Essays on intra-household resource control, subjective expectations and human capital investment

Alex ARMAND

A Dissertation submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PHD)
University College London (UCL)

London

May 2014

Declaration

I, Alex Armand, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in this thesis.

Alex Armand

Abstract

Empowering women and understanding its role within households is one of the main policy objectives in the developing world. However, while there is consensus on this objective, there is still little evidence on which is the most efficient way to approach this issue. The PhD thesis is based on understanding whether a variation in intra-household control of resources has an effect on household outcomes in developing countries and whether this effect is related to subjective expectations. This research question is approached using empirical microeconomics methods. Chapter "The Marriage market and intra-household allocation of expenditure" tests the assertion that the status of the marriage market impacts on intra-household allocation of expenditures. Chapter "Validation of Subjective expectations" analyses the validity of questions related to subjective expectations using the data collected among social financial recipients in the Republic of Macedonia during the 2010 and 2012 data collection waves of the Macedonian "Secondary School Conditional Cash Transfer" evaluation household survey. Chapter "Parental perceived returns to schooling and human capital investment" analyses how parental subjective expectations about the return to schooling of their children affect future decisions about schooling. Chapter "Who wears the trousers in the family?" studies how the interaction between intra-household allocation of resources and expected returns to schooling influences human capital investment among poor households.

Contents

List of Tables

List of Figures

Acknowledgments

Over the past years I have received support and encouragement from a great number of individuals. First and foremost I want to thank my advisors. I would like to thank Prof. Pedro Carneiro for his guidance throughout this rewarding journey. It has been an honour to be his Ph.D. student. I am also grateful to Dr. Valerie Lechene for her invaluable support and help throughout my Ph.D. studies. I'm grateful to all the Staff from the Conditional Cash Transfer 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. My time at UCL was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for the time spent with colleagues, for many other people and memories that I will bring with me for the next step.

Lastly, but not less importantly, I would like to thank my family for all their love and encouragement. I would like to thank my parents, Livio Armand and Daniela Sartor, for raising me up the way I am, for their presence and for supporting me in all my pursuits.

Funding acknowledgments

I gratefully acknowledge the funding sources that made my Ph.D. work possible. I was funded by the European Social Fund and the Regione Autonoma Valle d'Aosta for first 3 years of my Doctoral studies. My work was also supported by 3ie - Impact Evaluation through the Policy Open Window grant.

Conjoint work

All chapters are entirely my own, as of course is the responsibility for any errors.

Chapter 1

Introduction

Empowering women is one of the main policy objectives in the developing world. However, while there is consensus on this objective, there is still little evidence on which is the most efficient way to approach this issue and how dynamics work within households. This thesis is based on understanding whether a variation in intra-household allocation of resources has an effect on household decisions in developing countries and whether this has a relationship with information and subjective expectations. A first approach is to consider an exogenous variation in variables affecting the formation of the household, such as the marriage market. Another strategy is to study a policy that exogenously switch resources between household members. In the thesis I followed both approaches by looking at the Macedonian Conditional Cash Transfer for Secondary School Education, which is a program aiming at increasing secondary school attendance across poor households in the Republic of Macedonia and whose payments are randomised across the country in relation to the gender of the recipient (the mother versus the household head in the household). Although there are several "first generation" impact evaluation studies assessing the impact of conditional cash transfer programs, there are far fewer "second generation" studies focusing on the question of how to design CCTs more efficiently and specifically whether the gender of the recipient matters.

As a first approach to understand whether variables affecting the formation of the household, such as the marriage market, has an effect on intra-household dynamics, chapter 2 tests the assertion that the status of the marriage and remarriage market impacts on intra-household allocation of expenditure by looking at the pre-policy structure of expenditures of poor households in Macedonia. I use

measures of both the marriage and the re-marriage market to capture differences about these two variables to understand whether households take decisions as a unitary model or whether they conform to a collective model. To do so, I study the effect of local variation of sex ratios and of the shares of new weddings attributed to a previous divorce to capture variation in intra-household distribution of power.

Understanding whether the distribution of power within household members affects the decision to spend on specific good categories is central since, at the same time, there is still no consensus about how household members interact to take important investment decisions, such as determining human capital investments for the children. Understanding how this decision is taken is particularly important in developing countries, where households tend to underinvest due to market failures related to information, such as incomplete or asymmetric information. For this reason, in Macedonia, information about the parental perceived returns to schooling was collected to understand how information interact with intra-household dynamics to determine child human capital investment. Chapter 3 analyses the validity of questions related to subjective expectations using the data collected among social financial recipients in the Republic of Macedonia during the 2010 and 2012 data collection waves of the Macedonian "Secondary School Conditional Cash Transfer" evaluation household survey. Data on subjective expectations have been collected using the method proposed by Guiso et al. (2002). Under distributional assumptions, this method allows eliciting the subjective expected earning distribution for each respondent. This chapter analyses the validity of the data by testing whether respondents understand the question and use sufficient mental effort to report their answers. Using randomisation in the order of questions, I find that when asked about the probability to earn less than a certain threshold and more than the same threshold, respondents tend to reply with sufficient mental effort only to the first question they are asked.

To understand how subjective information affect human capital investment, chapter 4 investigates the role of parental expected returns to schooling as determinants of schooling decisions. The chapter analyses the relationship between schooling decisions and ex-ante parental perceived returns to secondary school. 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. In addition, by using the longitudinal dimension of the data, I provide evidence

against cognitive biases in expectation reporting and against endogeneity issues, which provide support for the use of subjective data in decision models.

Since both intra-household power allocation and the subjective expectations are credible determinants of schooling decisions, chapter 5 studies how the interaction between intra-household allocation of resources and expected returns to schooling influences human capital investment among poor households. To do this I use a nationally implemented randomised programme in the Republic of Macedonia. This programme provides cash transfers to poor households conditional on secondary school enrolment of their children and payments are transferred either to mothers or to household heads. I make use of a unique dataset with information on parental subjective expectations of returns to schooling and employment risk to estimate the heterogeneity of the effect of the intervention. This chapter provides evidence that targeting mothers allows increasing secondary school enrolment only for children whose returns are sufficiently high at the beginning of the program. This outcome is driven by increases in individual expenditure shares on education for children in the highest tercile of the return distribution. Heterogeneous effects along the expected return distribution are supported by large ex-ante heterogeneity in parental expectations. At the same time, no effect is recorded for other inputs, such as monitoring of children school attendance and parental time use.

