Are Parental Perceived Returns to Schooling predicting Future Schooling Decisions? Evidence from Macedonia

Alex Armand*
University of Navarra

Version: September 1st, 2014

Abstract

This paper investigates the role of parental expected returns to schooling as determinants of future schooling decisions. I show that when observing schooling decisions two years after the collection of information about perceived returns, parental subjective expectations are strong predictors for the probability of the child to be enrolled in secondary school. I provide evidence that this relation is distinctively different when looking at boys and girls. By using the unique longitudinal dimension of the dataset, I provide evidence against cognitive biases in expectation reporting and against endogeneity issues, which supports the use of subjective data in decision models.

JEL Codes: D13, J12, J16, D8, I2, J16, O15

Keywords: subjective expectations; returns to schooling; gender; cognitive biases.

1 Introduction

While taking decisions about human capital investment, it is reasonable to believe that students and/or their parents face situations of limited or imperfect information about their future income

*University of Navarra, NCID - ICS Edificio de Bibliotecas, 31009 Pamplona, Spain, and Institute for Fiscal Studies (e-mail: aarmand@unav.es).

I thank Pedro Carneiro, Valerie Lechene, Orazio Attanasio, Chris Elbers, Marcos Vera-Hernandez, Mariacristina De Nardi, Christian Dustmann, Uta Schönberg, the participants at EUDN 10th Workshop, CEP, IFS EdePo and UCL seminar for useful comments. I'm grateful to all the Staff from the CCT team of the Macedonian Ministry of Labour and Social Policy and to all the staff and interviewers from Publik, who made this project and this study possible. I thank the European Social Fund and the Regione Autonoma Valle d'Aosta for financial support. All remaining errors are my own.

possibilities and, as Manski (2004) noted, it is realistic to believe that individuals make schooling decisions based on subjective expectations rather than actual schooling returns, which have been extensively used and estimated in literature mainly using earning data. In absence of data on expectations, non-verifiable assumptions on expectations are needed, while there is little reason to believe that individuals with similar information form their expectations in the same way.

This chapter makes use of an unique dataset on subjective expectations about returns to secondary school education collected in Macedonia along with the CCT program evaluation and contributes to the growing literature linking educational choices with information about perceived returns to schooling in developing countries, where the issue of perceived returns is particularly important for developing countries, in which measured returns are high, but schooling tend to remain low (Jensen 2010, Attanasio and Kaufmann 2009). If learning about future income is happening locally by observing neighbours or friends, there is a larger chance of segregation in expectations; for instance, in rural areas, individuals might learn only about returns in agricultural-specific activities, rather than learning about returns in urban areas, where jobs related to higher levels of schooling are most probably be found. I provide evidence that ex-ante parental expectations are important in explaining schooling decisions for children.

While literature provides evidence on heterogeneity of expected returns to schooling, the use of subjective expectations in choice models has been limited in literature since data of this type has become only recently and because there is widespread belief that subjective data are flawed by cognitive biases. One type of such bias attributed to subjective expectation data is the cognitive dissonance, i.e. the tendency of respondent to report expectations that conform to their decisions rather than the real expectation (Festinger, 1962). Evidence on this type of cognitive bias is still scarce in economics literature. Mullainathan and Washington (2009) find evidence of cognitive dissonance in political support of candidates by comparing opinions on voting-age eligibles versus non-eligible after the presidential elections and providing evidence that eligibles tend to have higher polarisation than non-eligibles. In relation to subjective expectations related to schooling, Zafar (2011) provides instead evidence against cognitive biases in expectation reporting by comparing expectations on a different set of outcomes related to undergraduate major choice before and after the decision is taken. This chapter contributes to this branch of literature by providing evidence against cognitive dissonance by making use of the longitudinal dimension of the dataset and by analysing the updating process of expectations. In this chapter, cognitive dissonance would affect the updating of expectations such that expectations linked to choices made during the two data collection point would be systematically revised upward and the expectations for the educational option not taken would be systematically revised down. I provide evidence that respondents do not revise their expectations in such a way, but that the updating of expectations follows a similar pattern across individuals with different educational choices. This makes the results of the chapter robust to cognitive biases.

Section 2 presents a theoretical framework to describe how parental expectations affect investment in children's human capital. Section 3 describes the empirical strategy and Section 4 presents the data used in the chapter. Section 5 shows the main results and presents the robustness checks.

2 Enrolment model with subjective expectations

Following a Beckerian approach to schooling decision, we model secondary school enrolment as a choice based on the discounted streams of future income depending on the achieved level of schooling and on schooling cost[1]. Given the static nature of the data, we will model the decision process as a two-period model: in the first period each parent decide whether to have his child enrolled in secondary school facing a schooling cost or having his child out of school working with only primary school completed. In the second period, the child will earn an income depending on whether he enrolled in secondary school in the first period. The cost is characterised by a component that depends on individual and household characteristics, c_i, and by a random component, $\epsilon_\mathbf{i}$, which is assumed to be following a log-normal distribution $\ln \mathcal{N}(0, \sigma_\epsilon)$ such that $c_i \epsilon_i > 0$. Additionally, the model assumes that costs $c_i \epsilon_i$ scale the utility deriving from the income achieved with the completion of secondary school.

Given that the decision is made before period 1 and there is uncertainty on the future streams of income, each parent will decide depending on subjective expectations over future income conditional on educational choices. We assume that each parent has paternalistic altruism, caring about the income of the child, and that the parental utility function has a CRRA functional form in income:

$$U(y_j) = \frac{y_j^{1-\gamma}}{1-\gamma} \tag{1}$$

where y_j is the income earned by the child and $j = p$ $(j = s)$ indicates that the highest level of schooling completed is primary (secondary) school. The decision problem of parent i is then defined

[1] A detailed derivation of the results is presented in Appendix A.

by the following maximisation problem

$$\max_{\delta_i \in [0,1]} \delta_i [\beta_i \frac{U(y_s)}{c_i \epsilon_i}] + (1 - \delta_i)\,(1 + \beta_i)\,U(y_p) \tag{2}$$

where β_i is the time discount rate and δ_i is equal to 1 if the parent enrols the child in secondary school and equal to zero if he doesn't enrol the child. Assuming that income subjective expectations are distributed with probability density $f^i_{Y_s}(y_s)$ if secondary school is completed and $f^i_{Y_p}(y_p)$ if primary school is completed, we can observe that the child will be enrolled in secondary school ($\delta^*_i = 1$) if the (discounted) expected utility from the completion of secondary school is larger than the (discounted) expected utility from the completion on primary school only:

$$\beta_i \frac{1}{c_i \epsilon_i} \int U_i(y_s)\, f^i_{Y_s}(y_s)\, dy_s > (1 + \beta_i) \int U_i(y_p)\, f^i_{Y_p}(y_p)\, dy_p \tag{3}$$

Assuming that income after having completed the school level j follows a log-normal distribution $\ln \mathcal{N}(\mu_j, \sigma^2_j)$, we can write the expected utilities for each educational level as

$$\begin{aligned} E[U(y_j)] &= \int_0^{+\infty} U_i(y_j) f^i_{Y_j}(y_j)\, \mathrm{d}y_j \\ &= \frac{1}{1-\gamma} \exp\left[(1-\gamma) \left(\mu_j + \frac{\sigma^2_j\,(1-\gamma)}{2} \right) \right] \end{aligned}$$

By substituting for the expected utilities in equation (3) and taking logs, we can write the condition for enrolling the child in secondary school as

$$\Phi \equiv \ln \frac{\beta_i}{1+\beta_i} - \tilde{c} + (1-\gamma)\,\mu_s + \frac{(1-\gamma)^2}{2} \sigma^2_s - (1-\gamma)\,\mu_p - \frac{(1-\gamma)^2}{2} \sigma^2_p > \tilde{\epsilon}_i \tag{4}$$

where $\tilde{c} \equiv \ln c$ and $\tilde{\epsilon} \equiv \ln \epsilon$ and $\tilde{\epsilon} \sim N(0, \sigma_\epsilon)$. Using Φ and the symmetry of the distribution of $\tilde{\epsilon}_i$, we can therefore write the probability of the child to be enrolled in secondary school as

$$\begin{aligned} Pr\,[\delta_i = 1 \,|\, \Phi] &= Pr\,[\Phi > \tilde{\epsilon}_i \,|\, \mu, \sigma, \mathbf{X}] \\ &= 1 - F_{\tilde{\epsilon}}\,(\Phi) = F_{\tilde{\epsilon}}\,(-\Phi) \end{aligned}$$

where $F_{\tilde{\epsilon}}\,(\tilde{\epsilon})$ is the cumulative distributive function of the Gaussian distribution $N(0, \sigma_\epsilon)$. We can now analyse how the probability to be enrolled in secondary school, $Pr\,[\delta_i = 1 \,|\, \Phi]$, is affected by the first two moments characterising the subjective distribution of income conditional on completing primary or secondary school. First of all, if we look at the effect of changing the means of these distributions, it is straightforward to note that increasing the mean of expected income conditional on completing primary school reduces the probability to be enrolled. On the contrary, increasing the mean of expected income after secondary school increases the probability of being enrolled. Clearly, keeping fixed the expected income having completed primary school and all other variables, if individual i expects to receive a slightly higher income after completing secondary school education, he will have a higher incentive to invest in schooling. The derivative of the probability to be enrolled with respect to μ_s is then positive. These results are summarised by the following derivative:

$$\frac{\partial Pr\left[\delta_i = 1 \,|\, \Phi\right]}{\partial \mu_p} = -g_\epsilon\left[-\Phi\right]\,(1-\gamma) < 0 \tag{5}$$

$$\frac{\partial Pr\left[\delta_i = 1 \,|\, \Phi\right]}{\partial \mu_s} = g_\epsilon\left[-\Phi\right]\,(1-\gamma) > 0 \tag{6}$$

