of schooling might be very sensitive to these assumptions.

Unfortunately, our data does not contain direct questions on the returns to education on the marriage market. For this reason, we are forced to use some indirect proxies for such a

¹⁹While there is a lot of evidence on the importance of labor market returns and indirect evidence on preferences, there is much less evidence on marriage market returns, even though they can be of first-order importance, for example for the next generation, as shown in a recent paper by Kaufmann, Messner, and Solis (2015).

return. In this sense, our evidence on marriage market considerations might be interpreted as only suggestive and subject to a number of caveats. However, such evidence, at the very least, warrants further research on this issue, which is potentially very important and which has not received much attention in the literature so far.

Our results have important policy implications for the design of programs aiming at increasing schooling, such as conditional cash transfer programs, fellowship programs, information campaigns etc. For the effective design of such programs, it is indispensable to understand what are the key determinants of schooling choices and whether there are gender differences in the role of these determinants.

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.

References

- ABRAMITZKY, R., AND V. LAVY (2014): “How Responsive is Investment in Schooling to Changes in Redistribution Policies and in Returns?,” *Econometrica*, 82(4), 1241–1272.
- ALVAREZ, B., AND M. VERA-HERNANDEZ (2013): “Exploiting Subjective Information to Understand Impoverished Children’s Use of Health Care,” *The Journal of Health Economics*, 32(6), 1194–1204.
- 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*.
- (2009): “Educational Choices, Subjective Expectations and Credit Constraints,” *NBER Working Paper 15087*, July.
- (2014): “Education Choices and Returns to Schooling: Mothers’ and Youths’ Subjective Expectations and their Role by Gender,” *Journal of Development Economics*.
- 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*.
- AZMAT, G., M. GUELL, AND A. MANNING (2006): “Gender Gaps in Unemployment Rates in OECD Countries,” *Journal of Labor Economics*, 24(1).
- AZMAT, G., AND B. PETRONGOLO (2014): “Gender and the labor market: What have we learned from field and lab experiments?,” *Labor Economics*, 30, 32–40.
- BAILEY, M. J., AND S. M. DYNARSKI (2011): “Gains and gaps: changing inequality in U.S. college entry and completion,” *NBER Working Paper 17633*, NBER Working paper 17633.
- BOOTH, A., AND P. NOLEN (2012): “Choosing to compete: How different are girls and boys?,” *Journal of Economic Behavior and Organization*, 81(2), 542–555.

- BOTTICINI, M., AND A. SIOW (2008): “Are there Increasing Returns in Marriage Markets?,” *University of Toronto, Department of Economics, Working Paper 333*.
- 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.
- CASTILLO, M., R. PETRIE, M. TORERO, AND L. VESTERLUND (2013): “Gender Differences in Bargaining Outcomes: A Field Experiment on Discrimination,” *Journal of Public Economics*, 2013, 99, 435–482.
- CHARNESS, G., AND U. GNEEZY (2012): “Strong Evidence for Gender Differences in Risk Taking,” *Journal of Economic Behavior and Organization*, 83(1), 50–58.
- DELAVANDE, A., X. GINÉ, AND D. MCKENZIE (2011): “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 (2014): “Investing in Schooling in Chile: The Role of Information about Financial Aid For Higher Education,” *The Review of Economics and Statistics*, 96(2), 244–257.
- DOHMEN, T., AND A. FALK. (2011): “Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender,” *American Economic Review*, 101(2), 556–90.
- DOHMEN, T., A. FALK, D. HUFFMAN, F. MARKLEIN, AND U. SUNDE (2009): “Biased probability judgment: Evidence of Incidence and Relationship to Economic Outcomes from a Representative Sample,” *Journal of Economic Behavior and Organization*, 72(3), 903–915.
- DOMINITZ, J., AND C. F. MANSKI (1996): “Eliciting Student Expectations of the Returns to Schooling,” *Journal of Human Resources*, 31 (1), 1–26.
- DRERUP, T., B. ENKE, AND H.-M. VON GAUDECKER (2014): “Measurement Error in Subjective Expectations and the Empirical Content of Economic Models,” *Mimeo*.
- EDLUND, L. (2005): “Sex and the City,” *Scandinavian Journal of Economics*, 107(1), 25–44.
- FILIPPIN, A., AND A. ICHINO (2005): “Gender Wage Gap in Expectations and Realizations,” *Labour Economics*, 12(1), 125–145.