Chapter 2

The Marriage market and intra-household allocation of expenditure

2.1 Introduction

Groups don't usually behave as a single rational individual and it is known at least since Arrow's Impossibility theorem. As a group of individuals, households may not behave as a single entity whose objective is to channel resources towards the best use of the family. Rather, members of the households bargain for maximal shares of the household resources and relative power may have fundamental consequences on the final decision of consumption.

Understanding the role of each member in familial interactions is key to predicting the response of households to public policy. As an important example, we can note that virtually all CCT programs in the world transfer money to women, but it is still uncertain if the registered effect is due to the release of budget constraint or a change in the use of resources by the household following an increase in the decision power of the wife. This understanding is fundamental in countries where the empowerment of women may have important indirect implications on investments in children's human capital through behavioural changes within households. This chapter studies the intra-household allocation of expenditure among Social Financial Assistance recipients in Macedonia by analysing the role of the marriage and the re-marriage market. In order to understand the be-

haviour of households, it is fundamental to understand how these interact with their environment. Studying the marriage market allows understanding how the power of women within their families varies depending on the strength of its bargaining position. By looking at the effect of local variation of the sex ratio and of new weddings shares where one of the two partners is divorced, I can proxy for the bargaining power of each partner and study how the household allocates expenditure to different goods.

A large body of research reports evidence that the amount of resources that each household member contributes to the family affects its allocation of consumption (for a literature review, see Duflo, 2005), however, there is no clear consensus on the precise mechanism through which households take decisions and allocate consumption. Different studies in the literature tested for the validity of the unitary model by using measures of relative income between husband and wife. Thomas (1990) and Schultz (1990) test for the effect of individual unearned incomes on different outcomes[1], while Bourguignon et al. (1993), Bourguignon et al. (1994) and Phipps and Burton (1998) provide evidence against the unitary model by using measures of relative income. However, using this measure to underpin the distribution of power inside the household may be disturbed by the endogeneity problem, since the relative earned/unearned income might depend on decisions that are not independent from the observed outcome.

Evidence on the importance of each member's role inside the household gave support to theories of intra-household allocation based on relative power within the familial environment. The main theoretical contribution to this research line is the collective rationality model (Chiappori, 1992), which assumes that decisions carried out within households are Pareto efficient. However, there is still controversy around this assumption. Evidence from a number of studies shows that in developed countries, households' consumption decisions tend to be Pareto efficient (Bourguignon et al., 1994; Browning and Chiappori, 1998; Chiappori et al., 2002), while other recent studies in developing countries provide mixed evidence against and in favour of efficient decisions. Evidence of inefficiency has been provided for agricultural choices (Udry, 1996, shows evidence for Nigeria) and for the inability to provide full insurance against individual income shocks in the marital contract (Duflo and Udry, 2004), while Bobonis (2009) and Attanasio

[1]Thomas (1990) finds evidence for the effect of husband's and wife's unearned income on labour supply and fertility decisions using the 1981 Socioeconomic Survey of Thailand. Schultz (1990), using a survey on household conditions in Brazil, rejects income pooling hypothesis by observing an effect of mother's unearned income on family health and nutrition status.

and Lechene (2011) support the efficiency assumption for consumption choices in rural Mexico. Related to evidence of inefficient choices, a growing theoretical literature states that the presence of imperfect marital contracts (Lundberg and Pollak, 1993, 2003; Ligon, 2002; Basu, 2006) or asymmetric information (Bloch and Rao, 2002) can lead to inefficient allocation of resources.

Testing for individual rationality or for alternative models of intra-household behaviour (such as the collective model), we require an exogenous shift in the intra-household allocation of power. Examples of such variation used in literature are policy interventions shifting power towards one of the household members. Lundberg et al. (1997) and Ward-Batts (2008) use the 1979 UK Reform of Child benefits, which exogenously shifted the recipient of the benefit towards mothers; they both provide evidence towards the rejection of income pooling. Attanasio and Lechene (2002) use the Progresa Program and its experimental nature to test for income pooling by instrumenting the wife's income share with the transfer generated by the program (which was targeting women); they confirm previous rejection of the income pooling hypothesis. The same experimental feature of Progresa is exploited by Bobonis (2009), who combines randomised variation in women's income and local rainfall shocks as distribution factors; he finds evidence that wife-specific income changes increase expenditure on children's goods, while income shocks generated by rainfall shocks have a smaller effect on public goods expenditure. A large CCT program such as Progresa is used as well by Attanasio and Lechene (2011), who test the restrictions implied by collective rationality estimating a z-conditional demand system in rural Mexico. By using the random allocation of cash transfers to women and the relative size of husband and wife's family network as distribution factors, they reject the unitary model and cannot reject efficiency in household decisions.

Other sources of variation in the intra-household allocation of power that are readily observed and have been used in different studies focus on elements of the household environment. Chiappori et al. (2002) use the state of the marriage market (measured by the sex ratio by age, race and state and by the features of the legislation on divorce) as a distribution factor in the intra-household decision for labor supply. They find that sex ratios associated with stronger power for wives (fewer women are associated with stronger bargaining power within the marriage contract) significantly decrease the labour supply of women. Other distribution factors, such as the legislation of abortion (Oreffice, 2007), the generosity of single parent benefits (Rubalcava and Thomas, 2000), the distribution of wealth

by gender at marriage (Thomas et al., 1997) and the pension recipient's gender (Duflo, 2003), showed to have a significant effect on intra-household allocation of power.