Given the characteristics of the data, which allow eliciting a measure of variance of the distribution of subjective expectations over future income, we can look at the effect of a change in the variances (σ_p^2 and σ_s^2) to the probability to be enrolled in secondary school. An increase in the variance of the distribution of subjective expectations over future income having completed primary school will decrease the probability to be enrolled in secondary school. On the contrary, an increase in the variance for the distribution of income having completed secondary school (σ_s^2) will increase the probability to be enrolled.

$$\frac{\partial Pr\left[\delta_i = 1 \,|\, \Phi\right]}{\partial \sigma_p^2} = -g_\epsilon\left[-\Phi\right]\,\frac{(1-\gamma)^2}{2} < 0 \tag{7}$$

$$\frac{\partial Pr\left[\delta_i = 1 \,|\, \Phi\right]}{\partial \sigma_s^2} = g_\epsilon\left[-\Phi\right]\,\frac{(1-\gamma)^2}{2} > 0 \tag{8}$$

For brevity, the model is designed assuming that the parent knows the unconditional (in terms of employment) distribution of income, however in the data, all information on expected future income is expressed conditional on being employed. However, a change in the probability of being employed after having completed secondary school can be interpreted as an increase of the income conditional on completing secondary school, all else equal. We would therefore expect to observe an increase in the probability of being enrolled when the probability of finding the job after completing secondary school becomes larger.

3 Empirical Strategy

Following Attanasio and Kaufmann (2009), this chapter presents probit regressions about the probability of having completed or being enrolled in secondary school to parental perceived returns to schooling, both in monetary and in employment terms [2]. Since in this setting we cannot relate schooling decisions to the whole probability distribution of future earnings, I assume that such distribution can be proxied by a few moments of the parental (subjective) distribution at age 25 of earnings for their children, conditional on completing the two main educational achievements for children in the targeted households (primary and secondary school). Additionally, it is important to note that schooling decisions are observed two years later (when the follow-up database has been collected, in 2012) compared to the moment in which subjective expectations have been collected (during baseline, in 2010).

To model the probability for child i living in municipality m of being enrolled in secondary school in 2012, $\delta_{im,2012}$ (where $\delta_{im,2012}$ is equal to 1 if enrolled and 0 otherwise), this chapter uses a latent index model of individual and municipality level characteristics and information about the parental perceived return to secondary school. Specifically, this chapter estimate the probability to be enrolled using the following model

$$\begin{aligned} \delta_{im,2012} &= 1 \\ &\Leftrightarrow \\ \delta^*_{im,2012} &= \alpha + \beta_0 \cdot ExpIncPrim_{i,2010} + \beta_2 \cdot ExpIncSec_{i,2010} + \\ &+ \sum_{j=1}^{2} \tau_j \cdot VarInc_{ij,2010} + X_i'\gamma + M_m'\eta + \epsilon_{im} > 0 \qquad (9) \end{aligned}$$

where $ExpIncPrim_{im,2010}$ is the expected income at age 25 conditional on completion of primary school only, $ExpIncSec_{im,2010}$ is the expected income at age 25 conditional on completion of secondary school, $VarInc_{ijm,2010}$ is the variance of future income conditional on completion of education level j ($j = 1$ indicates completion of primary school only, while $j = 2$ indicates completion of secondary school), X_i is a vector containing individual and household characters and M_m is a vector of municipality characteristics which influence schooling decisions. As discussed in the previous section, we would expect that a higher perceived return to secondary school would increase

[2] A different approach is to estimate a full dynamic optimisation model of current schooling decisions as a function of current and future benefits. See Keane and Wolpin (1997); Attanasio et al. (2012).

the probability for the child to be enrolled in secondary school, while for a given return a higher expected income conditional on completion of primary school only would lead to lower probability. To control for other characteristics of parental perception of the returns to secondary school, we include in the model information about the return in terms of employment probability. Therefore I extend equation 9 estimating the following model

$$\begin{aligned}\delta^*_{im,2012} &= \alpha + \beta_0 \cdot ExpIncPrim_{i,2010} + \beta_2 \cdot ExpIncSec_{i,2010} + \\ &+ \sum_{j=1}^{2} \tau_j \cdot VarInc_{ij,2010} + \sum_{j=1}^{2} \gamma_j \cdot PrWork_{ij,2010} + \\ &+ X_i'\gamma + M_m'\eta + \epsilon_{im} > 0 \end{aligned} \tag{10}$$

where $PrWork_{ijm,2010}$ is the perceived probability the child will find a job at age 25 conditional on completion of education level j. It is important to control for the perceived employment possibilities since the expected return is conditional on being employed.

4 Data

The data used in the chapter comes from a different number of sources. The main datasets are the Macedonian Household Surveys collected by the Ministry of Labour and Social Protection (MLSP), which contains detailed information on a variety of household information (demographics, expenditures, durable goods, housing characteristics) and individual level information on household members (education, health, labour supply). For children enrolled in secondary school, the Household Survey is supplemented with administrative data about attendance and performance at school. Additionally, I make use of different aggregated data at municipality level, supplied by Macedonian State Statistical Office, to construct measures of sex ratios, local labour market characteristics and other marriage market indicators.

For the scope of CCT program evaluation, two household surveys were collected during the Winter 2010, at the beginning of the program, and in Fall 2012, after two years of implementation. The baseline survey was conducted between November and December 2010, coinciding with the beginning of the first school year in which CCT program became available. At baseline, households were interviewed during the first two months of the program, rather than before the start of the intervention. However, it is reasonable to believe that this timeline had no effect on baseline results, since the program implementation was very slow at the beginning and the first payments were processed only in March-April 2010. In contrast, the survey was quick and the last interviews were carried out by the end of December. In parallel with the household survey, administrative data on student attendance and performance was collected by visiting secondary schools and collecting school records. This allowed double-checking the validity of self-reported information on school enrolment.

At baseline, a sample of eligible households was produced using the Ministry of Labour and Social Policy's electronic database of the recipients of all types of financial assistance, which has been assembled during Summer 2010 along with the implementation of the program. The population frame has been produced using the hardcopy archives at Social Welfare Centres (SWCs), which are the main territorial units for social welfare provision. There are 27 inter-municipal SWCs and they function as the key public providers of professional services in social work. The use of the electronic database for sampling allowed identifying 12481 SFA households with at least one child of secondary school age, from which we drew a random sample.

The follow-up survey was collected during the Fall of 2012. In order to minimise attrition, we made use of the detailed tracking information collected at baseline[3]. This methodology proved to have worked acceptably well during the follow-up data collection. In terms of SFA recipients, 1205 households were interviewed at baseline and, among those, 126 households were not found or refused to answer at follow-up, resulting in an attrition rate of 11.7%.

For the purpose of this chapter, I restrict the sample to children in all Social Financial Assistance households born from 1993 to 1998, for which data about subjective expectations are available at baseline. Table 2 presents the main descriptive statistics on child and household characteristics.

4.1 Subjective expectation module

In order to collect information about the parental perceived returns to education a specific section of the questionnaire was designed. Considering the low level of schooling among most of the respondents, it was fundamental to select a methodology that allowed eliciting a credible measure of subjective expectations without mentioning directly the term "probability" (Attanasio et al., 2005;

[3]We collected and updated contact information of at least two relatives or neighbours of the surveyed households, including addresses and telephone numbers. This allowed us minimising the risk of not finding the household in case they moved to another address or are not present at home during the attempt to interview them and to limit attrition to non-response due to refusal.

Attanasio and Kaufmann, 2009). The questionnaire asked parents information over the expected income conditional on completion of primary or secondary school (and conditional on being employed at age 25) for at least one adolescent child in the household (in the case that two adolescents of different gender are present the information was collected for both). In order to collect information on subjective expectations, the interviewer picked the youngest male and female adolescent in the age range 10-17 years old (at baseline) and refer to them in each question. The specific set of questions asked to the respondent is the following:

1. Now imagine that your child completed only primary (secondary) school and he/she finds a job. Try to imagine which possible job could he/she be employed in and imagine which could be the maximum and the minimum that he/she could earn, given

   (a) In the worst of the cases, how much do you think he/she could earn per month?

