

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

Orazio P. Attanasio* and Katja M. Kaufmann†

April 2017

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. Moreover, we investigate whether the (relative) weights of these factors differ by gender. We use data on individuals' expectations regarding future labor market outcomes which we directly elicited from the youths, and two different measures of marriage market returns. First, marriage market returns are proxied by the (net-)supply of potential partners in the youths' local marriage markets. Second, we use data which elicits youths' beliefs about their future spouse's earnings conditional on their own education level. We find that labor market as well as marriage market returns are important determinants of the college enrollment decision. However, boys' and girls' preferences differ in terms of the relative role of the two determinants, in that the relative weight of labor market versus marriage market returns is larger for boys than 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 the editor and two anonymous referees for their valuable comments. We thank our two research assistants, Anna-Lena Wassmuth and Viviana Garcia-Gomez for their excellent work. We are grateful to Ran Abramitzky, Sandra Black, Taryn Dinkelman, Christina Gathmann, Caroline Hoxby, Rob Jensen, Victor Lavy, Lance Lochner, Shaun McRae, Aprajit Mahajan, Matthias Messner, Aloysius Siow, Jeff Smith and Michele Tertilt 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 (i.e. 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 enrollment 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 and whether the relative weights of labor and marriage market returns 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).

If marriage market considerations are important for educational decisions (i.e. individuals prefer a more educated or higher earning spouse and if the likelihood of matching with such a partner increases in their own education), this suggests that increases in the educational attainment of one gender might have important spillover effects on the other gender. This implies that if a policy or a shock (for example a labor market shock) were to increase, say, boys’ education and thus leads to a larger gender gap in education, marriage market returns might decrease this gap again by providing an additional incentive for girls to increase their educational attainment as well. Such positive spillovers should be taken into account in the cost-benefit analysis of different types of policies.

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 data on subjective expectations that directly reflect how Mexican youths perceive their labor market prospects (in terms of earnings, 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.

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).

¹For example, by enrolling in college individuals signal their ability and social status and may thereby become more attractive for 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)).

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 addition, we provide evidence on the role of returns in the marriage market in education choices. We pursue two different approaches. First we analyze the role of marriage market considerations based on marriage market proxies as explained below. Second, we present results based on direct data on individuals' beliefs about labor market as well as marriage market outcomes for different schooling scenarios.

In terms of the first approach, 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 future 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. Moreover, we conduct a number of robustness tests including one in which we add detailed controls for local labor market conditions (such as local earnings and unemployment rates by gender and education).

As a second approach, we make use of an additional dataset which contains direct information on individuals' beliefs about marriage market outcomes for different schooling scenarios in addition to labor market expectations. Individuals are asked what they believe their spouse would earn conditional on them having a high school or a college degree.

Our results indicate that both boys and girls expect high labor market returns to col-

lege and perceive lower unemployment and earnings risks for higher education levels. Also both boys and girls expect their spouse to earn more the higher their own education level. Moreover, we find that labor market as well as marriage market returns are important determinants of the college enrollment decision. However, boys' and girls' preferences differ in terms of the relative role of the two determinants, in that the relative weight of labor market versus marriage market returns is larger for boys than 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 provide supporting evidence for 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.

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.²

Papers that investigate the link between subjective expectations of earnings and education choices have analyzed school enrollment decisions (see, for example, Nguyen (2008), Jensen (2010), Attanasio and Kaufmann (2014) and Kaufmann (2014)), high school track choices (see Giustinelli (forthcoming)) and college major decisions (see, among others, 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)).

Our paper is most closely related to the first group, which investigates how perceived returns affect schooling investments. For example, Jensen (2010) finds that his sample of

²See 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) and Alvarez and Vera-Hernandez (2013)). 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).

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. Nguyen (2008) analyzes 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 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 marriage market considerations. Moreover, we also elicit and use information about the risk that individuals perceive in association with each alternative (such as earnings 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, for example, Charness and Gneezy (2012) or in terms of the willingness to compete, see, among others, 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 the survey articles by Azmat, Guell, and Manning (2006) and Azmat and Petrongolo (2014) and, for example, in terms of 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, for example, Dohmen and Falk. (2011)) or differences in probability judgements (see 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 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. 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 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.

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.³ 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.⁴

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 decisions 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.

³This 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 less strongly related to cognitive ability, such as beauty or health. This is supported for example by Hamermesh and Biddle (1994) and Lee (2015).

⁴Also see Chiappori, Iyigun, and Weiss (2009) for a structural model on the relationship between education choices and the marriage market.

In a recent paper, Wiswall and Zafar (2015) analyze the effect of an information experiment on the intended major choice by NYU college freshmen. In addition to students' major-specific expectations of labor market returns, they also elicit students' expectations regarding the earnings of their (future) spouse and how these expectations depend on the major. Wiswall and Zafar (2015) find that labor market outcomes are important for the choice of the major, while spousal earnings connected to specific majors are not. In this paper we focus on the college enrollment decision and hypothesize that marriage market considerations might be more relevant in this latter decision problem because people might meet potential partners at university even if they pursue different majors. Moreover, we are interested in gender differences in the determinants of college enrollment.

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 the construction of the marriage market proxies. Section 4 presents summary statistics, in particular in terms of individuals' expectations on labor market and marriage market outcomes and provides evidence on the validity of these data. Section 5 discusses the main results on the role of labor and marriage market returns for male and female youths (including a number of robustness tests related to our marriage market proxies and the data on subjective expectations). Section 6 concludes.

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 to 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 enrollment

To analyze the relationship between enrollment 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 enrollment choice. Instead of following this route, we present probit regressions that relate the probability of enrollment 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 labor market conditions through state dummies and –in further robustness checks– we also control for labor market conditions at the local level, such as average earnings and unemployment rates by gender and education.

Specifically, to model the decision to enroll in college, having completed senior high school, we use a latent index model which we estimate on the sample of senior high school graduates. Denoting with S the college enrollment 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,

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

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 $E(\log Y_z)$ 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.

For robustness, we also show specifications including the expected (gross) returns to college instead of expected log college and high school earnings separately, since the expected gross return is generally analyzed in the literature as the most important determinant of the college enrollment decision. We denote the expected (gross) returns to college as ρ , which is defined as $\rho = E(\log Y_2) - E(\log Y_1)$.

$Var(\log Y_z)$ 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. As measures for R^M we either use a proxy for marriage market returns or a direct measure of youths' beliefs about their future spouses' earnings conditional on their own educational degree.

We obviously expect higher expected log *college* earnings to have a positive effect on the decision to attend college, while higher expected log *high school* earnings (i.e. higher opportunity costs) should have a negative effect. Focusing on the return to college, we would 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 are difficult to capture 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 enrollment 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 pursue two different approaches. First, 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 that case an increased likelihood of finding a (suitable) spouse or finding a "higher-quality" (e.g. more educated) spouse would constitute the return

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.

of leaving one's locality to continue schooling.

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 (or "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. In fact, using the administrative dataset "Estadística de matrimonios (marriage statistics)" from the Mexican Statistical Institute INEGI, which contains information on all civil marriages that took place in Mexico in 2005,⁶ one can show that 69% of marriages took place between spouses from the same locality. As expected, this share decreases with individuals education levels (74% of individuals with primary education or less have a spouse from the same locality, 68% of individuals with secondary education and 59% of college-educated individuals). These figures lend support to our notion of a "local marriage market" and is also consistent with our hypothesis that individuals might obtain a higher schooling degree to meet individuals from different localities and get married.⁷

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 college enrollment decision is her beliefs about her future labor market outcomes conditional on having a college degree or not, and we have direct measures of these beliefs. For this reason we can analyze jointly the role of labor market and marriage market considerations in college attendance decisions. Since one might still be worried that the sex ratio could pick up current labor market conditions, which might be relevant for the college decision on top of the beliefs of future own labor market outcomes, we carefully control

⁶According to Census data from 2000 around 92% of marriages are civil marriages, while 8% are only religious marriages. Thus our data contains information on the large majority of marriages.

⁷Moreover, using the Mexican longitudinal data "Mexican Family Life Survey (MxFLS)", we show that few individuals who migrate to obtain a higher schooling degree return to their locality to live and get married there (less than 16%).

for labor market conditions at the local level.⁸ We make use of different measures of local labor market statistics based on Census data (as discussed in more detail in the following section), and corroborate the validity of these measures using the Mexican National survey of occupation and employment ENOE.⁹ In particular, we add measures of average earnings, of unemployment rates and of partial employment and underemployment rates by gender and schooling degree at the local level. We show that our results are robust to adding these controls. Moreover, we find that individuals' beliefs about their own future labor market outcomes are important determinants of their education choice, while general labor market conditions are substantially less important once individuals' labor market expectations are controlled for.

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.¹⁰ 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 education level of the spouse.

Our second measure of marriage market returns relies on eliciting youths' beliefs about the earnings of their future spouse conditional on their own educational degree. These data thus allow us to use a more direct measure of "marriage market returns" and results based on this measure can therefore lend further support to the validity of our findings based on marriage market proxies.

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.

⁸Heckman, Ichimura, Smith and Todd (1998) and Heckman, Ichimura and Todd (1997) have shown the importance of taking into account local labor market conditions instead of labor market conditions at a higher level of aggregation.

⁹We cannot use data from the ENOE survey directly to compute local labor market variables, since too few of the localities in our main dataset Jovenes can be matched

¹⁰When 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 enrollment decision which could arise from observing others who went to college.

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 information on around 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).¹¹ We use the data on the siblings sample for a number of

¹¹Unfortunately, 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). In section 4.1 we

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 practiced 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?*
 - (c) *From zero to one hundred, what is the probability that your earnings at that age will be at least x ?*

compare individual and family background characteristics as well as people's expectations for the full sample and the subsample for which we have Census data.