- GIUSTINELLI, P. (forthcoming): “Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choices of High School Tracks,” *International Economic Review*.
- GOLDIN, C. (1992): “The Meaning of College in the Lives of American Women: The Past One-Hundred Years,” *NBER Working Paper No. 4099*.
- 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.
- GUIISO, 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.
- GUPTA, N. D., A. POULSEN, AND M. C. VILLEVAL (2013): “Gender matching and competitiveness: experimental evidence,” *Economic Inquiry*, 51(1), 816–835.
- HAMERMESH, D. S., AND J. E. BIDDLE (1994): “Beauty and the Labor Market,” *The American Economic Review*, 84(5), 1174–1194.
- 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.
- JOENSEN, J. S., AND H. NIELSEN (2015): “Mathematics and Gender: Heterogeneity in Causes and Consequences,” *The Economic Journal*.
- KAUFMANN, K. M. (2014): “Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns,” *Quantitative Economics, The Econometric Society*.
- KAUFMANN, K. M., M. MESSNER, AND A. SOLIS (2012): “Returns to Elite Higher Education in the Marriage Market: Evidence from Chile,” *SSRN Working Paper*.
- (2015): “Elite Higher Education, the Marriage Market and the Intergenerational Transmission of Human Capital,” *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*.
- LEE, S. (2015): “Marriage and Online Mate-Search Services: Evidence From South Korea,” *Journal of Applied Econometrics*.
- LOCHNER, L. (2007): “Individual Perceptions of the Criminal Justice System,” *American Economic Review*, 97(1), 444–460.
- MAHAJAN, A., AND A. TAROZZI (2011): “Time Inconsistency, Expectations and Technology Adoption: The Case of Insecticide Treated Nets,” *mimeo, Stanford University and Duke University*.
- 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.
- MENGEL, F. (2015): “Gender differences in networking,” *Mimeo*.
- NGUYEN, T. (2008): “Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar,” *mimeo*.
- OREOPOULOS, P., AND K. G. SALVANES (2011): “Priceless: The Nonpecuniary Benefits of Schooling,” *The Journal of Economic Perspectives*.
- STINEBRICKNER, T., AND R. STINEBRICKNER (2012): “Learning about Academic Ability and the College Drop-Out Decision,” *Journal of Labor Economics*, 30(4), 707–748.
- (2014): “A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout,” *The Review of Economic Studies*, 81(1), 426–472.
- 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.
- ZAFAR, B. (2009): “College Major Choice and the Gender Gap,” *Federal Reserve Bank of New York Staff Reports*.

Appendix

Table 1: First-stage Regression for Whether the Youth Responds Herself

Dep Var:	Youth Respondent: Yes/No		
	Instruments 1 Coeff/(SE)	Instruments 2 Coeff/(SE)	Instr and Controls Coeff/(SE)
Interview Evening of Weekday	-0.284* (0.155)	-0.344* (0.177)	-0.295* (0.176)
Interview Evening	0.280* (0.143)	0.365** (0.157)	0.320** (0.156)
Interview Week 40 to 42	0.126*** (0.029)	0.175*** (0.039)	0.168*** (0.039)
Male	-0.091*** (0.028)	-0.048 (0.036)	0.305 (0.276)
Interv Evening Weekday * Male		0.340 (0.401)	0.273 (0.399)
Interv Evening * Male		-0.404 (0.384)	-0.334 (0.381)
Interv Week 40 to 42 * Male		-0.106* (0.059)	-0.073 (0.059)
GPA of Jr HS (0-100) *Male			-0.002 (0.003)
GPA * Female			0.003* (0.002)
Number of Siblings * Male			-0.038*** (0.013)
Number of Siblings * Female			-0.028** (0.011)
No Father in Household * Male			0.071 (0.052)
No Father in Household * Female			0.019 (0.046)
Mother's Educ - Jr HS * Male			0.001 (0.057)
Mother's Educ - Jr HS * Female			-0.083* (0.050)
Per cap Income - below 5k * Male			0.009 (0.054)
Per cap Income - below 5k * Female			0.125** (0.049)
Per cap Income - above 10k * Male			-0.041 (0.056)
Per cap Income - above 10k * Female			-0.025 (0.049)
Sex Ratio (Sr HS) * Male			0.058* (0.035)
Sex Ratio (Sr HS) * Female			0.001 (0.030)
Observations	1237	1237	1237
F-Statistic	8.270	5.414	3.425
R-Squared	0.026	0.030	0.061
Adjusted R-Squared	0.023	0.024	0.043

Notes: Table displays coefficients and standard errors in brackets. * p<0.1 ** p<0.05 *** p<0.01. Excl. categories: mother's education primary or less, per capita income less than 5000 pesos. Dummies for mother's education senior high school and college are included, but not shown due to space constraints (insignificant).