   (b) In the best of the cases, how much do you think he/she could earn per month?

2. Now using the ruler, could you indicate how likely it is that:

   (a) he/she is going to earn less than $[(2a) + (2b)]/2$ Denars?

   (b) he/she is going to earn more than $[(2a) + (2b)]/2$ Denars?

In order to elicit subjective probabilities, a 0-100 ruler was used as visual aid and was initially presented using an example linking the chances of rain with the chosen scale[4]. In order to reconstruct the probability density function, it is necessary to consider distributions that can be identified using available information: the lower (y^L) and the upper (y^U) bounds of the distribution and the reported mass probability between y^L and the midpoint $(y^L + y^U)/2$. Given the structure of the collected information and assuming a specific class of distribution functions[5], we can construct the distribution of the expected income and calculate its first moments[6] (Guiso et al., 2002). Specifically, assuming

[4]The precise text read by the interviewer is the following: We are now going to deal with events in the future that may happen or not. We have a RULER with a scale from 0 to 10 which we will use to indicate how likely do you think one event might happen. For example: If I ask you "How likely is it that tomorrow will rain?" and you are fully sure that it will rain, then you'll indicate 10. If, on the contrary, you think that it is not going to rain, you will indicate 0. In case you're not sure whether it is going to rain or not, you will give me a low value in the scale if you think that the event is not very likely, or a high value if you think it is very likely. Let's try now. "How likely is it that tomorrow will rain?"

[5]Among the distribution functions that are consistent with this setting are the step-wise uniform distribution, the triangular distribution and the bi-triangular distribution. All the data related to expectations reported in the chapter are generated assuming a triangular distribution, since we allow for the extremes to have lower density.

[6]For simplicity, in the following analysis we won't condition for education level. However, all expectations and variances are conditional on completion of either primary or secondary school.

that y^L and y^U are the reported income in the worst and the best scenario and $f_{Y|E}(y|E_i)$ is the assumed continuous density function of the expected income conditional on being employed, we can compute the expected value and the variance for the future income:

$$E[Y|E_i = 1] = \int_{y^L}^{y^U} y\, f_{Y|E}(y|E_i = 1)\,\mathrm{d}y \equiv \bar{y}_E \quad (11)$$

$$Var[Y|E_i = 1] = \int_{y^L}^{y^U} (y - \bar{y}_E)^2\, f_{Y|E}(y|E_i = 1)\,\mathrm{d}y \quad (12)$$

Table 1 reports the response rates for the section about expectations. We can note that response rates are high and above 90% for all type of questions. Response rates are slightly higher for boys and for questions that involve a single answer. When facing more complex questions, such as the ones to elicit subjective expectations of the income distribution, response rates tend to be lower. Additionally response rates are slightly higher at follow-up compared to baseline, but the reasons are not clear (learning from the respondent, selection of the respondents or higher experience from the interviewers).

Table 3 presents the descriptive statistics about parental subjective expectations in the sample of reference. In terms of income expectations, returns to secondary school range from 50.2 percent for girls in rural areas to 54.6 percent for boys in rural areas. Returns are higher for boys in both urban and rural areas, but the gap is larger for rural areas. In terms of probability of employment, the return is higher in urban areas rather than rural. It is interesting to note that for girls in urban areas attending secondary school has a larger return in terms of employment compared to boys.

In order to compare parental expected income and market returns, Figure 1 presents a comparison between the sample distribution of expected income conditional on completion of primary or secondary school with the Macedonian national average net wage for the correspondent education group. For both boys and girls the average sample expected income is lower than the national average. It is however important to note that no national data is currently available to compute average wages at age 25 for different education group, while the only available comparison is with the whole working population. It is therefore not possible to conclude whether parents over- or under-estimate market returns to schooling.

5 Results

This section presents the estimates of Equations 9 and 10 for the sample of children in all Social Financial Assistance households born from 1993 to 1998, for which data about subjective expectations are available at 2010. In all specifications, the dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Controls include gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size and number of children, household asset group and distance from the closest school[7] and indicator dummies for households living in rural areas and in the capital city Skopje. Year and semester of birth dummies and Regional dummies are included.

Table 5 shows the estimates of Equations 9 and 10 for the whole sample using a probit model. If we look at how ex-ante expectations matter for enrolment, returns to secondary school are significantly positive only for the component of expected income conditional on the completion of secondary school. Doubling the expected income conditional on completion of secondary school lead to an increase of 20.7 percent in the probability of being enrolled in secondary school. When we control for measures associated to the variance of expected income, we don't find any significant effect, while the coefficients associated with expected income are robust. Additionally, controlling for the probability of being employed at the age of 25 after completing primary or secondary school, shows that part of the effect of higher expected income conditional on completion of secondary school is captured by a higher probability to be employed when completing secondary school. The coefficients are robust when controlling for individual and municipality characteristics. This result is consistent with the recent literature providing evidence that perceived returns are important to explain how individuals take educational choices (Jensen, 2010; Attanasio and Kaufmann, 2009).

In order to control how the relationship between ex-ante parental expectations and secondary school enrolment is heterogeneous in child and household characteristics Equation 10 is estimated separately for boys and girls in urban and rural areas. Table 6 shows the estimates of the model and provides evidence of gender differences, especially when comparing rural and urban areas. Expected income conditional on completing secondary school is particularly important for girls in urban areas, while

[7] In order to construct a measure of distance from the household dwelling to the secondary school, I make use of geographic coordinates collected for each household and for each secondary school in the country. I compute road distance and time required to reach the school by car for each school in the country, in order to identify the closest secondary school.

for boys important determinants of schooling are the expected income conditional on completing primary school and the probability of being employed after secondary school. These results shows that the decision to enrol children in school is fundamentally different between boys and girls. One obvious reason is that boys and girls are affected by different local labour markets, which might make choices particularly responsive to expected returns under certain conditions. Another reason is that parental expectations might be related to other choices, especially for girls. If we look at the probability to be married for girls in the sample, we can note that expected return is a particularly strong predictor of the probability to be married within two years from reported data (see Figure 2).

Firstly, parental expectations might directly reflect the chances to go to secondary school, so that wealthier households would report higher returns to compensate for the fact that they can afford sending their children to school. Since most household adult members are unemployed, we cannot rely on income since at the moment of the interview the respondent's only official source of income is the social assistance benefit. In this case, it is very difficult to observe household's long run economic status, which is the main determinant of important choices like human capital investment.

Direct costs of attending school are often associated with the enrolment decision, especially when considering poor households. In Macedonia, as previously explained, up to secondary school, public education is free, therefore issues related to tuition and enrolment costs are not a concern in this study. In addition, recipients of Social Financial Assistance are entitled to free books. However, we need to consider transportation and living costs related to attending school, which rely directly on the accessibility of the school from the location where the household live. In order to understand how budget constraints relates to the decision to enrol in secondary school, I estimate Equation 10 for children in households from different asset groups. To allocate households in different groups I follow Filmer and Pritchett (2001) using principal-component approach and information collected on assets owned by the household to compute an asset index proxying wealth[8]. I make use of the rich information about household asset ownership collected in 2010 to build an ex-ante wealth index and divide households into three groups depending on the percentile position in distribution of the index. Table 7 presents estimates the estimates for each sub-group. Expected income conditional on completing secondary school is particularly important for children in households with low or middle level of assets, showing that households that are relatively poorer (conditional on being in a homogenous sample, e.g. all households are recipients on Financial Assistance) have higher

[8]See **??** for details on how the index was built and for robustness checks.

responsiveness to expected income compared to the households with higher wealth. If we look at the probability of employment after secondary school the effect is instead ambiguous, since it is driving enrolment in households with either low or high assets, while the coefficient is not significant for the households with middle assets.

5.1 Unconditional versus conditional expectations

So far I have considered expectations conditional on being employed. However, we noted that enrolment decisions depend as well on the probabilities of being employed after each education level. As a robustness check, I will then use jointly the information on the (point) expectation of the probability of employment conditional on completion of primary or secondary school and the expected income for the same educational level to compute unconditional income expectations (conditional only on the completed school degree). However, no information is available on the expected income in case the child is not going to be employed at age 25 and therefore we will need to build different unconditional expectations based on different assumptions about unemployment income. In this section, we will consider two levels of unemployment income equal to 1000 MKD (around 12.9 GBP) and 3000 MKD (around 38.6 GBP) per month[9]

Assuming that, in case on unemployment at age 25, each child would earn a fix amount y_{UN} provided from any sort of financial assistance (state or family) and independent from the completed level of education[10], we can then combine conditional expected income to obtain the unconditional expected income. For each level of education j, we can then write the expected income as

$$\begin{aligned} E[y_j] &= p\, E[y_j|E_i = 1] + (1-p)\, E[y_j|E_i = 0] \\ &= p\, \bar{y}_{j,E} + (1-p)\, y_{UN} \end{aligned} \tag{13}$$

where p is the probability of being employed and E_i is an indicator variable equal to 1 if the child will be employed and 0 otherwise. In addition, using the (observed) variance of income conditional

[9]Social Financial Assistance benefit is computed as a percentage of the average net salary of workers in Macedonia during the previous year. The percentage depends on the number of the family members: for 1 member families 13.59% of the basis for calculation (in 2013, 2841 MKD or around 37 GBP); for 2 members families 17.46% (in 2013, 3650 MKD or around 47 GBP); for 3 members families 23% (in 2013, 4808 MKD or around 62 GBP); for 4 members families 28.58% (in 2013, 5974 MKD or around 77 GBP); for 5 and more members families 33.34% (in 2013, 6969 MKD or around 90 GBP).