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.

These types of subjective expectations questions have 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. In section 4 we discuss 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 enrollment decisions had been made.

While the timing of the survey (which occurred just after the school enrollment decision had been taken) allows to use the expectations data to estimate an enrollment model, it raises the potential concern that individuals might try to rationalize their choices. If this were to be the case, individuals who decided to enroll in college would rationalize their choice by stating higher expected returns to college (that is higher college earnings and/or lower expected high school earnings), and those, who decided not to enroll, would state lower expected returns to college. This would lead to a more dispersed cross-section of returns after the decision compared to before the enrollment decision.¹²

We use this argument to check whether ex-post rationalization introduces biases in expectations. In particular, 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 is more spread out for the cohort of senior high school graduates who have decided already whether to enroll in college or not, compared to the distribution of the cohort that is one year younger and just starting grade 12. The same reasoning should

¹²This 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 enroll in college and now state high returns to college and vice versa.

hold for perceptions of the probability of working. In section E we discuss this issue at length and present evidence that our main results cannot be explained by ex-post rationalization.

Moreover, we address this concern by comparing the results of this first survey to results based on data from an additional survey we conducted several years later to elicit expectations of labor market and marriage market returns (as discussed in section 3.5). The additional survey elicited youths' expectations about labor and marriage market outcomes ex-ante, i.e. while the youths attended their last year of senior high school in 2010/11. We resurveyed those youths one year later in 2011/12 after their decision to continue schooling and enroll in college or to stop school. Thus results based on these supplementary data do not suffer from the problem that expectations are elicited after choices have been made.

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, $\rho = E(\ln(Y^2)) - E(\ln(Y^1))$.¹³

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

¹³Notice 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.

the extremes. In our companion paper, we show that the first moment of the individual distribution is extremely robust with respect to the underlying distributional assumption, while the second moment is obviously larger for the step-wise uniform distribution that puts more weight on extreme values. In this paper we present results based on the triangular distribution, but we perform robustness checks using the other two distributional assumptions and point out differences if they occur.

3.4 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 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). 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 D.1 in the Online Appendix).

3.5 Data on Marriage Markets Returns

To analyze the importance of marriage market considerations as a determinant of college enrollment decisions, we pursue two different approaches. First, we use a proxy based on the availability of (suitable) potential partners in the locality of residence. Second, we rely on direct data on individuals' beliefs of marriage market returns.

Marriage market proxy: 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 and provide evidence on how robust results are based on alternative age ranges. We assume that high school graduates decide based on the prevalent sex ratio of unmarried men and women of that age range.¹⁴

Due to data constraints, we do not observe the prevalent sex ratio in 2005, but only in 2000 based on Mexican Census data from the year 2000 (“XII Censo General de Poblacion y Vivienda” by the Mexican statistical institute, INEGI), for which information is available at the locality level.¹⁵ Having information at the locality level is important, because our “marriage market” proxy is most likely to capture people’s knowledge about availability of partners in the case of a small reference area (“marriage market”). As discussed in section 2.1, evidence based on match rates of individuals from the same locality and based on migration patterns, support the assumption 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. We compare summary statistics of the two samples in terms of expectations and individual and family background characteristics in section 4.2 below and provide robustness checks on whether our results are still externally valid for the Jovenes sample.

¹⁴For descriptive statistics on the marriage market in Mexico, such as the degree of assortative mating, see the Online Appendix section A and Table A.1.

¹⁵The sex ratio from the year 2000 should be a reasonably good proxy for the relevant sex ratio for our cohort of high school graduates if they either look at the prevalent sex ratio at age 18 and there is some persistence in the sex ratio or if the sex ratio they observe in their youths for those older cohorts who already date plays a role in terms of affecting their beliefs about the availability of suitable partners in their locality. Using Mexican Census data from the year 2010, we show that the sex ratio is relatively persistent in that the correlation between the sex ratio in 2000 and in 2010 (i.e. 10 years later) is around 0.3. Moreover, using a measure for 2000 instead of 2005 should introduce noise and might thereby bias our estimates downwards. Thus the true effects might even be bigger than the ones we find.

Beliefs about marriage market returns: Five years after the implementation of the survey "Jovenes con Oportunidades", we conducted a pilot study on youths from Progresia families in junior and senior high school and used the opportunity to include questions on individuals' beliefs about labor and marriage market outcomes related to different hypothetical schooling scenarios. In particular, in the academic year 2010/11, we collected baseline data on these youths including a module on their beliefs, while we resurveyed them in the following academic year 2011/12 to learn about their schooling decisions one year later.

We make use of the cohort of youths who were in their last year of senior high school in 2010/11 and who decided (or not) to enroll in college in 2011/12. Due to budgetary constraints the pilot study was conducted on a relatively small sample of youths and due to the eruption of drug violence in the academic year 2011/12, we were not able to follow up on part of the youths, so that –focusing on the cohort of high school graduates– we end up with only 76 observations.

The first survey round in 2010/11 contained a module with questions on individuals' expectations related to different outcomes for different schooling scenarios. In particular, we asked the same questions on labor market outcomes (probability of work, minimum and maximum earnings and probability of earnings above the midpoint) as discussed in section 3.2 (see survey questions 1 to 2 (c)) with the only exception that individuals are asked for outcomes at the age of 30 instead of 25.

In addition we added the following question supposed to elicit individuals' beliefs about marriage market outcomes. Again we considered the scenario 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 the following questions about their future spouse's earnings. For example, in the case of the second scenario the question is:

3 Assume that you complete College, and that this is your highest schooling degree. Assume that you are married or cohabiting at the age of 30. How much do you think will your spouse (or partner) earn per month?

We present descriptive statistics of these additional data on labor and marriage market expectations in section 4.3 and compare it to the data based on the Jovenes survey.

3.6 Data for Local Labor Market Conditions

To control more carefully for labor market opportunities which might be correlated with the sex ratio of unmarried men to women, we compute a series of measures of (un)employment and partial/underemployment rates and of earnings at the locality level (by gender and education). In particular, we rely on the same data used to calculate the proxy for marriage market concerns (as discussed above), namely the Mexican Census from 2000.

We construct local labor market controls following the definitions of the Mexican Statistical Institute (INEGI) who compute labor market statistics based on the Mexican labor market survey, the “National survey of occupation and employment” (ENOE).¹⁶ The first measure we use is the unemployment rate, which is computed by dividing the number of individuals unemployed by the size of the economically active population (see detailed definitions in the Online Appendix) and which –according to the Census– is around 1% for women and 2% for men (compared to 3% for both women and men based on ENOE).

While a large fraction of the population in Mexico (and other Latin American countries) work in the informal sector, few people who seek work do not find any work. Thus unemployment rates are low, but underemployment and partial employment are a problem in that people work fewer hours than they would like to work. For that reason a commonly used measure in this context is the “unemployment and partial work rate” (or “Tasa de ocupación parcial y desocupación”) based on people who are either unemployed or working less than 15 hours in the week of reference. The figures for the unemployment and partial work rates based on the Census basically coincide with those from the employment survey ENOE, that is around 15% of women are unemployed or partially employed, while the value is around 7% for men.

Conditional on education, the figures are as follows: 17% of women with less than secondary education are unemployed or partially employed compared to 13% of women with secondary education and 9% among women with more than secondary education. For men, these figures are 7%, 8% and 5%, respectively. In terms of monthly earnings, women with secondary education earn on average 1870 Pesos, while women with more than secondary education earn 3190 Pesos. For men the figures are 2440 Pesos and 6300 Pesos.

4 Descriptive Statistics

Before using the data on subjective expectations to model schooling choices, we describe the general patterns and provide some evidence on data 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 1 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

¹⁶We cannot use data from the ENOE survey directly to compute local labor market variables, since only few of the localities in our main dataset “Jovenes” can be matched. Instead, we use ENOE (and the relevant documentation of ENOE) to corroborate the validity of our labor market variables based on the Census data (by comparing summary statistics of the labor market variables based on the two data sets).

youths and report whether differences are significant. In the first three columns of Table 1, 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).

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 boys and girls are very similar in both the full sample and the subsample (once we have controlled 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).

Due to space constraints, we just summarize the evidence from Table 1 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 observed differences between girls and boys conform with 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 (for a detailed discussion of Table 1, see Online Appendix section B). We now turn to analyze the differences in subjective expectations.

4.2 Descriptive Statistics: Subjective Expectations

In Table 2, 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 Online Appendix section C). 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.

One might be surprised that the gender gap in expected earnings is not significant. This is due to the fact that girls and boys differ somewhat in terms of their family characteristics (as discussed in Section 4.1). Girls tend to come from slightly more privileged and educated families, which are characteristics that tend to be positively related to expected earnings. Thus controlling for those characteristics, there is a significant gender gap in terms of high school earnings (as shown in the Online Appendix in Table C.2).

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 25.¹⁷

As discussed in sections 3.5 and 3.6, to construct our proxies of marriage market returns and to control for local labor market conditions, we need to match our data from the Jovenes survey with Mexican Census data which contains information at the locality level, but not for all localities. For this reason, we lose about half of our original sample when we perform this match. While this does not bias our results for the localities which are included in our sample, the results are not necessarily representative anymore for Mexican high school graduates whose families are beneficiaries of Progresa/Oportunidades (which is the sample we start of with). To understand whether and in what dimensions this matched sample might differ from our original sample, we compare labor market expectations and individual and family background characteristics for the subsample of Jovenes that can be matched with Census Data versus the subsample that cannot be matched with Census information.