Table 2: Summary Statistics of Individual and Family-Background Characteristics

Sample	Total			Youth		
	Boys Mean/(SD)	Girls Mean/(SD)	Diff (P-Val)	Boys Mean/(SD)	Girls Mean/(SD)	Diff (P-Val)
Enrollment Rate: College	0.307 (0.462)	0.333 (0.472)	(0.326)	0.268 (0.444)	0.267 (0.443)	(0.707)
GPA of Jr HS (Scale 0-100)	80.272 (8.221)	82.885 (9.688)	(0.000)	80.172 (6.422)	83.462 (8.038)	(0.000)
Mother's Educ - Primary	0.814 (0.400)	0.787 (0.413)	(0.286)	0.822 (0.405)	0.804 (0.407)	(0.884)
Mother's Educ - Jr HS	0.163 (0.369)	0.171 (0.383)	(0.697)	0.164 (0.371)	0.151 (0.358)	(0.133)
Mother's Educ - Sr HS	0.018 (0.134)	0.039 (0.194)	(0.026)	0.007 (0.086)	0.040 (0.195)	(0.025)
Mother's Educ - Univ	0.005 (0.074)	0.003 (0.054)	(0.492)	0.007 (0.086)	0.005 (0.070)	(0.848)
Per Cap Income - 5 to 10k	0.327 (0.470)	0.307 (0.462)	(0.453)	0.309 (0.463)	0.287 (0.453)	(0.907)
Per Cap Income - above 10k	0.247 (0.432)	0.222 (0.416)	(0.302)	0.264 (0.442)	0.210 (0.408)	(0.537)
No Father in Household (Mother Single/Sep/Div)	0.216 (0.412)	0.209 (0.407)	(0.764)	0.253 (0.435)	0.220 (0.415)	(0.222)
Number of Siblings	2.660 (1.827)	2.828 (1.798)	(0.107)	2.401 (1.758)	2.715 (1.773)	(0.237)
Ratio Unmarried Men/Women	1.109 (0.204)	1.109 (0.232)	(0.988)	1.110 (0.211)	1.109 (0.236)	(0.375)
Ratio Unmarried Men/Women (With Sen HS)	1.057 (0.539)	1.074 (0.633)	(0.615)	1.025 (0.555)	1.067 (0.670)	(0.570)
Observations	547	690		269	404	

Notes: The first three columns (“Total”) refer to the whole sample including youth and mother respondents. In our analysis we will be using the “Youth” sample (columns 4 to 6). The difference between boys’ and girls’ characteristics in column 6 corrects for potential sample selection using a standard Heckman selection correction. As exclusion restriction we use variables that capture the timing of the interview, which are strong predictors for whether the youth is present when the interviewer arrives to answer the questions on subjective expectations herself.

Table 3: Summary Statistics of Subjective Expectations of Future Earnings

	Boys	Girls	Diff
	Mean/(Std Dev)		(P-Val)
Exp Log Earnings			
- Senior HS	7.633 (0.471)	7.532 (0.516)	(0.199)
- College	8.288 (0.461)	8.272 (0.496)	(0.605)
Exp Return			
- College	0.655 (0.359)	0.735 (0.436)	(0.025)
Std Dev of Log Earn			
- Senior HS	0.064 (0.041)	0.062 (0.038)	(0.534)
- College	0.052 (0.032)	0.054 (0.034)	(0.874)
Prob of Work			
- Senior HS	0.676 (0.180)	0.670 (0.189)	(0.844)
- College	0.817 (0.174)	0.823 (0.156)	(0.516)
Observations	269	404	

Notes: The difference between boys' and girls' expectations in column 3 corrects for potential sample selection using a standard Heckman selection correction. As exclusion restriction we use variables that capture the timing of the interview, which are strong predictors for whether the youth is present when the interviewer arrives to answer the questions on subjective expectations herself.