[10]This assumption might be restrictive since the expected income in case of unemployment might vary by level of education in case it is generated by activities in the informal sector rather than state assistance. However, we don't have enough information to develop further this difference.

on employment ($Var[y_j|E_i = 1] = E[y_j^2|E_i = 1] - \bar{y}_{j,E}^2$) and the assumption of fixed unemployment income ($Var[y_j|E_i = 0] = 0$), we can compute the variance of the expected income:

$$\begin{aligned} Var[y_j] &= E[y_j^2] - E[y_j]^2 \\ &= p\, E[y_j^2|E_i = 1] + (1-p)\, E[y_j^2|E_i = 0] - E[y_j]^2 \\ &= p\left[E[y_j^2|E_i = 1] + \bar{y}_{j,E}^2\right] + (1-p)\, \bar{y}_{UN}^2 - E[y_j]^2 \\ &= pVar[y_j|E_i = 1] + p\,(1-p)\,(\bar{y}_{j,E} - y_{UN})^2 \end{aligned} \tag{14}$$

Table 4 presents descriptive statistics for unconditional expected returns and incomes (and its variances) using two hypothesis for the unemployment income (1000 MKD and 3000 MKD). Similarly to Table 5, Table 8 presents the estimates of Equations 9 and 10 for the whole sample using a probit model, but using unconditional expectations based on different assumptions relative to unemployment income. If we look at how ex-ante expectations matter for enrolment, returns to secondary school are significantly positive only for the component of expected income conditional on the completion of secondary school. Doubling the expected income conditional on completion of secondary school lead to an increase of 15 to 26 percent in the probability of being enrolled in secondary school. At the same time, doubling the expected income conditional on completion of primary school only lead to a decrease 10 to 21 percent in the probability of being enrolled in secondary school. When we control for measures associated to the variance of expected income, we don't find any significant effect, while the coefficients associated with expected income are robust. Results provide evidence that using unconditional expectations rather than conditional lead to similar conclusions on the importance of expected income for enrolment. However, as previously discussed, this is based on assumptions relative to the expected income conditional on unemployment, which might not be perceived as certain.

5.2 Robustness checks

While we showed that subjective expectations are important for explaining education demand and heterogeneous program effects, we need to control whether we are measuring subjective returns associated to schooling or whether reported expectations are capturing other variables and incentives. This sub-section aims at showing that subjective expectations play an important role in explaining secondary school enrolment even after controlling for several indicators that could have generated omitted variable bias. Firstly, parental expectations might directly reflect the chances to go to secondary school, so that wealthier households would report higher returns to compensate for the fact that they can afford sending their children to school. Secondly, subjective returns could be affected by direct costs associated with distance to school and with availability of better schools. This is particularly important since direct costs of attending school are often associated with the enrolment decision, especially when considering poor households. In Macedonia, as previously explained, up to secondary school, public education is free, therefore issues related to tuition and enrolment costs are not a concern in this study. Thirdly, reported returns might be correlated with unobserved taste heterogeneity. Results show that the relation between the probability of enrolment and perceived returns to secondary school is robust to check for endogeneity. In order to understand the role of distance to school on enrolment and its relation with perceived returns, Tables 9 presents some sensitivity analysis of the coefficients on subjective expectations. We can note that controlling for distance to school has very little effect on the coefficients on perceived returns. Additionally, the coefficient on distance to school (measure in hours and standardised) is negative but not significant, showing that direct costs might not be strong determinants of secondary schooling. This result is consistent when looking at different measures of distance to school. In conclusion, Table 10 compares the coefficient by estimating the model using a linear probability model versus a probit model. Results are robust to the two estimation methods, before and after controlling for individual characteristics.

5.3 Cognitive dissonance bias

One of the main reasons why subjective expectations have not been used in choice models is that they might suffer from cognitive dissonance, i.e. respondents reports expectations that are consistent with their decisions. If the collected data suffer from cognitive dissonance we would therefore face the following situation. Imagine that $E^*[Y|E_i = 1, J]$ is the real expected income conditional on being employed after having achieved education level j, while $E[Y|E_i = 1, J]$ is the reported expectation. Data would suffer from cognitive bias if an individual who opted to enrol in education $J = j$ (in our case, secondary school) would report expectations such that the expected income consistent with the decision is higher than the real expectations. We would therefore have the following case:

$$E[Y|E_i = 1, J = j] > E^*[Y|E_i = 1, J = j] \tag{15}$$

Using subjective expectations affected by cognitive dissonance in choice models would therefore upward bias our estimates on reported subjective expectations. In order to test for cognitive dissonance, I make use of the panel dimension of the dataset and I compare the expectations reported at 2010 and the expectations for the same child reported at 2012, after a decision is taken. Zafar

(2011) provides a similar evidence against cognitive dissonance in his study on major choice and subjective expectations by comparing expectations before and after the decision in taken. I compare the expectations associated to children whose highest educational level achieved at 2010 is primary school (independently from the grade they have achieved) and it is unchanged at 2012, with children whose highest educational level achieved at 2010 is primary school and whose highest educational level achieved at 2012 is secondary school (independently from the grade they have achieved). In presence of cognitive dissonance we would expect expectations for children who transitioned from primary to secondary school to have a positive difference compared to the children who didn't transition from primary to secondary. Figure 3 presents the distribution of the change in expected return from secondary school education (defined as the difference between the expected return at 2012 and the expected return at 2010), while Figure 4 shows the change in probabilities to be employed after primary and secondary school. In both cases, I cannot reject the Kolmogorov-Smirnov test for equality of distributions (see Table **??**). This test would be invalid in the case in which parental expectations reported at baseline are already consistent with the enrolment decision of their children. This might be related to the fact that some students are already enrolled in secondary school at the time in which we collect subjective expectation. However, the decision to enrol at baseline is not permanent, since the cases of drop outs are high and the cost to enrol is relatively low.

To complement this test, I compare the reported expected return for children in primary school age and for children in secondary school age (older than 15) by looking at differences across age. Panel A of Figure 5 shows estimates of two local polynomial regressions of the return to secondary schooling for the children in primary school age (younger than 15) and for the children in secondary school age (older than 15). By comparing means at the cut-off point of 15 years old, we can observe that there is no significant difference across the two groups. Similarly, Panel B presents a local polynomial smooth for the returns to schooling in terms of employment. Both figures provides evidence that parents with children in primary school age at baseline had similar expectations compared with children in secondary school age, even when comparing children at the margin.

6 Conclusion

This chapter makes use of an unique dataset on subjective expectations about returns to secondary school education collected in Macedonia along with the CCT program evaluation and contributes to the growing literature linking educational choices with information about perceived returns to schooling in developing countries. The setting allows observing information on schooling decisions and on ex-ante parental perceived returns to secondary school (measured two years before the decision that is object of the study).

I provide evidence that ex-ante parental expectations are important in explaining secondary schooling decisions for children. Additionally, important differences exists across gender. This chapter shows that expected income conditional on completing secondary school is particularly important for girls' enrolment, while boy's enrolment is mainly driven by expected income conditional on completing primary school and by the probability of successfully finding a job after secondary school. However, since intra-household gender differences might be one of the drivers of gender inequality, future research needs to deepen the understanding of how parental expectations interact with other decisions, such as early weddings, which are clearly linked to human capital accumulation.

Additionally this chapter provides evidence on the absence of cognitive dissonance bias in self-reported income expectations. In this chapter, cognitive dissonance would affect the updating of expectations such that expectations linked to choices made during the two data collection point would be systematically revised upward and the expectations for the educational option not taken would be systematically revised down. By making use of the longitudinal dimension of the data on subjective expectation, I provide evidence that respondents do not revise their expectations following a cognitive dissonance pattern, but that the updating of expectations follow a similar pattern across individuals with different educational choices.