Table B.1 in the Online Appendix shows that differences in terms of youths’ labor market expectations between the Jovenes sample that can be matched with Census data and the subsample that cannot be matched are small and mostly insignificant. Also in terms of individual and family background characteristics differences are mostly small, for example

¹⁷In 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.

college enrollment rates are very similar for boys and somewhat larger in the matched sample for girls. But there are some significant differences in terms of family background, in that individuals in our matched sample come from somewhat more privileged families. Mothers are slightly more educated in the matched sample, per capita income is higher, the number of siblings is somewhat lower, and families are less likely to have the father in the household. One potential reason for these differences might be that individuals from smaller (poorer) localities tend to be overrepresented in the Jovenes data compared to the Mexican Census and thus the matched data contains more individuals from larger localities (and in this sense tends to be more representative of Mexican high school graduates in general).¹⁸ In section 5.3 we show that our results also hold for the poorer subsample of our matched dataset, i.e. they do not depend on focusing on the more privileged families in our Jovenes sample.

In Online Appendix Section C we provide evidence on the validity of the expectations data we use, by comparing summary statistics of expected labor market earnings to realized earnings and by showing how expected earnings and realized earnings compare in terms of their correlation with individual and family background characteristics.

In a nutshell, the evidence based on the validity checks suggests that our sample of Mexican youths understand the expectation questions and give meaningful answers. Even though the goal was not to show how rational or well-informed people are (or how people form expectations), we can tentatively conclude the following: We do not find clear evidence that girls are less well-informed about potential labor market outcomes. First, girls' perceived earnings uncertainty (i.e. the perceived standard deviation of log earnings based on the elicited individual distribution of earnings) is comparable to the one of boys. Second, girls' expectations about earnings are of a similar order of magnitude as observed earnings (as is the case for boys). Third, expectations vary with characteristics in a way one would expect.

4.3 Descriptive Statistics: Supplementary Data on Labor and Marriage Market Expectations

In this section we present summary statistics on individual and family background characteristics of the youths in our second main dataset and provide descriptive statistics on their subjective expectations on labor and marriage market outcomes. Table B.2 shows that youths in this supplementary data are relatively similar to the Jovenes sample, for example in terms of college enrollment rates, mothers' education, GPA of junior high school, number of siblings etc, but this sample is somewhat more urban (for a more detailed discussion, see Online Appendix Section B).

¹⁸Since we only have information on the size of localities in the Census data, we do not know the size for those localities that could not be matched, but the difference in characteristics points towards the above mentioned conclusion in terms of size.

Table 3 provides summary statistics on youths' labor and marriage market expectations. Both male and female youths expect earnings conditional on a college degree to be substantially higher than earnings conditional on a high school degree (for boys log earnings are 8.6 for college versus 8.1 for high school, for girls log earnings are 8.5 to 7.8). There is a gender earnings gap, in particular for high school earnings (significant on 5%).

Compared to the expectations of the youths in the Jovenes sample, earnings expectations are higher for this sample for both schooling degrees and both genders. This is most likely due to the fact that this second survey elicited beliefs about future earnings at age 30 instead of 25 and to the fact that the survey was conducted five years later and earnings of high-skilled individuals have grown further. Also the perceived probability of work with a high school degree is higher for this sample (0.77 versus 0.67).

In terms of spouses' earnings, youths expect their spouses to have close to 30% higher earnings if they themselves have a college degree instead of a high school degree. Girls expect higher earnings for their husbands than for themselves conditional on their own schooling degree, which is again consistent with an expected gender earnings gap.

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. Moreover, we are interested in whether the determinants of education choices differ between boys and girls. Therefore we estimate the schooling decisions jointly for boys and girls and interact all regressors with a gender dummy, but present the coefficients for boys and girls in two separate columns in the result tables.¹⁹

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 enrollment 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.

To test whether imposing a specific functional form by including expected returns (defined as the log difference between college and high school earnings) instead of expected earnings separately by schooling level, we show results based on expected log earnings in columns 1

¹⁹As discussed in section 4.2, we address the potential sample selection problem (which is only a problem in the case of the first main dataset used, i.e. the Jovenes sample) by using a Heckman selection correction in a non-linear context. In particular, we estimate jointly a selection equation and the main college attendance equation.

and 2 (for boys and girls, respectively) and based on including expected returns to college in columns 3 and 4 (for boys and girls, respectively).

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 important since it is a measure of the psychological costs or benefits of getting further education and it 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 strongly significant predictors for their decisions to enroll in college (significant on 1%). Similarly, boys’ expectations about log earnings conditional on schooling are highly significant (on 1 %) and have the expected sign, that is higher expected high school earnings, i.e. higher opportunity costs, decrease the likelihood to go to college, while higher expected college earnings increase the likelihood to enroll in college.

The coefficient on girls’ expected returns is only about a third of the size of boys’ coefficient, it is significant at 10% and significantly smaller than the one for boys. In terms of the effect of expected log earnings, the signs of the coefficients are again as expected, i.e. higher expected high school earnings decrease the likelihood of college, higher expected college earnings increase this likelihood, though the coefficients are not significant at conventional levels. Again the coefficient of girls on both expected high school and college earnings are smaller than for boys (significant on 5% for high school earnings and p-value of 0.16 for college earnings). Coefficients on our measures of the perceived probability of work or earnings risk are not significant.

One way to interpret our estimates on expected log earnings is in the context of an expected utility framework with log utilities in that individuals compare their expected utility with a college degree and with a high school degree as the highest degree and decide about enrolling in college based on this comparison.

As we will discuss in detail in section E, 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).

5.2 Marriage Market Returns: Proxies

In the last section, we found differences in the importance 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.

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).²⁰

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).

The results of the upper panel of Table 5 suggest that marriage market considerations are indeed important for girls' college attendance decision, while this determinant appears to play a smaller (or no) role for boys. Of course, we might be concerned that our proxies for marriage market returns could, potentially, hide different effects. 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.

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. In terms of labor market outcomes, the coefficients on expected log college earnings (expected log high school earnings) are again highly significantly positive (negative) for boys and the magnitude of the coefficients is basically unchanged. Also the coefficients on girls' expected log earnings of college and high school are 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 insignificant and significantly different from

²⁰While 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 they are available upon request. Moreover, in Table D.2 in the Online Appendix, we show the same analysis as in Table 5, but including expected returns to college instead of expected log earnings for high school and college (similarly for Tables 6 and D.3, as discussed in the following section).

the ones for girls.

To conclude, we find that labor market as well as marriage market returns are important determinants of the college enrollment decision. However, boys' and girls' preferences differ in terms of the relative role of the two determinants, in that the relative weight of labor market versus marriage market returns is larger for boys than for girls.

5.3 Robustness of the Results on Marriage Market Proxies

Peer effects: A first 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.

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 by peer effects in education.

Local labor market conditions: The second key concern is that our proxy for marriage market returns might be correlated with current local labor market characteristics which might affect the college enrollment decision, even after controlling for individuals' expectations about their idiosyncratic future labor market outcomes. For this reason, we add detailed controls for labor market conditions at the local level. In particular, we use Mexican Census data to construct controls for average earnings, unemployment and partial employment rates at the local level by gender and by educational degree (as discussed in

section 3.6).

The lower panel of Table 6 shows that –after controlling for individuals’ beliefs about their own potential labor market outcomes– coefficients on local labor market conditions are mostly insignificant, with the exception of local partial/unemployment rates for the college-educated which has a significant negative effect on boys’ decision to enroll in college. More importantly, our conclusions in terms of the role of expected earnings and marriage market considerations is unchanged. Both labor market expectations and marriage market considerations are important determinants in college enrollment decisions.

Moreover, our results suggest that there are some important gender differences with respect to their relative roles. 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 enrollment decision, consistent with anecdotal evidence (compare the Mexican acronym “MMC”s (*‘mientras me caso’*) for girls going to college to find a husband).

We conducted further robustness checks (i) related to the fact that we can only match part of our sample to Census data to construct sex ratios, (ii) to test for robustness with respect to the age range used to construct sex ratios as proxies for the (net-) supply of marriageable partners, (iii) to test for robustness with respect to different measures of unemployment rates, (iv) to test for the importance of the sex ratio in larger localities and (v) to test for interaction effects between expected returns and the probability of working. As we show in the Online Appendix Section E, our results are robust to using different specifications and different samples, while interaction effects do not seem to be important. In the next section we show results based on elicited expectations on labor market as well as marriage market outcomes.

5.4 Marriage Market Returns: Subjective Expectations

After providing evidence on the importance of marriage market considerations in college enrollment decisions based on two marriage market proxies, we now move to an analysis based on direct data on individuals’ beliefs about marriage market and labor market outcomes.

In this section we rely on data from an additional survey which we conducted in 2010/11 and 2011/12. We focus on the cohort of youths who were in their last year of senior high school in 2010/11 (which is when we elicited their subjective expectations), and who were then enrolled college (or not) in the following year 2011/2012 (which is when we resurveyed them to gather information on their educational decisions). As in the previous sections we are interested in the role of expected log earnings conditional on high school and college and the role of the probability of working with high school or college degree.

More importantly, we now include measures of individuals' expectations of future marriage market outcomes conditional on highest schooling degree. In particular, individuals' were asked about what they think their spouses would earn when they are 30 conditional on them having a high school (or college) degree (see section 3.5 for details on the survey questions and section 4.3 on descriptive statistics). Thus we include two variables, expected log earnings of the spouse conditional on the individual him/herself having a high school degree and conditional on having a college degree.