Table 4: College Attendance Choice: Labor Market

Dependent Variable	College Attendance	
	Labor Market Expect Boys	Girls
Expected Return - College	0.784*** (0.247)	0.274* (0.162)
Prob of Work - Sr HS	0.077 (0.590)	0.444 (0.476)
Prob of Work - College	-0.484 (0.617)	0.38 (0.574)
Log Var of Earn - Sr HS	6.331 (11.767)	14.554 (10.475)
Log Var of Earn - College	2.855 (18.996)	-3.734 (13.975)
GPA of Jr HS (0-100)	0.045*** (0.010)	0.036*** (0.010)
No Father in Household	0.517** (0.238)	-0.198 (0.176)
Mother's Educ - Jr HS	0.189 (0.233)	0.109 (0.199)
Mother's Educ - Sr HS	-4.402 (616.588)	0.988*** (0.360)
Mother's Educ - Univ	-5.131 (572.811)	6.723 (507.200)
Per cap Income - below 5k	0.336 (0.217)	-0.249 (0.188)
Per cap Income - above 10k	-0.071 (0.232)	0.441** (0.178)
Observations	1237	
Censored Obs	564	
Log Likelihood	-1169.079	
Sample Sel: Corr of Errors (P-Val)	-0.213 (0.686)	

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.

Table 5: College Attendance Choice: Labor Market and Marriage Market

Dependent Variable	College Attendance Decision			
	Marriage Market Proxy 1		Marriage Market Proxy 2	
	Boys	Girls	Boys	Girls
Without Exp Labor Market Returns				
Ratio of Unmarried Men to Women (Sr HS)	-0.150 (0.138)			
Ratio of Unmarried Women to Men (Sr HS)		0.184* (0.099)		
Ratio of Unmarried Men to Women			-0.747* (0.452)	
Ratio of Unmarried Women to Men				0.575** (0.279)
Indiv. and Fam. Background Controls		Yes		Yes
Observations		1237		1237
Censored Obs		564		564
Log Likelihood		-1166.771		-1166.723
Sample Sel: Corr of Errors (P-Val)		-0.339 (0.476)		-0.400 (0.373)
With Exp Labor Market Returns				
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)
Indiv. and Fam. Background Controls		Yes		Yes
Observations		1237		1237
Censored Obs		564		564
Log Likelihood		-1156.434		-1156.332
Sample Sel: Corr of Errors (P-Val)		-0.416 (0.357)		-0.471 (0.278)

Notes: Table displays coefficients and standard errors in brackets. * p<0.1 ** p<0.05 *** p<0.01. Coefficients on controls are not displayed due to space constraints (see Online Appendix for tables with the full set of coefficients).

Table 6: College Attendance Choice: Labor and Marriage Market - Local Labor Market Conditions

Dependent Variable	College Attendance Decision			
	Marriage Market Proxy 1 Boys	Marriage Market Proxy 1 Girls	Marriage Market Proxy 2 Boys	Marriage Market Proxy 2 Girls
Potential Peer Effects				
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			-0.500 (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)
Fraction of College Educated Men (Women)	1.440 (1.228)	3.491*** (1.225)	1.309 (1.223)	2.958** (1.222)
Indiv. and Fam. Background Controls		Yes		Yes
Controls Unemployment and Earnings Risk		Yes		Yes
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)
Local Labor Market Conditions				
Ratio of Unmarried Men to Women (Sr HS)	-0.038 (0.146)			
Ratio of Unmarried Women to Men (Sr HS)		0.223* (0.127)		
Ratio of Unmarried Men to Women			-0.062 (0.597)	
Ratio of Unmarried Women to Men				0.647* (0.361)
Expected Return - College	0.923*** (0.275)	0.451** (0.195)	0.917*** (0.278)	0.466** (0.195)
Average Log Census Earnings - Sr HS	-0.313 (0.347)	0.213 (0.260)	-0.338 (0.345)	0.190 (0.263)
Average Log Census Earnings - College	0.424 (0.331)	0.089 (0.302)	0.386 (0.331)	-0.050 (0.294)
Fraction of College Educated Men (Women)	1.486 (1.319)	2.565* (1.396)	1.566 (1.321)	2.116 (1.371)
Indiv. and Fam. Background Controls		Yes		Yes
Controls Unemployment and Earnings Risk		Yes		Yes
Observations		1237		1237
Censored Obs		564		564
Log Likelihood		-946.466		-946.900
Sample Sel: Corr of Errors (P-Val)		-0.009 (0.986)		-0.029 (0.952)

Notes: Table displays coefficients and standard errors in brackets. * p<0.1 ** p<0.05 *** p<0.01 .Coefficients on controls are not displayed due to space constraints. Perceptions of unemployment and earnings risk are included in all specifications but not displayed due to space constraints, see Online Appendix for tables with the full set of coefficients.