This chapter focuses on the specific role of the marriage and the re-marriage market in explaining intra-household allocation of expenditure by looking at measures of the relative bargaining power of both spouses and by looking at the environment that characterise their marital contract. However, understanding the role of the marriage market on the intra-household allocation of expenditure requires a precise measure of the relative power of partners in this market. The main idea is that factors influencing spouses' opportunities outside the wedding can have an effect over the intra-household distribution of power, even if the marriage is not dissolving. The idea was first proposed by Becker (1991), who proposed that the marriage market is an important factor influencing redistribution of utility across family members. The sex ratio, i.e. the relative scarcity of one gender versus the other, is a fundamental factor that explains the status of the market. One of the main results in the literature on intra-household decision and the marriage market is attributable to Chiappori et al. (2002). In this study, they used actual local sex ratios (together with indicators for the divorce law) to proxy for the distribution of power across spouses and for its effect on labour supply decisions for both partners; they find evidence in favour of the marriage market hypothesis, for which households living in areas where women are scarcer have on average more decision power allocated to wives, resulting in a lower number of hours worked. Other studies showed evidence that an increase in the rarity of women in the local marriage market is associated with a decrease in the female work participation (Angrist, 2002; Grossbard-Shechtman and Neideffer, 1997).

As proposed by Thomas et al. (1997), another direct measure of the bargaining power across partners that is related to the marriage market is the relative economic power (for instance, through the relative wage) at the moment in which the marital contract is stipulated, namely at wedding. However, it might be possible that the past distribution of income across spouse is correlated with unobservable factors and this would bias the effect of such distribution factor. Moreover, a questionable matter is whether the important factors affecting the intra-household distribution of power can be related to the marriage market or to the re-marriage market. In other words, we need to understand whether the past distribution of power (at the time of the wedding) matters for the current allocation of consumption or it is the current bargaining power in the re-marriage

market that matters for household decisions. It is reasonable to assume that this problem may be explained by the degree of commitment that the wedding contract includes. If we assume, as it is reasonable for the setting of this chapter, that commitment to marriage is not perfectly enforced since partners cannot commit not to divorce, then in order to understand relative positions of partners we need to study the current conditions of the marriage market. In fact, Thomas et al. (1997) provide evidence that the distribution of income at wedding time doesn't affect intra-household balance of power in those regions where wealth is traditionally pooled within the household.

To answer this issue, in this chapter I will use measures of both the marriage and the re-marriage market to capture differences about these two variables. I use data collected in the Republic of Macedonia among the poorest quintile of the population to test for the unitary model by exploiting variation of local sex ratios as a source of distribution factor. I find that sex ratios with a higher percentage of women affect intra-household allocation of expenditure through an increase in the expenditure share of education and health and through a decrease of the share allocated to food. Given general results on preferences differences across gender, this provide new evidence on the relationship of sex ratio with labour market characteristics since, we would expect rarity of women to have a positive effect on the decision power of women within the household. One possible explanation is that the relative power of spouses affects only labour supply decisions and then, through a lower relative income for women, it influences the allocation of consumption. I propose that, at least in the case of the Republic of Macedonia where unemployment is very high (in 2008, in some regions unemployment reached a peak of 58% with an average of 32%) variations in the local labour market characteristics influence the relative labour demand of women compared to men and this in turn affect the relative income of partners. The results are supported by IV estimation for sex ratios.

Using the sex ratio as a measure for the status of the wedding market doesn't allow us to discriminate on the precise effect of the distribution factor on the bargaining position of each partner. I then concentrate on more direct measures of marriage market status by focusing on the municipality-level share of weddings in which the spouse had previously divorced from another partner (differentiating between men and women), and I consider the two years previous to the survey. Using this measure I want to proxy for the real possibility for each partner to exit from the actual marriage and to find another partner. Moreover, it is reasonable

to assume that this measure is exogenous since divorce decisions are taken by other households[2]. Secondly, this allows distinguishing for the bargaining power of each partner by exploiting the variation of share of weddings attributable to divorces across genders. Using these measures as distribution factors I test for income pooling and I find evidence against the unitary rationality hypothesis. Moreover, using the possibility to distinguish among each partner's bargaining power, I provide evidence that is consistent with the effect of the sex ratio and in favour of the efficiency assumption.

The structure of the chapter is organised as follows. Section 2.2 presents the theoretical framework. Section 2.3 report information about the data and its descriptive statistics. Section 2.4 describes the empirical strategy and its validity. Section 2.5 discusses the results.

2.2 Theoretical framework

This section makes direct reference to the general version of the collective model proposed by Bourguignon et al. (2009), which allows intra-household allocation of consumption to be an efficient outcome whose position on the Pareto frontier depends on the relative power of decision makers within the household. I allow for the presence of different kind of goods, namely private and public goods. Consider a static version of the household collective model where husband and wife are the 2 decision makers $m = \{H, W\}$. Consumption can be allocated among n goods that are divided into private and assignable (n_a), public (n_p) and a Hicksian composite good C, that can be privately or publicly consumed, such that these are all possible goods that can be consumed by the household ($n = n_a + n_p + 1$). I denote the vector of privately-consumed goods as $q^m \in \mathbb{R}^n_+$ and the vector of publicly-consumed goods as $Q \in \mathbb{R}^n_+$. Each household member's preference is represented by $u^H(q^H, q^W, Q, C, a)$ and $u^W(q^H, q^W, Q, C, a)$ where a are preference factors, e.g. a vector of individual and household characteristics that affect preferences directly through utility functions. Preference factors are divided into observable (a_o) and unobservable (a_u) individual and household characteristics, such that $a = \{a_o, a_u\}$ and the utility functions are then defined by $u^m(q^H, q^W, Q, C; a_o, a_u)$. I define z as a K-vector of distribution factors such that every z_k influences the allocation of consumption only through the allocation of power among household members, but it is not influencing individual preferences and the budget constraint.

[2]The issue of possible endogeneity is discussed in Section 2.5.2.