These data also have the following important advantage over using marriage market proxies: In the case of the proxies, we have information on marriage market opportunities in the locality where the high school graduates live, but no direct evidence on the marriage opportunities when the individuals leave for college. We assume that the likelihood to meet highly educated potential partners go up when the individual enrolls in college, but we do not have direct data on that. When we use individuals' beliefs about marriage market outcomes conditional on highest schooling degree, we instead have direct information on both "states of the world", i.e. marriage outcomes with a high school degree and with a college degree. Thus results based on individuals' beliefs about marriage market outcomes can help in supporting our interpretation that it is "marriage market returns" that matter.

Table 7 displays the result of regressing the decision to enroll in college on individual and family background characteristics (gender, high school GPA, mothers' education, presence of the father in the household, number of siblings and whether the household has access to a health insurance, which we use as a proxy for how privileged the household is since we do not have direct information on household income). The coefficients on the controls have the expected sign, but only some of them are significant due to the small number of observations.

The main results of Table 7 are the following: In column 1 we pool male and female youths and control for gender in the regression. We find that higher expected log high school earnings significantly decrease the likelihood to enroll in college, as one would expect (the coefficient is significant on 1%). The coefficient on expected college earnings is not significant. In terms of the role of marriage market returns, we find that higher expected earnings of the spouse conditional on the individual him- or herself having a college degree significantly increases the likelihood to enroll in college (significant on 5%).

In order to analyze whether the role of the determinants differs by gender, we interacted our expectations variables with a gender dummy (instead of fully interacting all included variables because our sample size is small). Since coefficients on labor market expectations were similar for male and female youths, we show results based on interacting the expected marriage market outcomes with gender (see columns 2 and 3) and find that the coefficient on expected log earnings of the spouse conditional on college is only significant for female youths (on 5%) and is about 50% larger in size for girls than for boys.

To conclude, our results suggest an important role of expected labor market and marriage market outcomes as determinants of college enrollment decisions lending further support to our findings discussed in the previous section based on marriage market proxies. Moreover, using both approaches we find that boys' and girls' preferences differ in the relative weight they give to labor market versus marriage market outcomes.

5.5 Robustness: Subjective Expectations and Endogeneity (Ex-Post Rationalization)

In this section we discuss concerns that are sometimes raised with respect to data on subjective expectations, that is potential problems of endogeneity due to reversed causality (ex-post rationalization). We first conduct a test of whether ex-post rationalization appears to be present in our main data based on the Jovenes survey. Second, we discuss how our data based on the additional survey we conducted can help us address this concern and provide supporting evidence on the robustness of our results.

The term "ex-post rationalization" describes the behavior of people who state beliefs to justify their choices (ex-post), that is the decision affects the beliefs instead of beliefs affecting the decision. For example, an individual might eat a whole cake and justify this decision by stating the belief that otherwise the cake would have gone bad. In the context of this paper, one might be worried that people decide to go to college for reasons other than expected labor or marriage market returns, and that they justify this decision by stating high expected returns.

Test of ex-post rationalization: To test the possibility that our respondents answer the expected return questions to justify their choices ex-post, since the survey on expectations was conducted two-three months after the schooling decision, we conduct the test that we discussed in section 3.2 comparing the cross-section of expected earnings for our cohort of interest (i.e. students who had just graduated from senior high school and decided to enroll in college or not before the survey was conducted) with a cohort that is one year younger (and thus just starting the last grade of high school). Due to space constraints we discuss the full set of results of our tests in the Online Appendix section F. Based on those test results we do not find evidence for ex-post rationalization.

Further evidence based on the second main dataset used: Moreover, in the previous section 5.4 we provided further support for our findings by relating expectations about labor and marriage market returns which we elicited ex-ante to youths' college enrollment decisions, for which individuals were resurveyed one year later after the decision had been made. Thus

ex-post rationalization is much less of a concern for those results, which confirm our previous findings.

Lastly, recent papers have shown causal effects of earnings expectations on educational decisions by conducting randomized information experiments, such as the one by Jensen (2010) on the beliefs of eight-graders in the Dominican Republic or the paper by Wiswall and Zafar (2015) on college major decisions (as discussed in the introduction).

To conclude, we have applied different strategies and provided different pieces of evidence in support of the interpretation that individuals' earnings expectations affect their educational decision. Our results show that not only expected labor market expectations matter in college enrollment decisions, but also marriage market returns, in particular for female youths.

6 Conclusion

The results of this paper help to shed light on 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 but also the risks involved with such choices and the returns on the marriage market.

Our main result is that labor market as well as marriage market returns are important determinants of the college enrollment decision. However, boys' and girls' preferences differ in terms of the relative role of the two determinants, in that the relative weight of labor market versus marriage market returns is larger for boys than for girls.

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, since results on gender differences in determinants of schooling might be very sensitive to these assumptions.

The role of marriage market returns as a determinant of educational decisions has not received much attention in the literature so far. To analyze the importance of marriage market considerations, we used proxies for marriage market returns as well as direct data on individuals' beliefs about marriage market outcomes conditional on highest schooling degree. Both approaches led to the same conclusion, that is marriage market returns are an important determinant of the college enrollment decision, in particular for girls.

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 Progresá,” *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.
- BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2005): “Why the apple doesn’t fall far: Understanding the intergenerational transmission of human capital,” *American Economic Review*, 95, 437–449.
- 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.
- CHIAPPORI, P.-A., M. IYIGUN, AND Y. WEISS (2009): “Investment in Schooling and the Marriage Market,” *American Economic Review*, 99(5), 1689–1713.
- 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.
- 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*.
- 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*.
- MAHAJAN, A., AND A. TAROZZI (2011): “Time Inconsistency, Expectations and Technology Adoption: The Case of Insecticide Treated Nets,” *mimeo, Stanford and Duke University*.
- MANSKI, C. F. (2004): “Measuring Expectations,” *Econometrica*, 72 (5), 1329–1376.
- MENGEL, F. (2015): “Gender differences in networking,” *Mimeo*.
- MORETTI, E., AND G. DAHL (2008): “The Demand for Sons,” *Review of Economic Studies*, 75(4).
- 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 Perspective*.
- 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.

WISWALL, M., AND B. ZAFAR (2015): “Determinants of College Major Choice: Identification using an Information Experiment,” *Review of Economic Studies*, 82 (2), 791–824.

ZAFAR, B. (2009): “College Major Choice and the Gender Gap,” *Federal Reserve Bank of New York Staff Reports*.

Appendix

Table 1: Summary Statistics of Individual and Family-Background Characteristics (Jovenes data matched with Census data)

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 or more	0.024 (0.152)	0.042 (0.201)	(0.069)	0.015 (0.086)	0.044 (0.195)	(0.051)
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 2: Summary Statistics of Subjective Expectations of Future Labor Market Outcomes (Jovenes data matched with Census data)

	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)
Exp Std Dev 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.

Table 3: Summary Statistics of Subjective Expectations of Future Labor and Marriage Market Outcomes (Supplementary Data)

	Boys	Girls	Diff
	Mean/(Std Dev)		[P-Val]
Labor Market Expectations			
Exp Log Earn - HS	8.138 (0.554)	7.848 (0.595)	0.290 [0.031]
Exp Log Earn - Coll	8.641 (0.685)	8.517 (0.494)	0.124 [0.365]
Prob of Work - HS	0.785 (0.207)	0.759 (0.822)	0.026 [0.600]
Prob of Work - Coll	0.802 (0.168)	0.799 (0.224)	0.003 [0.958]
Marriage Market Expectations			
Exp Log Wage Spouse - HS	8.187 (0.537)	8.218 (0.436)	-0.030 [0.792]
Exp Log Wage Spouse - Coll	8.479 (0.643)	8.501 (0.521)	-0.022 [0.871]
Observations	36	40	

Table 4: College Attendance Choice: Labor Market

Dependent Variable	College Attendance			
	Labor Market Expectations			
	Exp Log Earn		Exp Return	
	Boys	Girls	Boys	Girls
Labor Market Exp				
Exp Log Earn - HS	-0.885*** (0.261)	-0.252 (0.174)		
Exp Log Earn - Coll	0.689*** (0.259)	0.272 (0.179)		
Expected Return - College			0.781*** (0.246)	0.276* (0.161)
Prob of Work - Sr HS	0.148 (0.599)	0.409 (0.474)	0.060 (0.590)	0.443 (0.476)
Prob of Work - College	-0.298 (0.637)	0.234 (0.581)	-0.497 (0.617)	0.350 (0.574)
Var of Log Earn - HS	7.939 (11.891)	14.791 (10.306)	6.054 (11.757)	14.551 (10.429)
Var of Log Earn - Coll	1.588 (19.528)	-4.658 (13.999)	2.888 (18.981)	-3.613 (13.993)
GPA of Jr HS (0-100)	0.053*** (0.013)	0.029*** (0.010)	0.043*** (0.010)	0.035*** (0.009)
No Father in Household	0.510** (0.240)	-0.221 (0.175)	0.517** (0.241)	-0.219 (0.176)
Mother's Educ - Jr HS	0.210 (0.236)	0.120 (0.199)	0.187 (0.233)	0.111 (0.199)
Mother's Educ - HS/Coll	-4.810 (5.001)	1.140*** (0.348)	-4.930 (4.776)	1.160*** (0.350)
Per cap Income - below 5k	0.337 (0.219)	-0.259 (0.186)	0.336 (0.217)	-0.258 (0.188)
Per cap Income - above 10k	-0.059 (0.235)	0.436** (0.177)	-0.071 (0.232)	0.442** (0.178)
Observations	1237 (564)		1237 (564)	
Log Likelihood	-1172.237		-1173.297	
Sample Sel: Error Corr (P-Val)	-0.220 (0.672)		-0.205 (0.706)	