References

Attanasio, O. and K. Kaufmann (2009). Educational choices and subjective expectations of returns: Evidence on intrahousehold decisions and gender differences. *NBER Working Paper 15087*. 2, 6, 9, 11

Attanasio, O., C. Meghir, and M. Vera-Hernandez (2005). Elicitation, validation, and use of probability distributions of future income in developing countries. *Paper prepared for the 2005 Econometric Society Meeting*. 8

Attanasio, O. P., C. Meghir, and A. Santiago (2012). Education choices in mexico: using a structural model and a randomized experiment to evaluate progresa. *The Review of Economic Studies 79*(1), 37–66. 6

Festinger, L. (1962). *A theory of cognitive dissonance*, Volume 2. Stanford university press. 2

Filmer, D. and L. Pritchett (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of india*. *Demography 38*(1), 115–132. 12

Guiso, L., T. Jappelli, and L. Pistaferri (2002). An empirical analysis of earnings and employment risk. *Journal of Business & Economic Statistics 20*(2). 9

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515. 2, 11

Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of political Economy 105*(3), 473–522. 6

Manski, C. (2004). Measuring expectations. *Econometrica*, 1329–1376. 2

Mullainathan, S. and E. Washington (2009). Sticking with your vote: Cognitive dissonance and political attitudes. *American Economic Journal: Applied Economics*, 86–111. 2

Zafar, B. (2011). Can subjective expectations data be used in choice models? evidence on cognitive biases. *Journal of Applied Econometrics 26*(3), 520–544. 2, 15

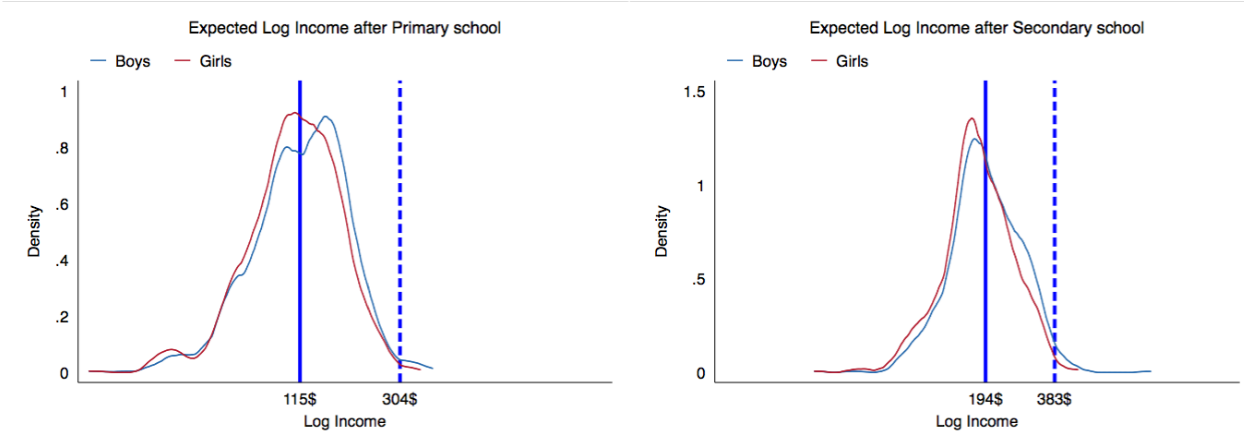
Table 1: Complete response rates for expectations related to schooling by gender of the child

	Baseline (2010)		Follow-up (2012)	
	Female	Male	Female	Male
Expectations for primary school	0.926	0.937	0.933	0.967
Expectations for secondary school	0.946	0.952	0.940	0.971
Expectations about employment	0.970	0.976	0.996	0.996
Probability to go to university	0.970	0.972	0.993	0.996

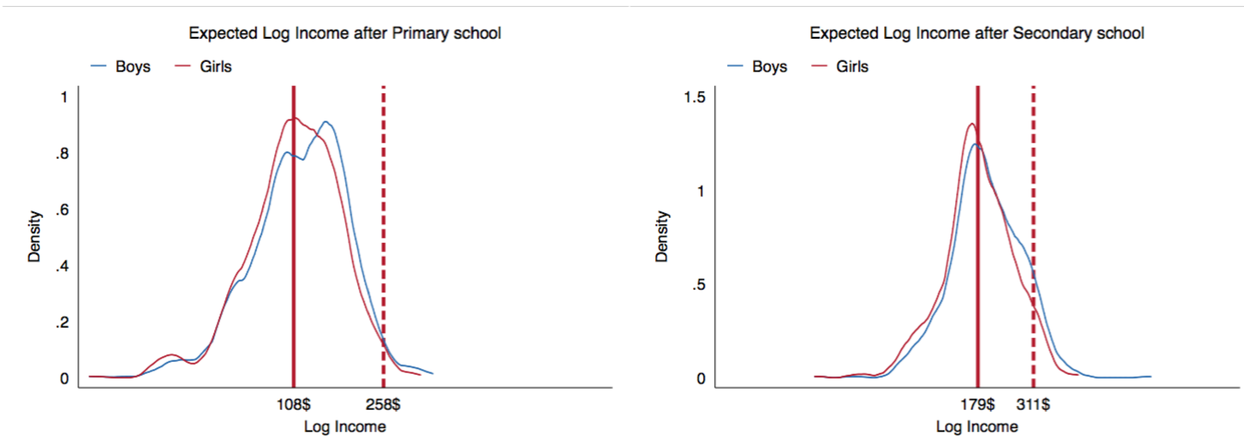
Note. An observation is considered complete if the respondent answers all requested information to compute expectations. Response rates are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Response rates are divided by gender since some households report expectations for more than one child when children in the age range for completing the expectations section have different gender.

Figure 1: Comparison between expected income and market (net) wages

Panel A. Boys

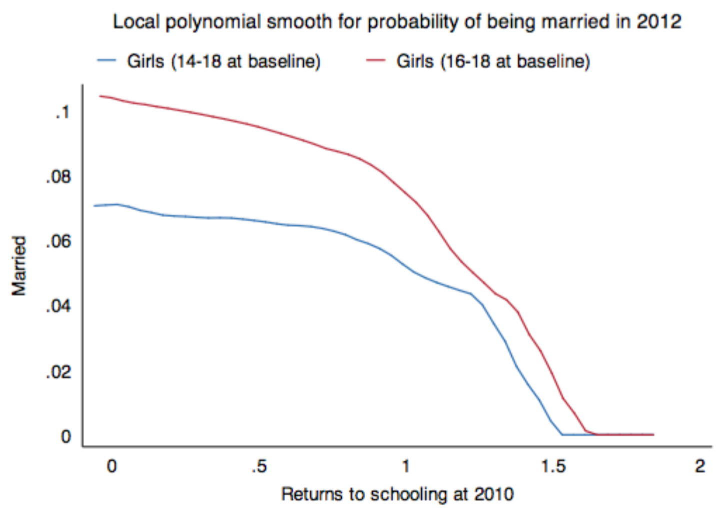


Panel B. Girls



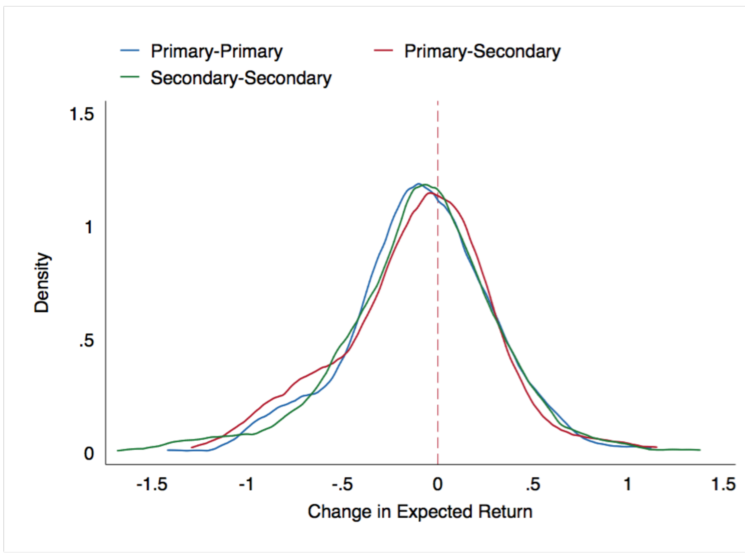
Note. The figure presents the sample distribution of expected (log)-income conditional on completing primary or secondary school, the national average net wage for the correspondent education group in 2010 (dotted line) and the correspondent sample mean in USD (solid line). It is important to note that expected income is asked for age 25, while average wages are reported for the whole population. Data about wages has been made available by the Macedonian State Statistical Office.