ONLINE APPENDIX (NOT FOR PUBLICATION)

Table 7: Expected Earnings and Observed Census Earnings

	Boys		Girls	
	Exp Earn Mean/(SD)	Return Mean/(SD)	Exp Earn Mean/(SD)	Return Mean/(SD)
Exp Log Earnings				
- Senior HS	7.633 (0.471)		7.532 (0.516)	
- College	8.288 (0.461)	0.655 (0.358)	8.272 (0.496)	0.734 (0.436)
Log Census Earnings				
- Senior HS	7.661 (0.302)		7.462 (0.323)	
- College	8.140 (0.309)	0.486 (0.375)	8.048 (0.252)	0.599 (0.407)
Observations	269		404	

Table 8: Correlation between Expectations and Individual and Family Background Characteristics

Dependent Variable	Expected Log Earnings			
	Boys		Girls	
	High School	College	High School	College
GPA of Junior HS (0-100)	-0.007 (0.004)	-0.004 (0.005)	0.005 (0.003)	0.006* (0.003)
Number of Siblings	0.051* (0.031)	0.031 (0.033)	0.020 (0.017)	0.026 (0.017)
Per cap Income - 5 to 10k	0.092 (0.073)	0.062 (0.079)	0.135** (0.068)	0.084 (0.065)
Per cap Income - more than 10k	0.161** (0.082)	0.130 (0.089)	0.179** (0.079)	0.099 (0.076)
Mother's Educ - Jr HS	-0.031 (0.077)	0.007 (0.084)	-0.113 (0.074)	-0.084 (0.071)
Mother's Educ - Sr HS	-0.398 (0.376)	-0.348 (0.389)	-0.078 (0.132)	0.030 (0.126)
Mother's Educ - Univ	0.235 (0.343)	0.592 (0.384)	-0.097 (0.393)	0.099 (0.375)
Observations	547	547	690	690
Cens. obs.	278	278	286	286
Chi-Square	13.089	8.848	11.972	8.739
Inverse Mills Ratio	0.093	0.410	0.118	-0.012
S.E. of Inv Mills	0.371	0.402	0.204	0.195

Notes: Table displays coefficients and standard errors in brackets. * p<0.1 ** p<0.05 *** p<0.01. Excl. categories: mother's education primary or less, lowest per capita parental income category.

Table 9: College Attendance Choice: Marriage Market

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)	-0.15 (0.138)			
Ratio of Unmarried Women to Men (Sr HS)		0.184* (0.099)		
Ratio of Unmarried Men to Women			-0.747* (0.452)	
Ratio of Unmarried Women to Men				0.575** (0.279)
GPA of Jr HS (0-100)	0.043*** (0.010)	0.038*** (0.010)	0.051*** (0.011)	0.035*** (0.010)
No Father in Household	0.474** (0.222)	-0.215 (0.172)	0.472** (0.222)	-0.233 (0.171)
Mother's Educ - Jr HS	0.200 (0.226)	0.107 (0.191)	0.173 (0.225)	0.170 (0.185)
Mother's Educ - Sr HS	-4.748 (3429.514)	0.950*** (0.359)	-5.845 (1.7e+04)	0.946*** (0.357)
Per cap Income - below 5k	0.307 (0.219)	-0.239 (0.192)	0.315 (0.219)	-0.278 (0.186)
Per cap Income - above 10k	-0.02 (0.231)	0.410** (0.178)	0.011 (0.230)	0.416** (0.176)
Observations		1237		1237
Censored Obs		564		564
Log Likelihood		-1166.771		-1166.723
Sample Sel: Corr of Errors (P-Val)		-0.339 (0.476)		-0.400 (0.373)

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).