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 Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Without Labor Market Expectations				
Ratio of Unmarried Men to Women (Sr HS)	-0.150 (0.138)			
Ratio of Unmarried Women to Men (Sr HS)		0.176* (0.098)		
Ratio of Unmarried Men to Women			-0.755* (0.455)	
Ratio of Unmarried Women to Men				0.580** (0.280)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Observations	1237 (564)		1237 (564)	
Log Likelihood	-1170.670		-1170.384	
Sample Sel: Corr of Errors (P-Val)	-0.330 (0.504)		-0.387 (0.404)	
With Labor Market Expectations				
Ratio of Unmarried Men to Women (Sr HS)	-0.15 (0.135)			
Ratio of Unmarried Women to Men (Sr HS)		0.173* (0.100)		
Ratio of Unmarried Men to Women			-0.544 (0.451)	
Ratio of Unmarried Women to Men				0.592** (0.287)
Exp Log Earn - HS	-0.864*** (0.264)	-0.265 (0.170)	-0.808*** (0.265)	-0.275 (0.168)
Exp Log Earn - Coll	0.634** (0.256)	0.263 (0.173)	0.639** (0.257)	0.261 (0.171)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Controls for Prob. of Work and Var. of Log Earn	Yes	Yes	Yes	Yes
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1159.147		-1159.478	
Sample Sel: Corr of Errors (P-Val)	-0.417 (0.350)		-0.462 (0.292)	

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

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Peer Effects				
Ratio of Unmarried Men to Women (Sr HS)	-0.111 (0.145)			
Ratio of Unmarried Women to Men (Sr HS)		0.203* (0.106)		
Ratio of Unmarried Men to Women			-0.473 (0.474)	
Ratio of Unmarried Women to Men				0.558* (0.300)
Exp Log Earn - HS	-0.869*** (0.263)	-0.308* (0.178)	-0.828*** (0.265)	-0.310* (0.177)
Exp Log Earn - Coll	0.693*** (0.263)	0.340* (0.184)	0.701*** (0.264)	0.330* (0.182)
Frac of Coll Educ Men (Women)	1.268 (1.237)	3.272*** (1.223)	1.219 (1.229)	2.838** (1.221)
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1154.600		-1155.981	
Sample Sel: Error Corr (P-Val)	-0.207 (0.685)		-0.242 (0.624)	
Local Labor Market Conditions				
Ratio of Unmarried Men to Women (Sr HS)	-0.140 (0.155)			
Ratio of Unmarried Women to Men (Sr HS)		0.193* (0.109)		
Ratio of Unmarried Men to Women			-0.772 (0.534)	
Ratio of Unmarried Women to Men				0.651** (0.323)
Exp Log Earn - HS	-0.923*** (0.268)	-0.328* (0.181)	-0.875*** (0.269)	-0.347* (0.182)
Exp Log Earn - Coll	0.779*** (0.271)	0.363* (0.189)	0.795*** (0.272)	0.357* (0.188)
Frac of Coll Educ Men (Women)	0.549 (1.375)	3.586*** (1.304)	0.439 (1.375)	3.259** (1.304)
Average Census Earn - HS	0.203 (0.127)	-0.003 (0.037)	0.213* (0.127)	-0.002 (0.038)
Average Census Earn - Coll	-0.012 (0.033)	-0.002 (0.070)	-0.016 (0.033)	-0.023 (0.070)
Partial/Unempl Rate Men - HS	-0.012 (0.018)	0.002 (0.008)	-0.013 (0.017)	-0.002 (0.008)
Partial/Unempl Rate Men - Coll	-0.032* (0.018)	-0.013 (0.011)	-0.035* (0.018)	-0.016 (0.011)
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1146.448		-1146.890	
Sample Sel: Error Corr(P-Val)	0.001 (0.999)		-0.038 (0.944)	

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 7: College Attendance Choice: Expectations on the Labor and Marriage Market (Supplementary Data)

Dependent Variable	College Attendance Decision		
	All	Boys	Girls
Labor Market Exp			
Exp Log Earn - HS	-1.304*** (0.457)		-1.335*** (0.498)
Exp Log Earn - Coll	-0.384 (0.544)		-0.387 (0.581)
Prob of Work - HS	0.127 (0.908)		0.081 (0.885)
Prob of Work - Coll	0.548 (1.193)		0.591 (1.215)
Marriage Market Exp			
Exp Log Earn Spouse - HS	0.476 (0.551)	0.339 (0.768)	0.880 (0.683)
Exp Log Earn Spouse - Coll	1.111** (0.464)	0.891 (0.612)	1.375** (0.645)
GPA - HS	0.329 (0.247)		0.263 (0.247)
Male	0.612 (0.430)		9.207* (5.261)
No Father in Household	0.281 (0.371)		0.333 (0.386)
Mother's Yrs Educ	0.015 (0.049)		0.026 (0.050)
Health Insurance	0.749** (0.365)		0.889** (0.377)
Number of Siblings	-0.070 (0.098)		-0.065 (0.099)
Observations	76		76
Log Likelihood	-34.754		-34.085
Joint Sig: Exp (P-Val)	0.0319		0.0432

Notes: Table displays coefficients and standard errors in brackets. * p<0.1 ** p<0.05 *** p<0.01. Excl. categories: female, no health insurance, father in household.

ONLINE APPENDIX: NOT FOR PUBLICATION

A Background Information

A.1 Local Labor Market Controls: Definitions

- Economically active population: persons of 15 years or more that were attached to an economic activity or were looking for a job in the reference week.
- Working population: persons of 15 years or more that in the week of reference made some economic activity during at least one hour. It include the working population that had job but did not work for a temporary reason and people that work without a wage.
- Unemployed: persons of 15 years or more that looked for a job in the week of reference because they were not attached to any economic activity or job.
- Partial work: people that worked less than 15 hours in the week of reference.

A.2 Assortative Mating in Mexico

Table A.1: Assortative Mating in Mexico

	Husbands' Education			<i>Total</i>
	Prim or Less	High School	College	
Women's Education				
Primary educ or less	64.75%	33.32%	1.93%	100%
High school	15.04%	74.83%	10.13%	100%
College or more	2.68%	28.62%	68.70%	100%
<i>Total</i>	25.38%	56.29%	18.33%	

Notes: Table based on administrative data on marriage in Mexico (for details, see section 3.5).

B Descriptive Statistics

Detailed discussion of Table 1 In the youth sample, around 27% of male and female high school graduates enroll 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 1 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 three education dummies: for primary education, junior high school and senior high school or more. About 80% of mothers have only some primary education, while around 16% have attended junior high school and only 4% attended senior high school or university. Mothers are slightly more likely to have a senior high school degree (or a higher degree) in the case of girls. This is likely due to the fact that i) the focus of this paper is on a cohort of high school graduates (i.e. relatively highly educated youths) and ii) that mothers' education is particularly strongly correlated with girls' education (see, e.g., Black, Devereux, and Salvanes (2005) and see Table 4 according to which mothers' education is more strongly related to girls' decision to enroll in college than boys' decision). For this reason girls who have managed to obtain a high school degree come from relatively more educated families (in terms of mothers' education) compared to boys.

We create three per capita income categories, where the thresholds are equal to twice and four times the minimum wage.²¹ 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).

²¹Per 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)).

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 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 1 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 observed differences between girls and boys conform with 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.

Table B.1: Summary Statistics: Jovenes Sample Matched with Census Data versus Subsample that cannot be Matched

	Summary Statistics					
	Boys			Girls		
	Census	No Census	Diff	Census	No Census	Diff
	Mean/(SD)		[P-val]	Mean/(SD)		[P-val]
Exp Log Earn - HS	7.633 (0.471)	7.607 (0.520)	0.025 [0.539]	7.538 (0.516)	7.511 (0.497)	0.026 [0.468]
Exp Log Earn - Coll	8.288 (0.461)	8.249 (0.485)	0.039 [0.328]	8.273 (0.496)	8.196 (0.470)	0.076 [0.029]
Exp Return Coll	0.655 (0.359)	0.642 (0.360)	0.013 [0.661]	0.735 (0.436)	0.685 (0.374)	0.050 [0.090]
Var of Log Earn - HS	0.064 (0.041)	0.066 (0.039)	-0.002 [0.485]	0.062 (0.038)	0.064 (0.041)	-0.002 [0.449]
Var of Log Earn - Coll	0.052 (0.032)	0.056 (0.035)	-0.004 [0.147]	0.054 (0.034)	0.054 (0.037)	0.000 [0.938]
Prob of Work - HS	0.676 (0.180)	0.650 (0.166)	0.026 [0.068]	0.680 (0.189)	0.647 (0.186)	0.033 [0.014]
Prob of Work - Coll	0.817 (0.174)	0.824 (0.148)	-0.006 [0.640]	0.833 (0.156)	0.816 (0.165)	0.017 [0.130]
Enrollment Rate: College	0.268 (0.444)	0.223 (0.417)	0.045 [0.210]	0.267 (0.443)	0.206 (0.405)	0.061 [0.044]
GPA of Jr HS (0.100)	80.172 (6.422)	81.769 (6.895)	-1.598 [0.004]	83.462 (8.038)	82.545 (6.664)	0.917 [0.085]
Mother's Educ - Jr HS	0.164 (0.371)	0.115 (0.319)	0.049 [0.087]	0.151 (0.358)	0.110 (0.313)	0.041 [0.088]
Mother's Educ - HS/Coll	0.015 (0.121)	0.022 (0.148)	-0.007 [0.512]	0.045 (0.207)	0.013 (0.115)	0.031 [0.010]
Per Cap Income - 5 to 10k	0.309 (0.463)	0.204 (0.403)	0.105 [0.004]	0.287 (0.453)	0.251 (0.434)	0.036 [0.262]
Per Cap Income - above 10k	0.264 (0.442)	0.121 (0.327)	0.143 [0.000]	0.210 (0.408)	0.144 (0.352)	0.066 [0.016]
No Father in HH	0.253 (0.435)	0.121 (0.327)	0.132 [0.000]	0.220 (0.415)	0.107 (0.309)	0.113 [0.000]
Number of Siblings	2.401 (1.758)	2.799 (1.905)	-0.398 [0.009]	2.715 (1.773)	2.746 (1.716)	-0.031 [0.807]
Observations	269	314		404	374	

Descriptive statistics of the supplementary data Table B.2 shows that youths in this supplementary data have very similar college enrollment rates to the Jovenes sample (around 27%). Youths' average GPA of senior high school is 7.8 out of 10 points (for the Jovenes sample we have information on GPA of junior high school, on which youths scored 80 points out of 100). Mothers have on average 6 years of education (72% have less than junior high school, 22% have a junior high school degree and 5% have a senior high school degree or more, i.e. individuals in this sample have mothers that are slightly more educated than the youths from the Jovenes sample). At the same time, the youths of this sample are less likely to live in a household with the father present. In particular, 31% of boys and 45% of girls live in a household without father present (compared to 25% in the Jovenes sample). This is likely due to the fact that this survey is more urban than the Jovenes sample.²² In terms of number of siblings the two sample are again quite similar (youths have around 2.8 siblings).