Any efficient allocation of consumption can be characterised as a solution of the following program

$$\begin{aligned} &\max_{\{q^H,q^W,Q,C\}} && u^H(q^H,q^W,Q,C;a_o,a_u)+\lambda u^W(q^H,q^W,Q,C;a_o,a_u) \\ &\textit{subject to} && p^Hq^H+p^Wq^W+p^QQ+p^CC\leq y \\ & && \lambda=\lambda(y,\mathbf{p},\mathbf{z};a_o,a_u) \\ & && y=y^H+y^W+y^J \end{aligned} \tag{2.1}$$

where $\mathbf{p}=(p^H,p^W,p^Q,p^C)$ is a vector of prices for all kind of goods, λ is the Pareto weight, y is total household income, y^H and y^W are individual incomes of husband and wife and y^J is income jointly held by both partners. The resulting demand is then a function of total resources (or expenditure, denoted by x), prices, individual and household characteristics and the Pareto weight which influence the decision power of each partner in the household:

$$g=\xi(x,\mathbf{p},\lambda(y,\mathbf{p},\mathbf{z};a_o,a_u);a_o,a_u) \tag{2.2}$$

where $g=(q^H,q^W,q^Q,C)$. The Pareto weight $\lambda=\lambda(y,\mathbf{p},\mathbf{z};a_o,a_u)$ affect the sharing rule which determines the allocation of available resources to partners and the sharing rule is influenced by distribution factors that determines the power in the decision making process for each decision maker. It is common in literature to consider partner-specific incomes (y^H,y^W,y^J) as distribution factors, while this choice, as I will explain later, may generate empirical problems due to the endogeneity of incomes. In order to overcome this problem I will use alternative distribution factors by exploiting the variation in the characteristics of the wedding market.

The first testable prediction of the collective model follows from the demand system 2.2 and is the rejection of the unitary rationality hypothesis. In other words, we want to test that the allocation of consumption is not affected by any distribution factor. In contrary, the demand system is compatible with unitary rationality if and only if it satisfies the following condition for any $i=1,..n$ and $k=1,..K$:

$$\frac{\partial\xi_i}{\partial z_k}=0 \tag{2.3}$$

Testing for income pooling allows us to reject the unitary model under the assumption that the chosen distribution factor affects the allocation of consumption only through the Pareto weight and not through partner's preferences. In our case, testing for individual rationality means to test for the joint significance of the sex ratio and of measures of divorce rates in the demand system. Joint significance of these measures, allow us to claim that the distribution of power across partners do matter for the allocation of expenditure to different consumption goods. If we assume two distribution factors, (z_1 and z_2), we can express the demand in the form $g = \xi(x, \mathbf{p}, \lambda(y, \mathbf{p}, z_1, z_2; a_o, a_u); a_o, a_u)$. By computing the partial derivatives of the demand with respect to the two distribution factors, we obtain:

$$\begin{aligned} \frac{\partial \xi}{\partial z_1} &= \frac{\partial \xi}{\partial \lambda} \cdot \frac{\partial \lambda}{\partial z_1} \\ \frac{\partial \xi}{\partial z_2} &= \frac{\partial \xi}{\partial \lambda} \cdot \frac{\partial \lambda}{\partial z_2} \end{aligned} \tag{2.4}$$

If we consider one specific good i, we can compute the ratio of the first derivatives with respect to the two distribution factors and we can show that such ratio is constant since $\lambda(.)$ doesn't depend on i. The ratio is then equal to:

$$\frac{\partial \xi_i / \partial z_1}{\partial \xi_i / \partial z_2} = \frac{\partial \lambda / \partial z_1}{\partial \lambda / \partial z_2} \tag{2.5}$$

The ratio of partial derivatives of demand for good i with respect to each distribution factor and conditional on total household expenditure, is equal across all goods and equal to the ratio of partial derivatives of the Pareto weight with respect to the distribution factors. This condition is know as the proportionality condition. The main idea behind such condition is that the effect of distribution factors is equally proportional for all goods since each distribution factor affects demand allocation only through its effect on the Pareto weight. Hence, their impact is one-dimensional, in the sense that they affect allocation as if there was only one distribution factor. In addition, recently, Bourguignon et al. (2009) showed that this condition is necessary and sufficient for efficiency. The main testable prediction for collective rationality is then that for any good $i, j = 1, ..n$ with $i \neq j$ and any distribution factor $k, k' = 1, ..K$ with $k \neq k'$

$$\frac{\partial \xi_i / \partial z_k}{\partial \xi_i / \partial z_{k'}} = \frac{\partial \xi_j / \partial z_k}{\partial \xi_j / \partial z_{k'}} = \frac{\partial \lambda / \partial z_k}{\partial \lambda / \partial z_{k'}} \tag{2.6}$$

Bourguignon et al. (2009) shows that, if we consider two distribution factors (z_1 and z_2) and we assume that the Pareto weight λ is increasing function of the

distribution factor z_1 and decreasing function of the distribution factor z_2 , we can conclude that demand functions consistent with any bargaining model are such that for any good the ratio of partial derivatives of the demand with respect to the two distribution factors is equal across goods and is negative. For any $i = 1, ..n$, the condition can then be expressed in the following way.

$$\frac{\partial \xi_i / \partial z_k}{\partial \xi_i / \partial z_{k'}} = \frac{\partial \xi_j / \partial z_k}{\partial \xi_j / \partial z_{k'}} < 0 \tag{2.7}$$

In the rest of the chapter I will estimate an empirical version of the model presented in this section and perform tests of individual rationality and collective rationality based on different distribution factors. In order to understand the environment and the main characteristics of the sample, the following section presents the sources of data used and its main descriptive statistics.

2.3 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. Specifically the secondary data used in the chapter comes from the 2002 Census of Population (State Statistical Office, 2002), the 2007 Census of Agriculture and the 2007-2010 Population and Vital statistics (all provided by the Macedonian State Statistical Office).