Figure 2: Relation between parental perceived returns to schooling and marriage status



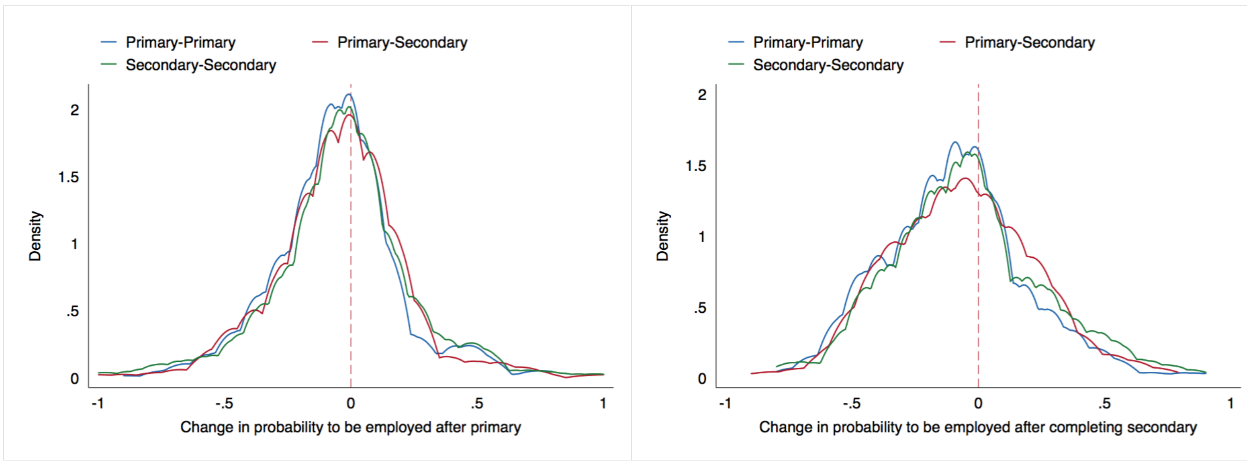
Note: The graph shows the local polynomial smooth of the probability to be married in 2012 on the perceived parental return to schooling at 2010. Dependent variable is equal to 1 if the girl is married in 2012 and 0 otherwise. In 2010 none of the girls is married.

Figure 3: Change in expected return from baseline to follow-up



Note. Change is expected return is defined as the difference between the monetary return to secondary school education collected in 2012 and the one collected in 2010 for the same child. "Primary-Secondary" refers to children that went from being in primary school in 2010 to being enrolled or having completed secondary school in 2012. "Primary-Primary" refers to children that were enrolled or had completed primary school in 2010 and their status is unchanged in 2012.

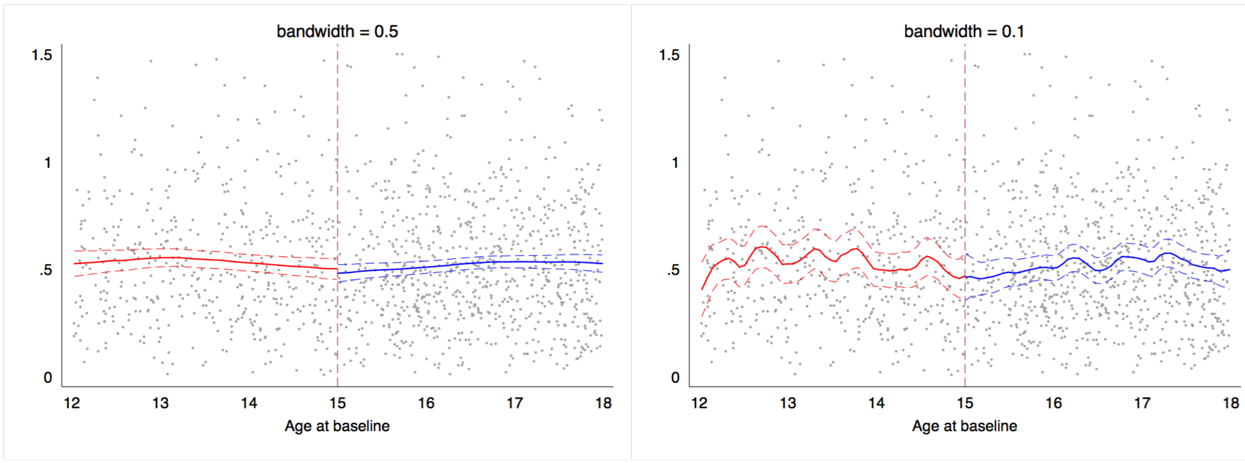
Figure 4: Change in expected probability of being employed at age 25 from baseline to follow-up



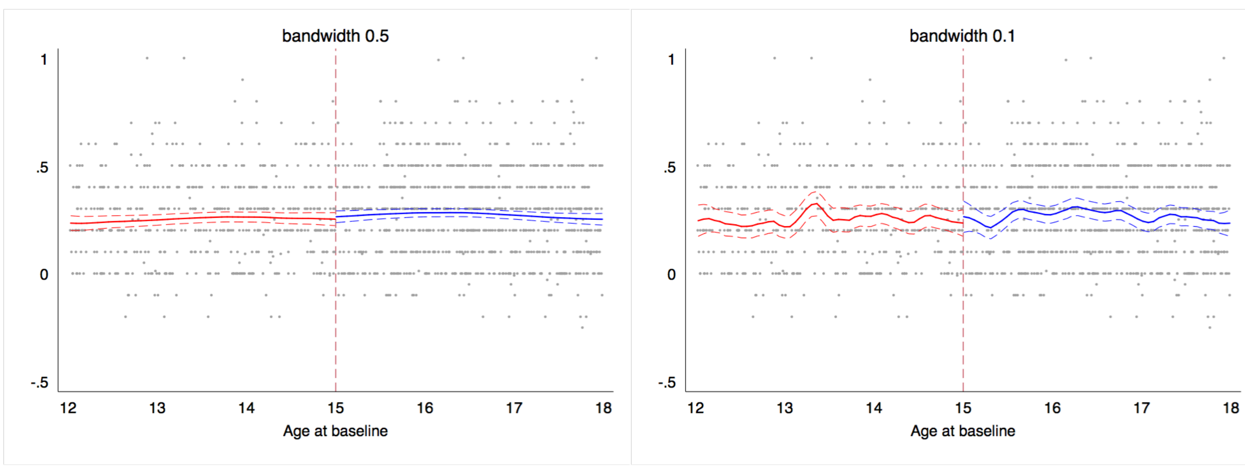
Note. The change in expected probability of being employed is defined as the difference between the probability of being employed after having completed primary school (left panel) or having completed secondary school (right panel) collected in 2012 and the one collected in 2010 for the same child. "Primary-Secondary" refers to children that went from being in primary school in 2010 to being enrolled or having completed secondary school in 2012. "Primary-Primary" refers to children that were enrolled or had completed primary school in 2010 and their status is unchanged in 2012.

Figure 5: Local polynomial regression for Expected Returns by age of the child

Panel A. Returns to schooling in monetary terms



Panel B. Returns to schooling in employment terms



Note. The Figure present local polynomial regressions (at different bandwidth) around the cut-off age of 15, which divides the age group 12-17 years old into a primary school age group and a secondary school age group. Panel A presents the return to secondary school, computed as the difference between expected incomes after primary and secondary school (reported in logarithms and computed using triangular distribution). Panel B presents the return to schooling in employment terms, defined as the difference in the probability to find a job after secondary and after primary school. 95% confidence interval is represented using dotted lines, while the local regression is represented by the solid line. Age is determined from date of birth at December 31st 2010 and is expressed in years as a continuous variable.

Table 2: Descriptive statistics on child and household characteristics

	Urban		Rural	
	Boys	Girls	Boys	Girls
Age	17.39	17.36	17.31	17.31
	(1.604)	(1.684)	(1.608)	(1.690)
Male household head	0.807	0.795	0.892	0.888
	(0.395)	(0.405)	(0.311)	(0.316)
Education of household head				
Lower primary or less	0.182	0.144	0.193	0.181
	(0.387)	(0.352)	(0.395)	(0.386)
Upper primary	0.512	0.540	0.583	0.572
	(0.501)	(0.499)	(0.494)	(0.496)
Secondary school or more	0.305	0.316	0.224	0.247
	(0.461)	(0.466)	(0.418)	(0.432)
Age (head)	46.25	46.46	46.75	46.80
	(5.330)	(5.934)	(5.473)	(5.604)
Ethnicity				
Macedonian	0.477	0.513	0.336	0.367
	(0.500)	(0.501)	(0.473)	(0.483)
Albanian	0.228	0.221	0.452	0.447
	(0.420)	(0.415)	(0.499)	(0.498)
Roma	0.214	0.190	0.0386	0.0465
	(0.411)	(0.393)	(0.193)	(0.211)
Turkish	0.0912	0.0875	0.178	0.140
	(0.288)	(0.283)	(0.383)	(0.347)
Muslim	0.526	0.490	0.726	0.684
	(0.500)	(0.501)	(0.447)	(0.466)
Household members	4.596	4.688	4.873	4.884
	(1.260)	(1.331)	(1.246)	(1.264)
Number of children	2.519	2.605	2.726	2.763
	(0.966)	(1.075)	(1.015)	(1.104)
Asset group				
Low	0.291	0.274	0.402	0.400
	(0.455)	(0.447)	(0.491)	(0.491)
Middle	0.295	0.300	0.344	0.363
	(0.457)	(0.459)	(0.476)	(0.482)
High	0.414	0.426	0.255	0.237
	(0.493)	(0.495)	(0.437)	(0.426)
Distance from closest school (hours)	0.157	0.150	0.307	0.306
	(0.254)	(0.244)	(0.222)	(0.223)

Note. Standard deviations in parenthesis. Characteristics are reported in 2012 for children born from 1993 to 1998 and for which data about subjective expectations are available in 2010. Sample includes children born from Asset groups are defined by using principal component analysis and using indicators of asset and land ownership at the time in which expectations are reported. The distance from school is determined using geo-coordinates of households and schools and by computing road distance in terms on time from the household dwelling to the closest school teaching a Macedonian program.