Table B.2: Summary Statistics of Controls in Supplementary Data

	Summary Statistics		
	Boys	Girls	Diff
	Mean/(Std Dev)		[P-Val]
Enrollment Rate: College	0.278 (0.454)	0.250 (0.439)	0.028 [0.787]
GPA HS (0-10)	7.706 (0.772)	7.930 (0.732)	-0.022 [0.197]
Mother's Yrs Educ	5.861 (3.863)	6.075 (2.749)	-0.214 [0.780]
Health Insurance	0.639 (0.487)	0.525 (0.506)	0.114 [0.322]
No Father in Household	0.306 (0.467)	0.450 (0.504)	-0.144 [0.201]
Number of Siblings	3.083 (2.075)	2.675 (1.670)	0.408 [0.346]
Observations	36	40	

²²The fact that this figure is larger for girls is consistent with what people have found in the literature, see, e.g. Moretti and Dahl (2008), though the difference in our case is not significant.

C Data Validation: Expected Earnings, Realized 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 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.

‘rational’ is probably impossible. And, in a sense, it is not too important: for modeling education choices what matters is whether the elicited expectations reflect the subjective expectations that people base their decisions on.

Appendix Table C.1 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.

According to Appendix Table C.1, girls expect lower high school earnings than boys and similar college earnings, suggesting that the expected gender gap is smaller than the observed gender gap. However, as discussed in Section 4.1, girls and boys differ somewhat in terms of family characteristics. In particular, girls tend to come from slightly more privileged and educated families, which are characteristics that tend to be positively related to expected earnings. Thus controlling for those characteristics, there is a significant gender gap, in particular in terms of high school earnings (as shown in Table C.2 and as discussed below).

The second interesting exercise consists of relating the expected value of future earnings to a number of individual and family characteristics and –in addition– to compare this to how realized earnings relate to those characteristics. For this comparison, we focus on characteristics for which we have information in the Jovenes and in the Census data, which are used for the regression of expectations and earnings realizations, respectively. Such an exercise serves the purpose 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.²³

Appendix Table C.2 shows both expected earnings and realized earnings are positively correlated with the individual being male, pointing towards a gender gap in realized as well

²³It 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.

as in expected earnings (as discussed before). Moreover, expected and realized earnings are strongly positively correlated with the families' per capita income (both in the case of earnings with a high school and with a college degree). One possible explanation is 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. Mothers' education is not a significant predictor of earnings expectations, even though it is positively correlated with realized earnings.

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.²⁴

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 clear evidence that girls are less well-informed about potential labor market outcomes. First, girls' perceived earnings uncertainty (i.e. the perceived standard deviation of log earnings based on the elicited individual distribution of earnings) is comparable to the one of boys. Second, girls' expectations about earnings are of 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.

²⁴An 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.

Table C.1: 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 C.2: Correlation between Expected (Realized) Earnings and Individual and Family Background Characteristics

Dependent Variable Data Source	Realized Log Earnings (Mexican Census)		Expected Log Earnings (Jovenes Data)	
	High School	College	High School	College
	Male	0.024*** (0.005)	0.036*** (0.004)	0.087* (0.045)
Mother's Educ - Jun HS	0.149*** (0.008)	0.129*** (0.005)	-0.065 (0.060)	-0.018 (0.058)
Mother's Educ - Sen HS	0.355*** (0.013)	0.171*** (0.006)	0.077 (0.108)	0.011 (0.105)
Per Cap Income - 5 to 10k	0.804*** (0.020)	0.689*** (0.008)	0.125** (0.049)	0.081* (0.047)
Per Cap Income - more than 10k	0.340*** (0.011)	0.347*** (0.007)	0.177*** (0.055)	0.121** (0.054)
Observations	81596	171910	1237	1237
Censored Obs			564	564
R-squared (Chi-Squared)	0.088	0.079	17.741	11.369
Lambda			0.051	0.040
S.E. (Lambda)			0.145	0.141

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

D Tables complementing main analysis

Table D.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)	-0.294* (0.177)
Interview Evening	0.280* (0.143)	0.365** (0.157)	0.318** (0.157)
Interview Week 40 to 42	0.126*** (0.029)	0.175*** (0.039)	0.169*** (0.039)
Male	-0.091*** (0.028)	-0.048 (0.036)	0.303 (0.277)
Interview Evening Weekday * Male		0.340 (0.401)	0.265 (0.399)
Interview Evening * Male		-0.404 (0.384)	-0.327 (0.382)
Interview Week 40 to 42 * Male		-0.106* (0.059)	-0.078 (0.059)
GPA Jr HS * Male			-0.002 (0.003)
GPA Jr HS * Female			0.003 (0.002)
Number of Siblings * Male			-0.039*** (0.013)
Number of Siblings * Female			-0.028** (0.011)
No Father in Household * Male			0.072 (0.052)
No Father in Household * Female			0.017 (0.046)
Mother's Educ - Jr HS * Male			0.002 (0.057)
Mother's Educ - Jr HS * Female			-0.083 (0.051)
Per cap Income - below 5k * Male			0.011 (0.054)
Per cap Income - below 5k * Female			0.124** (0.049)
Sex Ratio (HS) * Male			0.057* (0.035)
Sex Ratio (HS) * Female			-0.001 (0.030)
Observations	1237	1237	1237
F-Statistic	8.270	5.414	3.280
R-Squared	0.026	0.030	0.059
Adjusted R-Squared	0.023	0.024	0.041

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. The last specification includes state dummies, which are also interacted with the gender dummy.

Table D.2: College Attendance Choice: Labor Market and Marriage Market (using expected returns)

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Without Labor Market Expectations				
Ratio of Unmarried Men to Women (Sr HS)	-0.150 (0.138)			
Ratio of Unmarried Women to Men (Sr HS)		0.176* (0.098)		
Ratio of Unmarried Men to Women			-0.755* (0.455)	
Ratio of Unmarried Women to Men				0.580** (0.280)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1170.670		-1170.384	
Sample Sel: Corr of Errors (P-Val)	-0.330 (0.504)		-0.387 (0.404)	
With Labor Market Expectations				
Ratio of Unmarried Men to Women (Sr HS)	-0.147 (0.136)			
Ratio of Unmarried Women to Men (Sr HS)		0.177* (0.100)		
Ratio of Unmarried Men to Women			-0.596 (0.453)	
Ratio of Unmarried Women to Men				0.646** (0.282)
Expected Return - College	0.748*** (0.247)	0.280* (0.158)	0.721*** (0.248)	0.280* (0.156)
Controls for Prob. of Work and Var. of Log Earn	Yes	Yes	Yes	Yes
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1172.237		-1173.297	
Sample Sel: Corr of Errors (P-Val)	-0.220 (0.672)		-0.205 (0.706)	

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 D.3: College Attendance Choice: Labor and Marriage Market - Local Labor Market Conditions (using expected returns)

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Peer Effects				
Ratio of Unmarried Men to Women (Sr HS)	-0.101 (0.147)			
Ratio of Unmarried Women to Men (Sr HS)		0.210** (0.107)		
Ratio of Unmarried Men to Women			-0.510 (0.477)	
Ratio of Unmarried Women to Men				0.609** (0.295)
Expected Return - College	0.779*** (0.248)	0.339** (0.166)	0.763*** (0.250)	0.331** (0.164)
Frac of Coll Educ Men (Women)	1.454 (1.224)	3.395*** (1.220)	1.327 (1.219)	2.901** (1.219)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Controls for Prob. of Work and Var. of Log Earn	Yes	Yes	Yes	Yes
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1155.427		-1156.388	
Sample Sel: Corr of Errors (P-Val)	-0.153 (0.776)		-0.213 (0.676)	
Local Labor Market Conditions				
Ratio of Unmarried Men to Women (Sr HS)	-0.131 (0.156)			
Ratio of Unmarried Women to Men (Sr HS)		0.199* (0.109)		
Ratio of Unmarried Men to Women			-0.807 (0.534)	
Ratio of Unmarried Women to Men				0.679** (0.317)
Expected Return - College	0.849*** (0.255)	0.354** (0.168)	0.835*** (0.255)	0.358** (0.168)
Frac of Coll Educ Men (Women)	0.775 (1.343)	3.624*** (1.302)	0.535 (1.357)	3.279** (1.299)
Average Census Earnings - HS	0.199 (0.126)	-0.001 (0.037)	0.212* (0.127)	-0.002 (0.038)
Average Census Earnings - Coll	-0.016 (0.032)	0.005 (0.069)	-0.019 (0.033)	-0.021 (0.070)
Partial/Unempl Rate - HS	-0.012 (0.017)	0.002 (0.008)	-0.014 (0.017)	-0.002 (0.008)
Partial/Unempl Rate - Coll	-0.034* (0.018)	-0.014 (0.011)	-0.036** (0.018)	-0.017 (0.011)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Controls for Prob. of Work and Var. of Log Earn	Yes	Yes	Yes	Yes
Observations (Censored)	1237 (564)		1237 (564)	
Log Likelihood	-1146.987		-1147.026	
Sample Sel: Corr of Errors (P-Val)	0.067 (0.913)		-0.013 (0.981)	

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.