For the scope of the program evaluation of the Macedonian CCT program for secondary education, two household surveys were collected during the Winter 2010, at the beginning of the program, and in Fall 2012, after two years of implementation. This chapter makes use of the baseline household survey was conducted between November and December 2010, coinciding with the beginning of

the first school year in which CCT program became available. The Macedonian "Conditional Cash Transfer (CCT) for Secondary School Education" is a social protection program aiming at increasing secondary school enrolment and completion rate among children in the poorest households of the population. It was first implemented by the Macedonian Ministry of Labour and Social Policy in Fall 2010 to provide cash transfers to poor households conditional on having children in school-age attending secondary school at least 85% of the time. In order to target poor households, the Ministry of Labour and Social Policy decided to offer the program to the beneficiaries of the Social Financial Assistance (SFA) benefit, which is the most significant income support program, accounting for around 0.5 percent of GDP and 50 percent of total spending on social assistance (Verme, 2008). SFA is a mean-tested monetary transfer granted to people who are fit for work, are socially not provided for and cannot support themselves. The amount paid for SFA is equivalent to the difference between household income and the social assistance amount determined for the household, depending on household size and time spent in SFA, varying from 1 825 MKD (around 40 USD) for one-member household to 4 500 MKD (around 98 USD) for households with 5 or more members. It is considered as the benefit of last resort, meaning it is provided after other benefits if the household income is still below a certain living standards threshold. It is mainly collected by households in the poorest tail of the income distribution; in 2009, the World Bank[3] reports that total SFA benefits are collected for 55 percent by the poorest quintile, 22 percent by the second poorest quintile and 11 percent to the middle quintile.

At baseline, a sample of Financial Assistance households (recipients of SFA and Child Benefit) 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.

A sample of recipients of SFA and Child Benefit 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

[3]I make reference to the "Project Appraisal Document - Report No: 47195-MK" between the Former Yugoslav Republic of Macedonia and the World Bank.

the implementation of the program[4]. 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[5].

The survey 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). Moreover information on family networks (given by the number of visits from/to relatives for each member) and on domestic violence attitudes are collected. The survey asks for expenditure on food, alcohol and tobacco over the last week, on education, health, home, entertainment and services over the last month, on clothes in the last 3 months. The presence of questions with different length of time for different goods depending on their frequency of purchase reduce the problem of imperfect coincidence between actual consumption, which is directly measured, and observed expenditure (seeBlundell and Meghir, 1987, for an exposition of the purchase-infrequency model).

Table 2.1 and Table 2.2 presents descriptive statistics about household structure, divided by type of municipality of residence of the household (rural versus urban) and across different ethnic groups. Households are composed on average by 4.7 members with an average age of the household head of 44.9. Each household has 1.6 children in the age group 13-18 years old and fewer children with a lower age (0.15) in the 0-6 class age and 0.44 in the 7-12 class age). If we look at ethnic composition[6], the majority of households in the sample is Macedonian (45.6 percent), versus a 30.5 percent of Albanian households, 12.9 percent of

[4]All estimations are weighted by sampling weights in order to reconstruct the SFA and Child Benefit population at the time of the interview.

[5]We aimed for a sample size of 17 households eligible for the CCT (recipients of social and financial assistance with children of secondary school age) per municipality, although in practice there was some variation in this number due to the fact that in some municipalities the eligible population was smaller then 17. For power calculations, we considered a power of 0.8 and a significance level of 0.05. With 42 clusters per arm and an inter-cluster correlation of 0.25, using 17 households per municipality it would be possible to detect a difference in expenditures in children's education (or in any other item) of 0.33 of a standard deviation and an increase in the proportion of students attending 85% or more of the classes of roughly 10% points.

[6]Ethnicity refers to the ethnic group reported by the household head. The main ethnicities present in the Republic of Macedonia are the Macedonian, Albanian, Roma and Turk ethnicities. Mixed marriages are rare exceptions in the sample.

Roma households and 11.0 percent of Turk households. Macedonian households have less members than Albanian and Roma households (4.3 versus 5.2 and 4.9 members) and on average have a lower number of children in the 13-18 years old age group (1.5 versus 1.75 and 1.53 children). Additional descriptive statistics are presented in Table 2.11.

In order to understand the distribution of decision power across partners, it is necessary to look at the characteristics of husbands and wives. Table 2.3 presents the main individual characteristics for these individuals in terms of age, education and wage share at the time of the wedding. Wives are less educated than husbands, with 20.8 percent who completed at least secondary school, compared to 33.0 percent for husbands. On average, wives are 2.5 years younger than husbands.

2.3.1 Expenditure and budget shares

Expenditure data was collected using a recall method, based on the frequencies reported by the Macedonian Household Budget Survey, which is collected by the Macedonian State Statistical Office with the purpose to identify expenditure attitudes by the average Macedonian household. The expenditure section of the questionnaire was divided into sub-sections depending on the characteristics of the goods and their proposed frequency of purchase. For food items, the questionnaire initially asked about the quantities purchased and expenditure in the last 7 days. The food aggregate is composed by the following categories: cereals and wheat products, vegetables, fruit, meat, fish and other foods. Other foods contains oils and fats, sugar, salt, honey and jam, chocolate, sweets and cookies, salties, coffee and cocoa, infant food, spices, teas, sauces, instant soups and meals and other beverages (mineral water, soft drinks, fruit juices). For non-food items, information was collected on the purchases during the period of reference and about the items that were received for free or as a payment in kind. For this kind of goods it is more complicated to impute values of items that were not purchases, since the heterogeneity in quantity and quality is higher. To this extent, the respondent was then asked to imagine to be in the situation to buy these goods and to report at which price they could have found them at the local market.

In order to have a measure of comparison and to compute expenditure shares, all figures have been converted to monthly expenditures. This allowed constructing an aggregate measure for total household expenditure. The expenditure shares allocated to different categories are then determined as the ratio between the amount spent on the category and the total household expenditure. In order to exclude problems of infrequency related to the decision to purchase durable goods, the expenditures on durable goods were removed from the analysis. Table 2.4 presents summary statistics on monthly household expenditure and on budget shares allocated to different categories. On average, households spend 22.126 MKD monthly[7](roughly 290 British pounds). The highest budget share is allocated to food with an average share of 47.2 percent. We can note that some households spent the whole expenditure on food, showing that the living conditions of families are fairly poor. Among sub-categories, cereals and vegetables are the mainly component of the food expenditure with an average share of respectively 17.69 and 10.93 percent. The share allocated to meat is lower (5.83 percent), while the one for fish and fish-related products drops to 0.42 percent. Dairy products captures around the same share as meat, with 5 percent of expenditure on average. Clothing corresponds to around 4.7 percent, while expenditure on children education collect on average 13.0 percent of total household expenditure, with a slightly bigger average share for girls. It is interesting to note that on average households spent around 3.1 percent on alcohol and tobacco. Health expenditures reach instead an average share of 8.6 percent, while house and hygiene-related expenses capture 7.4 percent, compared to 14.1 percent for utilities.