Table 10: College Attendance Choice: Labor Market and Marriage Market

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)	-0.147 (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)
GPA of Jr HS (0-100)	0.046*** (0.011)	0.033*** (0.010)	0.053*** (0.012)	0.031*** (0.010)
Number of Siblings	0.056 (0.060)	0.021 (0.046)	0.07 (0.059)	0.030 (0.044)
No Father in Household	0.510** (0.232)	-0.194 (0.172)	0.505** (0.233)	-0.211 (0.171)
Mother's Educ - Jr HS	0.140 (0.230)	0.063 (0.193)	0.129 (0.228)	0.129 (0.186)
Mother's Educ - Sr HS	-4.475 (1877.176)	0.904** (0.362)	-7.333 (2.8e+04)	0.900** (0.358)
Per cap Income - below 5k	0.280 (0.221)	-0.275 (0.189)	0.285 (0.220)	-0.315* (0.183)
Per cap Income - above 10k	-0.024 (0.233)	0.451** (0.179)	0.006 (0.232)	0.458*** (0.178)
Observations		1237		1237
Censored Obs		564		564
Log Likelihood		-1156.434		-1156.332
Sample Sel: Corr of Errors (P-Val)		-0.416 (0.357)		-0.471 (0.278)

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).

Table 11: College Attendance Choice: Labor and Marriage Market - Potential Peer Effects

Dependent Variable	College Attendance Decision			
	Marriage Market Proxy 1 Boys	Marriage Market Proxy 1 Girls	Marriage Market Proxy 2 Boys	Marriage Market Proxy 2 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 12: College Attendance Choice: Labor and Marriage Market - Local Labor Market Conditions

Dependent Variable	College Attendance Decision			
	Marriage Market Proxy 1 Boys	Marriage Market Proxy 1 Girls	Marriage Market Proxy 2 Boys	Marriage Market Proxy 2 Girls
Ratio of Unmarried Men to Women (Sr HS)	-0.038 (0.146)			
Ratio of Unmarried Women to Men (Sr HS)		0.223* (0.127)		
Ratio of Unmarried Men to Women			-0.062 (0.597)	
Ratio of Unmarried Women to Men				0.647* (0.361)
Expected Return - College	0.923*** (0.275)	0.451** (0.195)	0.917*** (0.278)	0.466** (0.195)
Prob of Work - Sr HS	0.038 (0.653)	0.638 (0.553)	0.021 (0.651)	0.663 (0.557)
Prob of Work - College	-0.428 (0.702)	0.037 (0.680)	-0.408 (0.700)	0.099 (0.679)
Log Var of Earn - Sr HS	4.778 (13.758)	13.543 (11.672)	4.463 (13.738)	17.497 (11.606)
Log Var of Earn - College	-0.797 (20.421)	-5.396 (17.938)	-0.589 (20.541)	-5.130 (18.014)
Average Log Census Earnings - Sr HS	-0.313 (0.347)	0.213 (0.260)	-0.338 (0.345)	0.190 (0.263)
Average Log Census Earnings - College	0.424 (0.331)	0.089 (0.302)	0.386 (0.331)	-0.050 (0.294)
Fraction of College Educated Men (Women)	1.486 (1.319)	2.565* (1.396)	1.566 (1.321)	2.116 (1.371)
GPA of Jr HS (0-100)	0.059*** (0.015)	0.038*** (0.012)	0.058*** (0.015)	0.039*** (0.012)
Number of Siblings	0.053 (0.069)	-0.021 (0.054)	0.052 (0.069)	-0.004 (0.053)
No Father in Household	0.697*** (0.241)	-0.266 (0.198)	0.691*** (0.241)	-0.249 (0.197)
Mother's Educ - Jr HS	0.074 (0.253)	-0.046 (0.219)	0.077 (0.252)	0.021 (0.214)
Mother's Educ - Sr HS	-4.472 (328.963)	0.853** (0.375)	-7.109 (0.000)	0.844** (0.372)
Per cap Income - below 5k	0.396 (0.251)	-0.242 (0.220)	0.394 (0.250)	-0.284 (0.219)
Per cap Income - above 10k	0.006 (0.261)	0.232 (0.212)	0.014 (0.260)	0.272 (0.213)
Observations	1237		1237	
Censored Obs	564		564	
Log Likelihood	-946.466		-946.900	
Sample Sel: Corr of Errors (P-Val)	-0.009 (0.986)		-0.029 (0.952)	