Table 3: Descriptive statistics of parental subjective expectations

	Urban		Rural	
	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)
Income expectations				
Return to secondary school	0.520	0.519	0.546	0.502
	(0.341)	(0.318)	(0.373)	(0.332)
Expected income (prim.)	8.566	8.513	8.550	8.494
	(0.478)	(0.427)	(0.448)	(0.436)
Expected income (sec.)	9.086	9.032	9.095	8.996
	(0.355)	(0.330)	(0.339)	(0.338)
Var. income (prim.)	0.0232	0.0242	0.0192	0.0197
	(0.0305)	(0.0332)	(0.0265)	(0.0253)
Var. income (sec.)	0.0155	0.0152	0.0133	0.0146
	(0.0206)	(0.0177)	(0.0206)	(0.0210)
Probability of employment				
Return to secondary school	0.273	0.284	0.253	0.253
	(0.215)	(0.210)	(0.198)	(0.207)
Prob. of employment (prim.)	0.218	0.185	0.242	0.218
	(0.206)	(0.189)	(0.183)	(0.181)
Prob. of employment (sec.)	0.491	0.471	0.494	0.470
	(0.229)	(0.222)	(0.196)	(0.193)
Observations	1022	1022	1022	1022

Note. Standard deviations in parenthesis. Returns to secondary school are computed assuming a triangular distribution. Return in terms of probability of employment is defined as difference between the probability of being employed conditional on completing secondary school and conditional on completing primary school.

Table 4: Descriptive statistics of parental subjective expectations using unconditional returns

	Urban		Rural	
	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)
Unemployment income: 1000 MKD				
Return to secondary school	0.692	0.717	0.689	0.647
	(0.465)	(0.488)	(0.462)	(0.435)
Expected income (prim.)	7.305	7.231	7.320	7.264
	(0.460)	(0.394)	(0.377)	(0.356)
Expected income (sec.)	7.981	7.935	7.999	7.901
	(0.589)	(0.552)	(0.483)	(0.464)
Var. income (prim.)	1.325	0.945	1.375	1.101
	(2.161)	(1.769)	(1.997)	(1.788)
Var. income (sec.)	4.521	4.485	4.863	4.478
	(2.905)	(2.867)	(2.740)	(2.762)
Unemployment income: 3000 MKD				
Return to secondary school	0.397	0.395	0.400	0.357
	(0.273)	(0.285)	(0.265)	(0.236)
Expected income (prim.)	8.159	8.119	8.163	8.138
	(0.245)	(0.202)	(0.206)	(0.185)
Expected income (sec.)	8.545	8.504	8.558	8.489
	(0.358)	(0.334)	(0.291)	(0.277)
Var. income (prim.)	1.325	0.945	1.375	1.101
	(2.161)	(1.769)	(1.997)	(1.788)
Var. income (sec.)	4.521	4.485	4.863	4.478
	(2.905)	(2.867)	(2.740)	(2.762)
Observations	1022	1022	1022	1022

Note. Standard deviations in parenthesis. Returns to secondary school are computed assuming a triangular distribution. Return in terms of probability of employment is defined as difference between the probability of being employed conditional on completing secondary school and conditional on completing primary school.

Table 5: Enrolment regression and parental perceived returns

	Dep.var.: Enrolled or completed secondary school					
	Probit	Probit	Probit	Probit	Probit	Probit
	(1)	(2)	(3)	(4)	(5)	(6)
Expected income (prim.)	-0.066	-0.070	-0.073	-0.077	-0.027	-0.043
	(0.048)	(0.047)	(0.055)	(0.053)	(0.058)	(0.057)
Expected income (sec.)	0.243***	0.202***	0.247***	0.207***	0.190**	0.163**
	(0.071)	(0.066)	(0.073)	(0.067)	(0.076)	(0.075)
Var. income (prim.)			-0.211	-0.185	-0.178	-0.225
			(0.536)	(0.520)	(0.533)	(0.532)
Var. income (sec.)			0.218	0.279	0.308	0.458
			(0.690)	(0.695)	(0.695)	(0.690)
Prob. of employment (prim.)					-0.225*	-0.174
					(0.137)	(0.110)
Prob. of employment (sec.)					0.269**	0.220**
					(0.107)	(0.108)
Regional and birthyear dummies	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓
Observations	1022	1022	1022	1022	1022	1022

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. The dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. Where indicated, I include controls for gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size, number of children, rural and Skopje dummies, household asset group and distance from the closest school. Year and semester of birth dummies and Regional dummies are included.

Table 6: Enrolment regression and parental perceived returns, by gender and type of municipality

	Dep.var.: Enrolled or completed secondary school					
		Urban			Rural	
	Girls	Boys	All	Girls	Boys	All
Expected income (prim.)	-0.049	0.085	0.035	0.037	-0.144***	-0.123*
	(0.110)	(0.095)	(0.082)	(0.170)	(0.045)	(0.066)
Expected income (sec.)	0.288**	0.218	0.174*	-0.041	0.070	0.122
	(0.133)	(0.150)	(0.097)	(0.241)	(0.068)	(0.099)
Var. income (prim.)	-1.116	4.045***	0.746	-3.727	-0.651	-1.608***
	(1.236)	(1.293)	(0.870)	(2.296)	(0.644)	(0.592)
Var. income (sec.)	-0.648	-1.073	-0.672	3.263	0.035	1.099
	(2.155)	(1.972)	(1.287)	(2.091)	(0.454)	(0.761)
Prob. of employment (prim.)	-0.281	-0.456*	-0.258	0.121	-0.074	-0.044
	(0.174)	(0.260)	(0.173)	(0.351)	(0.112)	(0.145)
Prob. of employment (sec.)	0.208	-0.073	0.035	0.358	0.365***	0.425***
	(0.187)	(0.158)	(0.135)	(0.311)	(0.095)	(0.156)
Regional and birthyear dummies	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Observations	261	268	545	215	259	474

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. The dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. Where indicated, I include controls for gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size, number of children, rural and Skopje dummies, household asset group and distance from the closest school. Year and semester of birth dummies and Regional dummies are included.

Table 7: Enrolment regression and parental perceived returns, by asset group

	Dep.var.: Enrolled or completed secondary school		
	Low	Middle	High
Expected income (prim.)	-0.127	-0.080	0.052
	(0.132)	(0.083)	(0.055)
Expected income (sec.)	0.304**	0.237**	-0.032
	(0.149)	(0.097)	(0.074)
Var. income (prim.)	-0.988	0.313	0.946
	(1.394)	(0.873)	(0.590)
Var. income (sec.)	-3.522	0.663	0.989*
	(2.333)	(1.036)	(0.555)
Prob. of employment (prim.)	-0.378	-0.124	-0.061
	(0.263)	(0.097)	(0.105)
Prob. of employment (sec.)	0.498**	0.075	0.259**
	(0.216)	(0.115)	(0.106)
Regional and birthyear dummies	✓	✓	✓
Controls	✓	✓	✓
Observations	345	328	286

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. The dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. Where indicated, I include controls for Gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size, number of children, rural and Skopje dummies, household asset group and distance from the closest school. Year and semester of birth dummies and Regional dummies are included. Asset groups are defined by using principal component analysis and using indicators of asset and land ownership at the time in which expectations are reported.

Table 8: Enrolment regression and parental perceived unconditional returns

	Dep.var.: Enrolled or completed secondary school			
Unemployment income hypothesis	1000 MKD		3000 MKD	
	Probit	Probit	Probit	Probit
	(1)	(2)	(3)	(4)
Expected income (prim.)	-0.103**	-0.123*	-0.210**	-0.208**
	(0.048)	(0.068)	(0.087)	(0.106)
Expected income (sec.)	0.138***	0.152***	0.249***	0.264***
	(0.038)	(0.042)	(0.063)	(0.075)
Var. income (prim.)		0.004		-0.001
		(0.012)		(0.010)
Var. income (sec.)		-0.003		-0.002
		(0.008)		(0.008)
Regional and birthyear dummies	✓	✓	✓	✓
Controls		✓		✓
Observations	1022	1022	1022	1022

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. The dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. The expected income is unconditional with respect to employment status at age 25 and is built using information on conditional expected income and employment expectations and different hypothesis relative to the income in case of unemployment. Where indicated, I include controls for gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size, number of children, rural and Skopje dummies, household asset group and distance from the closest school. Year and semester of birth dummies and Regional dummies are included.