I will refer to budget shares as monetary expenditure allocated to a specific good as a percentage of total household expenditure. An alternative strategy would be to consider consumption shares, since expenditure decisions might differ from actual consumption. Using expenditure shares allows avoiding measuring error linked to the imputation of values for consumption, but if household expenditure differs systematically from consumption this could lead to a misinterpretation of the results. To check whether there are substantial differences between expenditure and consumption decision, a household consumption aggregate was built using information on food consumption and non-food consumption. In order to build an aggregate for food consumption, we take into account not only what the household spent on food purchases, but what the household actually consumed during the last 7 days. In order to impute the value of consumed food (either purchased, self- produced, stored or received as gift), the quantities reported by

[7]1 Macedonian Denar = 0.015 British pounds.

the household are multiplied by a set of prices using a criterion of proximity to the household. The first choice is to use prices directly paid by the household, but for a large number of households, goods are either purchased or self-produced, so that prices at household level might not always be available. In order to overcome this problem, we compute median prices using observations on other households at different geographical levels (municipality, region and Macedonia). To be sure, in order to compute values of self-produced goods, we should use farm-gate prices, that would more precisely estimate the value of such goods, while using market prices would include intermediaries markup into the purchasing price. However, given the size of the country and its relative degree of closeness to international market, it is reasonable to assume that observed food prices are very close to farm-gate prices, since there is large availability of locally produced green markets, where prices are relatively lower than goods purchased at other distribution chains. For non-food items, consumption is computed by adding the value of expenditures and the value of good received without paying, using as a value the self-reported price at which the respondent could buy the same good in the local market. Table 2.4 presents a comparison of expenditure and consumption aggregates (and its composition) for households in urban and rural areas. We can notice that there are little differences between the two measures, showing that expenditure and consumption decisions are not very dissimilar. For this reason, we proceed analysing only expenditure shares since they present a lower degree of measurement error.

2.4 Empirical strategy

This Section presents the econometric model used to asses the validity of the theoretical framework presented in Section 2.2. As previously explained, in order to test for unitary model of household decision, I express household demand as a function of total (log) household expenditure, distribution factors and a set of individual and household specific characteristics[8]:

$$w_{ih} = \alpha_i + \beta_i ln(exp)_{ih} + \gamma \mathbf{z}_h + \mathbf{\Omega X_h} + \epsilon_{ih} \tag{2.8}$$

where w_{ih} is the budget share spent on good i in household h, z_h is a vector of distribution factors, X_h is a vector of wife, husband and household characteristics

[8]In this specification I allow prices to be part of the error term. The potential problems are discussed along with the chapter.

and ϵ_{ih} are unobservable determinants of demand for each good. Controls include household composition, gender and ethnicity of the household head, husband and wife age and schooling, regional dummies, dummies for whether the household live in rural areas and live in the capital city (Skopje) and number of financial assistance recipients in the municipality.

To the extent that total expenditure is determined by unobserved factors affecting household expenditure allocation, ordinary least squares (OLS) estimates would be affected by omitted variable bias. The presence of unusually high (or low) values of expenditure on a consumption good would generate a correlation between (total) expenditure and the unobservable factors in the error term. As shown by Browning and Chiappori (1998) this problem may be solved by using total household income as an instrumental variable for expenditure, since it is correlated with the total level of household expenditure, but conditioning on income should have no effect on the consumption shares allocated to different goods. As in most studies in literature, I consider the possibility of having total expenditure as endogenous and I instrument it with total household income[9], an asset index built using principal component analysis (see Section 5.8) and regional unemployment rate in 2010 (U_m). I allow for more flexible forms by higher order polynomials for total income and I find that the best fit is quadratic in total household income and linear in unemployment and asset index. Since household income is zero in some cases, I don't use a log specification to match for total expenditure[10]. First stage OLS regression of total expenditure using the chosen instruments is reported in Table 2.10; all regressors are significant at 1% level and joint-significance test of instruments reject the hypothesis of weak instrument.

In order to test for income pooling I then proceed by focusing on the marriage market and by using, as previously discussed, a measure of rarity of women, the sex ratio at husband's age, and a measure of the weddings share from divorces. I compute sex ratios at municipality level using the 2010 Population Statistics available from the Macedonian State Statistical Office. Population statistics allow me to know for each age class[11] the number of male and female individuals living in each municipality[12]. I compute the corresponding sex ratio x_h for household h

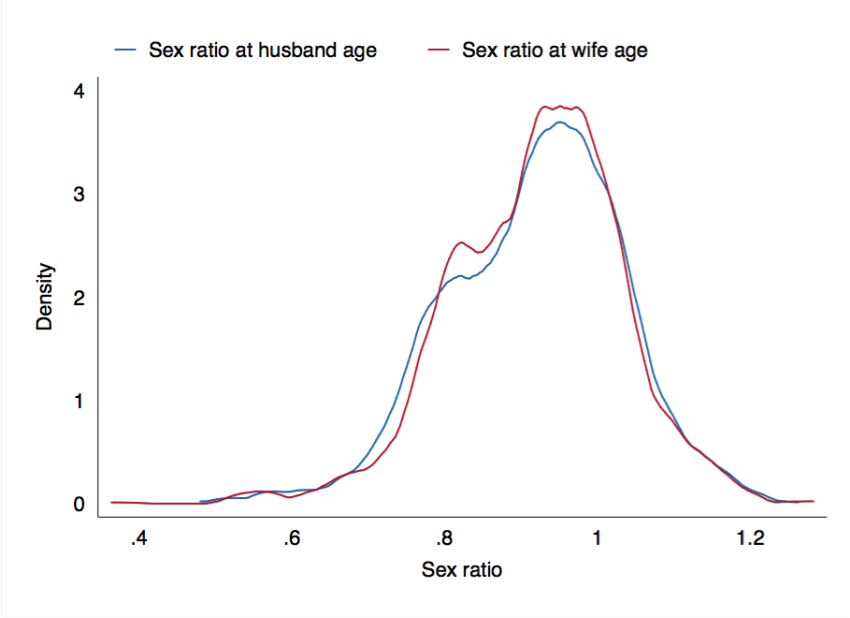
[9] I compute (yearly) total household income by using information on labour income (monthly salary paid and months worked), agricultural and breeding income and other sources of income.