E Robustness checks

Full Jovenes sample versus matched Jovenes sample: As discussed in section 3.5, when matching our data from the Jovenes survey with Mexican Census data to construct marriage market proxies and to control for local labor market conditions, we unfortunately lose a substantial fraction of our data, since the Census did not survey all of the (relatively small) localities in which individuals from the Jovenes sample live.

In section 4.2 we compared summary statistics of the subjective expectations and of the characteristics of the individuals in the matched and the unmatched sample and found that labor market expectations and college enrollment rates are similar for the two samples, but that individuals from the subsample that could be matched with the Census (and on which our analysis relies) are from somewhat more privileged families, in particular in terms of higher per capita income.

To test whether our results are more generally applicable we conduct the following two robustness tests. First, we test whether our results also hold for the poorer half of our sample (i.e. for those that are underrepresented in our matched sample compared to the Jovenes sample), i.e. we run the regression on labor market expectations and marriage market proxies for individuals from families with below median per capita income. The upper panel of Table E.1 shows that results are very similar for this subsample. Coefficients on expected log earnings are of the same magnitude as before (for girls even slightly higher) and still mostly significant (some of the coefficients are slightly less significant due to the fact that we need to rely on only around 350 uncensored observations). Coefficients on marriage market proxies are even larger for this subsample and significant for girls on 5% and 10%.

Second, we test whether our results also hold for individuals from smaller localities, since the Census did not survey all of the (relatively small) localities in which individuals from the Jovenes sample live. As can be seen from the lower panel of Table E.1, results are again very similar for individuals from localities with less than 20,000 inhabitants. Coefficients on expected log earnings are of similar magnitude and significant for boys, while coefficients on marriage market proxies remain unchanged in size and significance for girls.

These robustness tests provide further evidence that our results are valid for Mexican youths from relatively disadvantaged families who graduated from high school.

Table E.1: College Attendance Choice: Robustness Matched versus Full Jovenes Sample

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Poor Households (Income Below Median)				
Ratio of Unmarried Men to Women (Sr HS)	-0.053 (0.188)			
Ratio of Unmarried Women to Men (Sr HS)		0.340** (0.141)		
Ratio of Unmarried Men to Women			-0.808 (0.682)	
Ratio of Unmarried Women to Men				0.639* (0.386)
Exp Log Earn - HS	-0.845* (0.493)	-0.412 (0.302)	-0.907*** (0.347)	-0.519** (0.236)
Exp Log Earn - Coll	0.536 (0.375)	0.317 (0.249)	0.600* (0.354)	0.387 (0.244)
Prob of Work - HS	0.346 (0.703)	0.637 (0.711)	0.463 (0.766)	0.764 (0.676)
Prob of Work - Coll	-0.201 (0.864)	-0.604 (0.829)	-0.229 (0.911)	-0.725 (0.857)
Var of Log Earn - HS	8.237 (14.588)	4.412 (12.165)	11.182 (16.335)	11.052 (12.255)
Var of Log Earn - Coll	-1.743 (23.155)	18.902 (16.461)	-3.142 (26.065)	18.290 (17.334)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Observations (Censored)	688 (316)		688 (316)	
Log Likelihood	-642.399		-643.385	
Sample Sel: Corr of Errors (P-Val)	-0.580 (0.592)		0.063 (0.957)	
Households in Small Loc (Below 20,000)				
Ratio of Unmarried Men to Women (Sr HS)	-0.054 (0.130)			
Ratio of Unmarried Women to Men (Sr HS)		0.171* (0.087)		
Ratio of Unmarried Men to Women			-0.496 (0.488)	
Ratio of Unmarried Women to Men				0.513* (0.305)
Exp Log Earn - HS	-0.760*** (0.224)	-0.249 (0.155)	-1.029*** (0.315)	-0.280 (0.198)
Exp Log Earn - Coll	0.417* (0.242)	0.222 (0.163)	0.636** (0.317)	0.312 (0.211)
Prob of Work - HS	0.343 (0.617)	-0.122 (0.455)	0.641 (0.841)	-0.214 (0.553)
Prob of Work - Coll	-0.733 (0.665)	-0.019 (0.536)	-0.956 (0.866)	0.304 (0.664)
Var of Log Earn - HS	16.101 (10.169)	13.690* (7.620)	23.640 (14.798)	20.523* (11.842)
Var of Log Earn - Coll	1.093 (15.220)	-7.161 (13.543)	-3.078 (23.302)	-7.722 (16.538)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Observations (Censored)	866 (392)		866 (392)	
Log Likelihood	-792.531		-800.596	
Sample Sel: Corr of Errors (P-Val)	-0.529 (0.481)		-0.037 (0.946)	

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.

Interactions of expectations: Table E.2 displays results in which we include interaction terms between expected log earnings of college and high school with the probability of working conditional on a college and high school degree. We find that the interaction terms are never significant (all p-values are above 0.9). Coefficients on expected log earnings for college and high school are somewhat larger in absolute values (compared to excluding the interaction) and have the expected signs, but standard errors are now so large that they are not significant at conventional levels. Coefficients on the marriage market proxies are unchanged in terms of magnitude and significance, i.e. the two proxies are still significant for girls but not for boys. Given these findings we focus in our main analysis on a linear and additively separable specification, since interaction terms are not significant and leaving them out does not bias coefficients on the main variables of interest, namely labor market expectations and marriage market proxies (nor on other variables).

Table E.2: College Attendance Choice: Relevance of Interaction Terms

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Ratio of Unmarried Men to Women (Sr HS)	-0.140 (0.136)			
Ratio of Unmarried Women to Men (Sr HS)		0.180* (0.100)		
Ratio of Unmarried Men to Women			-0.496 (0.448)	
Ratio of Unmarried Women to Men				0.553* (0.288)
Exp Log Earn - HS	-0.920 (0.802)	-0.558 (0.548)	-0.754 (0.791)	-0.554 (0.540)
Exp Log Earn - Coll	0.780 (0.838)	0.777 (0.663)	0.696 (0.824)	0.747 (0.652)
Prob of Work - HS	-1.207 (8.386)	-2.516 (5.645)	0.201 (8.277)	-2.464 (5.567)
Prob of Work - Coll	-0.211 (8.162)	6.139 (6.591)	-0.770 (8.018)	5.726 (6.479)
Prob Work * Exp Log Earn - HS	0.188 (1.112)	0.385 (0.749)	-0.001 (1.098)	0.379 (0.739)
Prob Work * Exp Log Earn - Coll	-0.022 (1.007)	-0.720 (0.800)	0.053 (0.990)	-0.666 (0.786)
Var of Log Earn - HS	10.405 (11.764)	12.004 (10.063)	11.110 (11.699)	14.376 (9.883)
Var of Log Earn - Coll	-2.149 (18.873)	-7.171 (13.389)	-3.171 (18.537)	-6.306 (13.207)
Controls for Ind and Background Charact.	Yes	Yes	Yes	Yes
Observations	1237		1237	
Censored Obs	564		564	
Log Likelihood	-1157.716		-1158.512	
Sample Sel: Corr of Errors	-0.427		-0.471	
P-val: LR Test of Indep Eqns	0.326		0.275	

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.

Different measure of unemployment: Table E.3 presents results controlling for local unemployment rates instead of controlling for the local unemployment and partial employment rate. The latter is our preferred measure, since unemployment rates are generally very low (everybody works) and there is less variation in the unemployment rate compared to the unemployment and partial work rate. Table E.2 shows that results are unchanged when using one or the other measure of (un)employment (for the construction of the local labor market variables, see section 3.6).

Table E.3: College Attendance Choice: Labor and Marriage Market - Local Labor Market Conditions (Unemployment Rate)

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Ratio of Unmarried Men to Women (Sr HS)	-0.124 (0.143)			
Ratio of Unmarried Women to Men (Sr HS)		0.228** (0.108)		
Ratio of Unmarried Men to Women			-0.594 (0.496)	
Ratio of Unmarried Women to Men				0.604** (0.301)
Expected Return - Coll	0.771*** (0.253)	0.347** (0.171)	0.755*** (0.255)	0.347** (0.170)
Prob of Work - HS	-0.090 (0.628)	0.400 (0.488)	-0.075 (0.627)	0.426 (0.487)
Prob of Work - Coll	-0.531 (0.647)	0.326 (0.599)	-0.447 (0.646)	0.295 (0.592)
Var of Log Earn - HS	5.860 (12.697)	14.201 (10.736)	7.882 (12.781)	17.350 (10.568)
Var of Log Earn - Coll	-4.987 (19.991)	-6.908 (14.411)	-7.053 (19.910)	-6.855 (14.370)
Frac of Coll Educ Men (Women)	1.106 (1.403)	3.028** (1.305)	0.936 (1.406)	2.752** (1.302)
Average Census Earn - HS	0.228* (0.126)	0.000 (0.036)	0.238* (0.126)	0.001 (0.036)
Average Census Earn - Coll	-0.010 (0.032)	-0.010 (0.071)	-0.012 (0.032)	-0.041 (0.071)
Unempl Rate - HS	-0.097 (0.064)	0.008 (0.038)	-0.089 (0.065)	0.007 (0.037)
Unempl Rate - Coll	-0.024 (0.056)	0.094* (0.048)	-0.027 (0.058)	0.079* (0.047)
Controls for Ind and Background Charact.	Yes	Yes	Yes	Yes
Observations	1233		1233	
Censored Obs	563		563	
Log Likelihood	-1143.720		-1144.456	
Sample Sel: Corr of Errors	-0.199		-0.253	
P-val: LR Test of Indep Eqns	0.701		0.609	

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.