Table 9: Enrolment regression, parental perceived returns and distance to school

	Dep.var.: Enrolled or completed secondary school		
	Probit (1)	Probit (2)	Probit (3)
Expected income (prim.)	-0.027 (0.058)	-0.027 (0.057)	-0.040 (0.057)
Expected income (sec.)	0.190** (0.076)	0.190** (0.076)	0.161** (0.075)
Var. income (prim.)	-0.178 (0.533)	-0.172 (0.533)	-0.179 (0.529)
Var. income (sec.)	0.308 (0.695)	0.308 (0.697)	0.440 (0.691)
Prob. of employment (prim.)	-0.225* (0.137)	-0.225 (0.137)	-0.177 (0.110)
Prob. of employment (sec.)	0.269** (0.107)	0.269** (0.107)	0.223** (0.108)
Distance to school		-0.004 (0.018)	-0.126 (0.085)
Regional and birthyear dummies	✓	✓	✓
Distance to school		✓	✓
Controls			✓
Observations	1022	1022	1022

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. The dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. Where indicated, I include controls for Gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size, number of children, rural and Skopje dummies, household asset group and distance from the closest school. Year and semester of birth dummies and Regional dummies are included. Asset groups are defined by using principal component analysis and using indicators of asset and land ownership at the time in which expectations are reported. Distance from the closest school is computed using geo-coordinates and is standardised.

Table 10: Enrolment regression and different estimation method

	Dep.var.: Enrolled or completed secondary school			
	OLS		Probit	
	(1)	(2)	(3)	(4)
Expected income (prim.)	-0.024	-0.025	-0.027	-0.040
	(0.048)	(0.043)	(0.058)	(0.057)
Expected income (sec.)	0.160**	0.114**	0.190**	0.161**
	(0.063)	(0.057)	(0.076)	(0.075)
Var. income (prim.)	-0.092	0.011	-0.178	-0.179
	(0.463)	(0.468)	(0.533)	(0.529)
Var. income (sec.)	0.244	0.047	0.308	0.440
	(0.567)	(0.650)	(0.695)	(0.691)
Prob. of employment (prim.)	-0.180	-0.129	-0.225*	-0.177
	(0.115)	(0.089)	(0.137)	(0.110)
Prob. of employment (sec.)	0.214**	0.155*	0.269**	0.223**
	(0.089)	(0.085)	(0.107)	(0.108)
Regional and birthyear dummies	✓	✓	✓	✓
Controls		✓		✓
Observations	1022	1022	1022	1022

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. The dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. Where indicated, I include controls for Gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size, number of children, rural and Skopje dummies, household asset group and distance from the closest school. Year and semester of birth dummies and Regional dummies are included. Asset groups are defined by using principal component analysis and using indicators of asset and land ownership at the time in which expectations are reported.

Appendix A The Model

The decision problem of parent i is defined by the following maximisation problem:

$$\max_{\delta_i \in [0,1]} \delta_i [\beta_i \frac{U(y_s)}{c_i \epsilon_i}] + (1-\delta)\,(1+\beta_i)\,U(y_p)$$

where β_i is the time discount rate and δ_i is equal to 1 if the parent enrols the child in secondary school and equal to zero if he doesn't enrol the child. Assuming that income subjective expectations are distributed with probability density $f^i_{Y_s}(y_s)$ if secondary school is completed and $f^i_{Y_p}(y_p)$ if primary school is completed, we can observe that the child will be enrolled in secondary school ($\delta^*_i = 1$) if the (discounted) expected utility from the completion of secondary school is larger than the (discounted) expected utility from the completion on primary school only:

$$\beta_i \frac{1}{c_i \epsilon_i} \int U_i(y_s)\, f^i_{Y_s}(y_s)\, dy_s > (1+\beta_i) \int U_i(y_p)\, f^i_{Y_p}(y_p)\, dy_p$$

where y_j follows a log-normal distribution $\ln N(\mu_j, \sigma_j)$ and $U_i(y_j)$ is a CRRA utility function in income. Therefore the expected value for y_s is equal to

$$\begin{aligned} E[U(y_j)] &= \int_0^{+\infty} U_i(y_j) f^i_{Y_j}(y_j)\, \mathrm{d}y_j \\ &= \frac{1}{\sqrt{2\pi}\sigma_j} \int_0^{+\infty} \frac{y_j^{1-\gamma}}{1-\gamma} \frac{1}{y_j} \exp\left[-\frac{(\ln y_j - \mu_j)^2}{2\sigma_j^2} \right] \mathrm{d}y_j \end{aligned}$$

We can now apply the transformations $\ln y_j = x_j \Rightarrow \frac{1}{y_j}\mathrm{dy_j} = dx_j$ and $y_j^{1-\gamma} = \exp\left((1-\gamma)\ln y_j\right)$ to rewrite

$$
\begin{aligned}
E[U(x_j)] &= \frac{1}{\sqrt{2\pi}\sigma_j}\int_{-\infty}^{+\infty}\frac{\exp\left((1-\gamma)\,x_j\right)}{1-\gamma}\exp\left[-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right]\mathrm{d}x_j \\
&= \frac{1}{1-\gamma}\frac{1}{\sqrt{2\pi}\sigma_j}\int_{-\infty}^{+\infty}\exp\left[(1-\gamma)\,x_j-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right]\mathrm{d}x_j \\
&= \frac{1}{1-\gamma}\frac{1}{\sqrt{2\pi}\sigma_j}\int_{-\infty}^{+\infty}\exp\left[\frac{2\sigma_j^2\,(1-\gamma)\,x_j-x_j^2-\mu_j^2+2x_j\mu_j}{2\sigma_j^2}\right]\mathrm{d}x_j \\
&= \frac{1}{1-\gamma}\frac{1}{\sqrt{2\pi}\sigma_j}\int_{-\infty}^{+\infty}\exp\left[\frac{-x_j^2+2x_j\left[\mu_j+\sigma_j^2\,(1-\gamma)\,x_j\right]-\mu_j^2}{2\sigma_j^2}\right]\mathrm{d}x_j \\
&= \frac{1}{1-\gamma}\exp\left[\frac{\left[\mu_j+\sigma_j^2\,(1-\gamma)\right]^2-\mu_j^2}{2\sigma_j^2}\right]\frac{1}{\sqrt{2\pi}\sigma_j}\cdot \\
&\quad\cdot\int_{-\infty}^{+\infty}\exp\left[\frac{\left[x_j-\left(\mu_j+\sigma_j^2\,(1-\gamma)\right)\right]^2}{2\sigma_j^2}\right]\mathrm{d}x_j \\
&= \frac{1}{1-\gamma}\exp\left[\frac{\left[\mu_j+\sigma_j^2\,(1-\gamma)\right]^2-\mu_j^2}{2\sigma_j^2}\right] \\
&= \frac{1}{1-\gamma}\exp\left[(1-\gamma)\left(\mu_j+\frac{\sigma_j^2\,(1-\gamma)}{2}\right)\right]
\end{aligned}
$$

We can now use the result for the expected utility to rewrite the condition for the child to be enrolled in secondary school ($\delta_i^*=1$):

$$
\frac{\beta_i}{1+\beta_i}\frac{1}{c_i\epsilon_i}\exp\left[(1-\gamma)\left(\mu_s+\frac{\sigma_s^2\,(1-\gamma)}{2}\right)\right]>\exp\left[(1-\gamma)\left(\mu_p+\frac{\sigma_p^2\,(1-\gamma)}{2}\right)\right]
$$

Taking logs of both sides we can rewrite and rearranging we obtain

$$
\Phi\equiv\ln\frac{\beta_i}{1+\beta_i}-\tilde{c}_i+(1-\gamma)\,\mu_s+\frac{(1-\gamma)^2}{2}\sigma_s^2-(1-\gamma)\,\mu_p-\frac{(1-\gamma)^2}{2}\sigma_p^2>\tilde{\epsilon}_i
$$

where $\tilde{c}_i\equiv\ln c_i$ and $\tilde{\epsilon}_i\equiv\ln\epsilon_i$. We can note that since ϵ follows a lognormal distribution, then $\tilde{\epsilon}$ follows a Gaussian distribution,$\tilde{\epsilon}\sim N(0,\sigma_\epsilon)$. Using Φ and the symmetry property of the distribution of $\tilde{\epsilon}_i$, we can then write the probability of the child to be enrolled in secondary school as

$$
\begin{aligned}
Pr\left[\delta_i = 1 \,|\, \Phi\right] &= Pr\left[\Phi > \tilde{\epsilon}_i \,|\, \mu, \sigma, \mathbf{X}\right] \\
&= 1 - F_{\tilde{\epsilon}}\left(\Phi\right) = F_{\tilde{\epsilon}}\left(-\Phi\right)
\end{aligned}
$$