[10] Ward-Batts (2000) follows a similar approach.

[11] Age classes are mainly of 5 years.

[12] Interracial marriages in the sample are rare, but it would be an improvement to use sex ratios at municipality and at ethnicity level. However, this information is not available.

Figure 2.1: Distribution of sex ratios at husband and wife age (2010)



Note. The graph shows a density estimation of the sample distribution of the sex ratio at wife and husband age. Sex ratios are defined as the ratio between the number of women and the number of men in the municipality of residence of the household and in the corresponding age group.

living in municipality m by

$$z_{h,m} = \frac{female_{m,a}}{male_{m,a}}$$

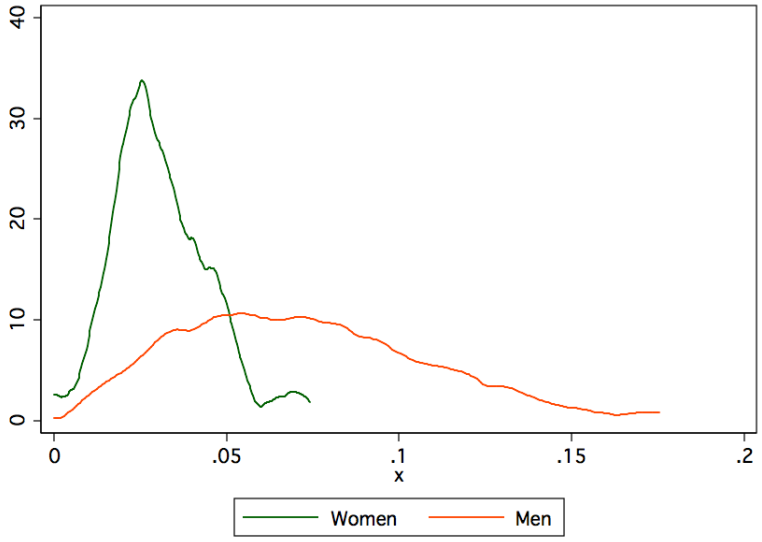
where $female$ is the number of female individuals and $male$ is the number of male individuals living in municipality m and being in the same age class a of husband in household h. Figure 2.1 presents the kernel density of sex ratios associated to wives' and husbands' age. Sex ratios at husband's age range from 0.48 to 1.263 percent, with an average value of 0.918. The distribution of sex ratios for husband's and wife's ages is very similar due to the close relationship between the age both spouses and the availability of data for 5-years age class.

Using the computed sex ratios, I estimate the following model for each good[13]:

$$w_{i,h} = \alpha_i + \beta_i ln(exp)_{i,h} + \gamma_i z_{h,m} + \mathbf{\Omega X_h} + \epsilon_i \tag{2.9}$$

[13]I estimated the model with a more flexible functional form by introducing higher orders polynomial in expenditure. The best fit is obtained using with a linear specification.

Figure 2.2: Distribution of weddings share attributed to divorces



Note. The graph shows a density estimation of the sample distribution of the share of weddings attributed to divorces for women and men. The share is defined as the total number of men (women) that got married after a divorce in the municipality of residence during the reference period.

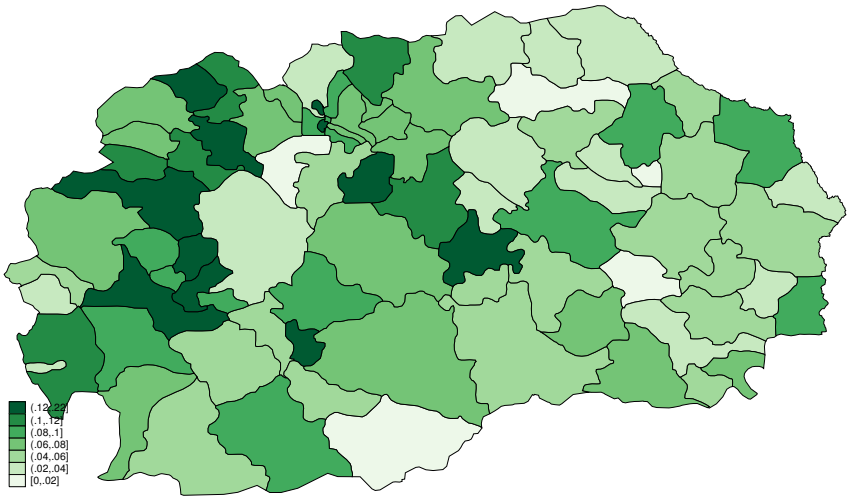
and I proceed testing for the unitary rationality of the household by checking for the joint significance of $z_{h,m}$.