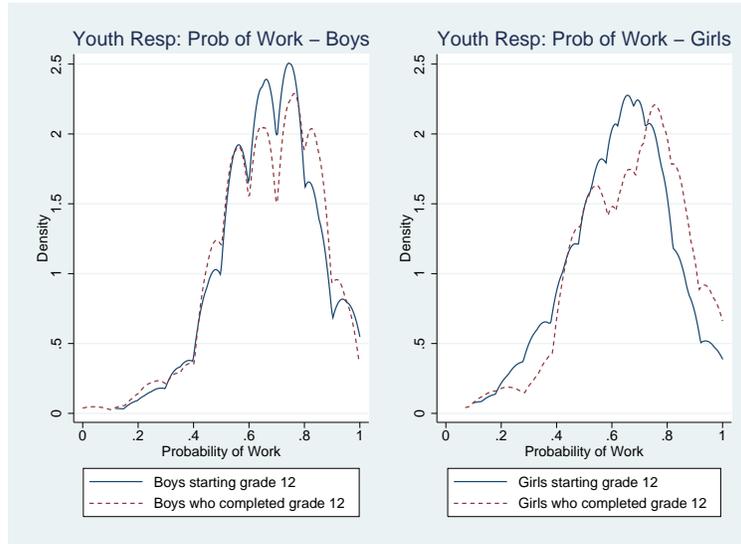
Notes: Table displays coefficients and standard errors in brackets. * p<0.1 ** p<0.05 *** p<0.01. Excl. categories: father in household, mother's education primary or less, per capita income between 5000 and 10000 pesos. A dummy for mother's education university is included but not displayed due to space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

Table 13: College Attendance Choice: Rationalization of Choices

	P-Val of KS-Test	
	Boys	Girls
Exp Log Earnings		
- Senior HS	0.495	0.714
- College	0.826	0.112
Exp Return		
- College	0.211	0.142
Prob of Work		
- Senior HS	0.906	0.003
- College	0.995	0.165
Log Var of Earnings		
- Senior HS	0.204	0.638
- College	0.369	0.142
Observations		
(Sen HS Grads/Grade 12)	269/212	404/257

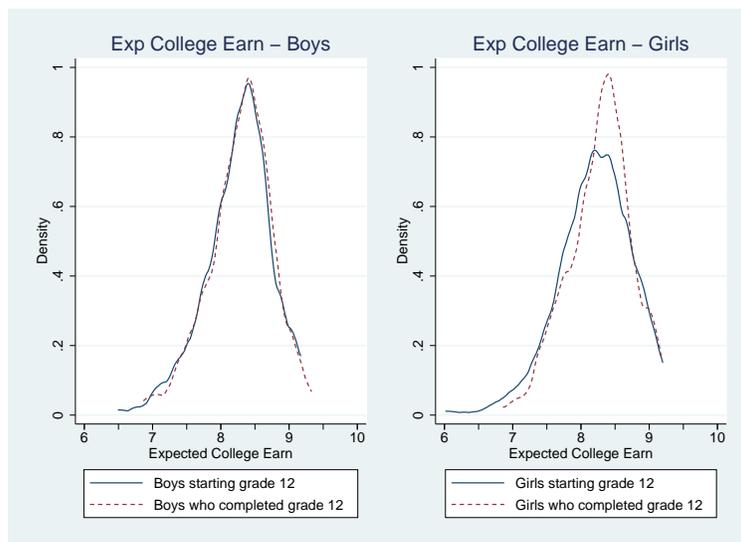
Notes: Table displays the p-values of Kolmogorov-Smirnov tests of equality of distributions. The null hypothesis is that the cross-sectional distribution of -for example- expected returns is the same for the sample of senior high school graduates (whose schooling decision we are analyzing) and the sample 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).

Figure 1: Comparing Expectations of Senior High School Graduates with a One-Year Younger Cohort: Probability of Working with a College Degree



Notes: In this figure we illustrate the cross-sectional distribution of the probability of working with a college degree, comparing our cohort of analysis, the senior high school graduates, with a cohort that is one year younger (for boys and girls respectively).

Figure 2: Comparing Expectations of Senior High School Graduates with a One-Year Younger Cohort: Expected College Earnings



Notes: In this figure we illustrate the cross-sectional distribution of expected college earnings, comparing our cohort of analysis, the senior high school graduates, with a cohort that is one year younger (for boys and girls respectively).