Marriage market proxies and robustness concerning age range: In Table E.4 we show results using different age ranges to construct our marriage market proxies. In addition to constructing the sex ratio of unmarried men to women based on individuals between 18 and 30 (i.e. individuals of the same age or older than the youths whose college decision we analyze), we also construct sex ratios based on unmarried women and men between age 16 and 25 (see section 3.5). Table E.4 shows that our results are robust to different age ranges and estimated coefficients on the sex ratio and on expected log earnings are very similar.

Table E.4: College Attendance Choice: Marriage Market Proxies (Robustness Age Range)

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Age 18-30	Age 16-25	Age 18-30	Age 16-25
Boys				
Expected Return - Coll * Male	0.748*** (0.247)	0.716*** (0.249)	0.721*** (0.248)	0.722*** (0.250)
Prob of Work - HS * Male	0.041 (0.570)	0.032 (0.554)	0.073 (0.564)	0.081 (0.575)
Prob of Work - College * Male	-0.527 (0.601)	-0.517 (0.582)	-0.425 (0.594)	-0.385 (0.605)
Var of Log Earn - HS * Male	6.695 (11.403)	6.307 (11.025)	8.827 (11.404)	9.506 (11.629)
Var of Log Earn - Coll * Male	0.023 (18.523)	0.038 (17.886)	-2.279 (18.290)	-2.212 (18.632)
Ratio of Unmarried Men to Women (HS)	-0.147 (0.136)	-0.143 (0.151)		
Ratio of Unmarried Men to Women			-0.596 (0.453)	-0.708 (0.478)
Girls				
Expected Return - Coll * Female	0.280* (0.158)	0.259* (0.156)	0.280* (0.156)	0.287* (0.158)
Prob of Work - HS * Female	0.438 (0.462)	0.420 (0.449)	0.428 (0.458)	0.449 (0.466)
Prob of Work - College * Female	0.352 (0.554)	0.339 (0.536)	0.334 (0.546)	0.289 (0.555)
Var of Log Earn - HS * Female	12.988 (10.275)	12.474 (10.001)	15.462 (10.033)	16.130 (10.132)
Var of Log Earn - Coll * Female	-5.951 (13.426)	-6.118 (12.976)	-5.803 (13.205)	-5.088 (13.424)
Ratio of Unmarried Women to Men (HS)	0.177* (0.100)	0.176* (0.098)		
Ratio of Unmarried Women to Men			0.646** (0.282)	0.734** (0.336)
Observations	1237	1237	1237	1237
Censored Obs	564	564	564	564
Log Likelihood	-1160.461	-1155.942	-1160.137	-1159.004
Sample Sel: Corr of Errors	-0.402	-0.503	-0.457	-0.401
P-val: LR Test of Indep Eqns	0.390	0.302	0.310	0.390

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.

Relevance of marriage market proxy in large localities: As an additional robustness check, we test whether the local sex ratio has a smaller impact in larger localities. Since those are more likely to contain a university, individuals would not have to move to a different city to go to college. In fact, most of our sample lives in relatively small localities (70% live in localities with less than 20,000 inhabitants, 80% in localities with less than 50,000 inhabitants), while less than 4% live in cities with more than 500,000 people. To have enough power to be able to detect any potential effect, we therefore restrict our sample to individuals living in localities of more than 50,000 inhabitants (even though these are still not very large). Table E.5 shows that coefficients on the sex ratio (conditional on education and unconditionally) are indeed not significant in larger localities.

Table E.5: College Attendance Choice: Robustness Larger Cities

Dependent Variable	College Attendance Decision			
	Marriage Proxy 1		Marriage Proxy 2	
	Boys	Girls	Boys	Girls
Expected Return - Col	1.043*	0.386	0.967*	0.348
	(0.615)	(0.304)	(0.502)	(0.225)
Prob of Work - HS	-1.157	0.341	-1.069	-0.049
	(1.282)	(0.737)	(1.020)	(0.625)
Prob of Work - Coll	0.117	1.477	0.363	1.427
	(1.458)	(1.184)	(1.165)	(0.988)
Var of Log Earn - HS	-3.750	-9.169	-8.906	-5.194
	(21.997)	(21.421)	(17.543)	(15.473)
Var of Log Earn - Coll	-45.699	-11.108	-51.663	-7.540
	(44.027)	(20.038)	(32.487)	(15.943)
Ratio of Unmarried Men to Women (Sr HS)	-1.044			
	(0.893)			
Ratio of Unmarried Women to Men (Sr HS)		0.481		
		(0.461)		
Ratio of Unmarried Men to Women			-2.244	
			(2.228)	
Ratio of Unmarried Women to Men				0.287
				(1.402)
Frac of Coll Educ Men (Women)	1.119	4.448	0.893	2.775
	(3.074)	(3.120)	(2.557)	(2.760)
Average Census Earn - HS	0.055	0.268	0.010	0.185
	(0.265)	(0.209)	(0.228)	(0.162)
Average Census Earn - Coll	-0.044	-0.968	-0.039	-0.722
	(0.046)	(0.604)	(0.034)	(0.525)
Partial/Unempl Rate - HS	0.030	-0.019	0.021	-0.020
	(0.081)	(0.030)	(0.071)	(0.028)
Partial/Unempl Rate - Coll	-0.249**	0.111	-0.207**	0.064
	(0.112)	(0.082)	(0.091)	(0.068)
Controls for Ind. and Background Charact.	Yes	Yes	Yes	Yes
Observations (Censored)	371 (172)		371 (172)	
Log Likelihood	-319.328		-322.425	
Sample Sel: Corr of Errors (P-Val)	-0.287 (0.576)		-0.353 (0.486)	

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. We focus on localities with more than 50,000 inhabitants (even though many are still relatively small), because more than 80% of individuals live in localities with below than 50,000 inhabitants (less than 4% in localities with more than 500,000 inhabitants).

F Ex-post rationalization

To test the possibility that our respondents answer the expected return questions to justify ex-post their choices, since 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 12.²⁵ 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 F.1 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 and perceived risks. We find that the distribution of expectations is the same for the 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 F.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

²⁵As 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.

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 F.2).²⁶

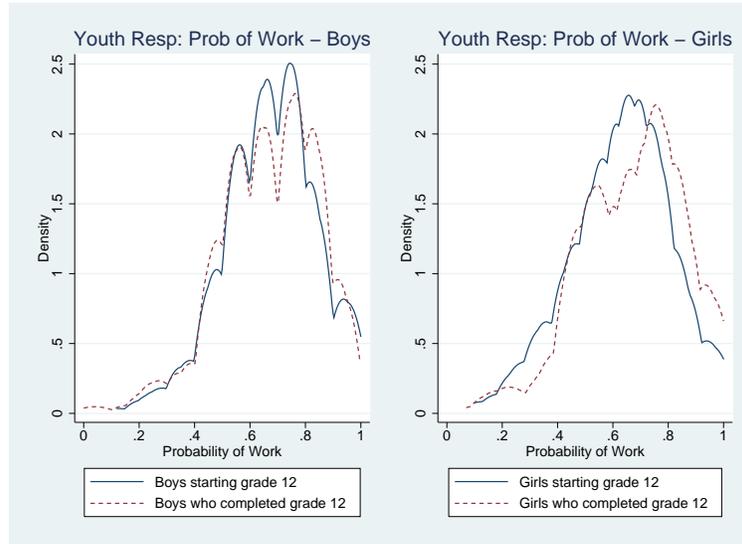
Table F.1: College Attendance Choice: Rationalization of Choices

	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).

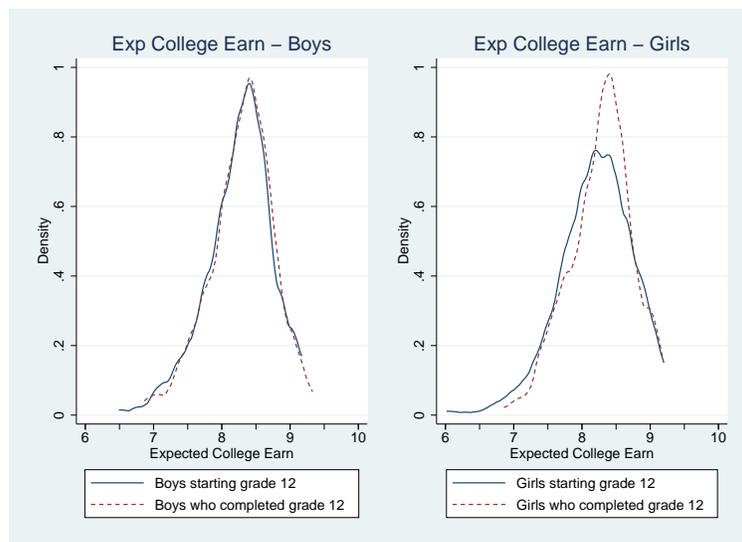
²⁶We illustrate graphically only the two instances, in which distributions are (close to) significantly different. The other graphs can be obtained from the authors upon request.

Figure F.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 F.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).