The use of sex ratio as a distribution factor doesn't allow us to discriminate among the power of husband and wife within the household since the measure is reflecting the relative position of both partners. Moreover, the high correlation of the measure at husband's and wife's age doesn't allow us to use these two measure to proxy for the power of both partners. In order to test for collective rationality, I then divide the bargaining power in the marriage market by using as distribution factors the share of total weddings (for men and for women) in which the spouse had a previous divorce and I consider the three years previous to the survey (namely 2007-2009). In other words, if I indicate W_H (W_W) as the total number of men (women) that get married in the period 2007-2009 and W_H^{div} (W_W^{div}) as the total number of men (women) that get married in the same period but after a divorce, the computed distribution factors will be equal to:

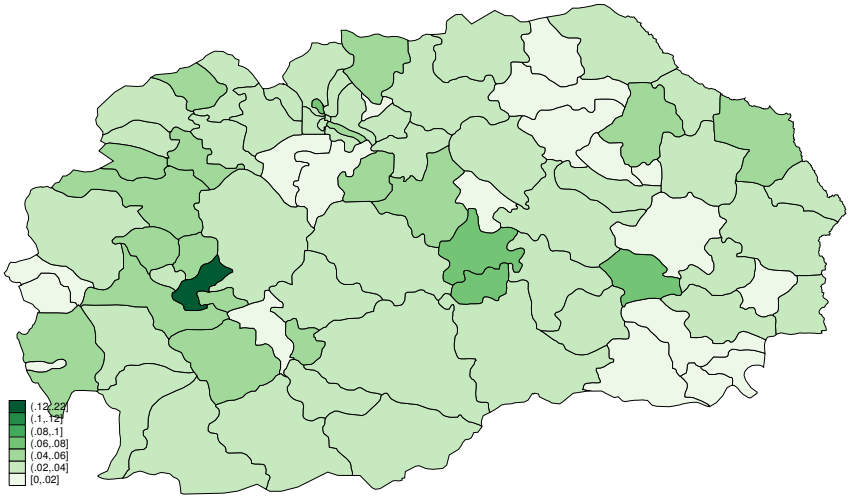
$$\begin{aligned} z_H &= {}^{W_H^{div}}/_{W_H} \\ z_W &= {}^{W_W^{div}}/_{W_W} \end{aligned} \tag{2.10}$$

Figure 2.3: Regional distribution of share of total weddings after a divorce (2007-09)

4a. Men



4b. Women



Note. The pictures show the geographical distribution of the share of weddings attributed to divorces for women and men. The share is defined as the total number of men (women) that got married after a divorce in the municipality of residence during the reference period.

It is reasonable to assume that the distribution factors z_H and z_W are correlated with the bargaining positions of wives ad husbands since they are measuring the possibility (or the individual perception) to exit the marriage and to re-marry with another partner. Nevertheless, the interpretation of the coefficient is ambiguous, since a higher number of men that re-marry after a divorce (compared to women) could be explaining the higher possibility of a husband to exit the marriage (increasing the bargaining power of men) and, at the same time, the higher availability of men in the marriage market (increasing the bargaining power of women). However, this paper aims at testing for the individual and collective rationality hypothesis and therefore will not be looking at different mechanisms through which different divorce rates affect intra-household decisions.

Figure 2.2 compares the distributions of z_H and z_W in the sample, while Figure 2.3 reports the regional distribution of shares of weddings after a divorce for men and women across Macedonia. It is easy to observe that the two distributions are quite different: the distribution of shares of weddings attributed to divorces is more concentrated for women and with a lower average, while the distribution for men is more dispersed. In the sample, z_W has an average of 3.09 percent (standard deviation 0.014) with a maximum value equal to 7.41 percent, while z_H has an average of 7.13 percent (standard deviation 0.036) with a maximum of 17.54 percent. This shows that if we look at the collective behaviour in the marriage market, we can obtain two measures that are reflecting wife and husband's position in the re-marriage market and in the meantime with a differential effect among the two partners. This allow us to test for the collective model since we can use as distribution factors two measures which affect the position of wife and husband in a different way. In this case, I use the two distribution factors z_H and z_W and I estimate for each good the following model:

$$w_{i,h} = \alpha_i + \beta_i ln(exp)_{i,h} + \eta_{i,H} z_{H,h} + \eta_{i,W} z_{W,h} + \mathbf{\Omega X_h} + \epsilon_i \tag{2.11}$$

The possibility to include two distribution factors in the estimation of the demand system allow us to test for the two following hypothesis:

1. **Individual rationality**. As reported in Section 2, the model is consistent with individual rationality if and only if for any consumption good $i = 1, ..n$

$$\frac{\partial w_i}{\partial z_H} = \frac{\partial w_i}{\partial z_W} = 0 \tag{2.12}$$

2. **Collective rationality.** The model is consistent with the efficiency assumption if and only if the proportionality condition is not rejected, namely if and only if for any consumption good $i, j = 1, ..n$, $i \neq j$

$$\frac{\partial w_i / \partial z_W}{\partial w_i / \partial z_H} = \frac{\partial w_j / \partial z_W}{\partial w_j / \partial z_H} \tag{2.13}$$

The following Section presents the main results obtained by following the proposed empirical strategy.

2.5 Results

Given that households decisions to allocate expenditure in different goods may be interrelated, we need to take into account the plausible cross-equation correlation of the error terms. This matter is solved by estimating the econometric model presented in Section 2.4 with system estimation methods. To understand the structure of household expenditure allocation, all tables present the estimated coefficients not only for the distribution factors, but for the total household expenditure. When the endogeneity problem related to total household expenditure is relaxed I estimate the system by seemingly unrelated regression (SUR), while for the majority of the results the system is estimated with Instrumental Variable via Full Information Maximum Likelihood (FIML). In every specification, I allow for standard errors to be clustered at municipality level. In all estimations of demand system, the sample is restricted to only couples and single parent households are excluded. I will consider the following good categories: food, adult clothing (male and female), child clothing, education, health, tobacco and alcohol and other goods (utilities, home expenditures and residual expenses).

Table 2.6 presents the estimated coefficients for the sex ratio at husband's age and for total expenditure[14]. In column (1), I estimate the model with SUR without instrumenting for total expenditure, while in column (2) total expenditure is instrumented using the IV strategy previously proposed. The test for unitary rationality is rejected at 1% in both cases. Under the assumption that the sex ratio conforms to the definition of distribution factor, it has a significant effect on the allocation of expenditure through the distribution of power across partners. Specifically, we observe a stronger effect on the allocation to expenditures on

[14]The whole regression is reported in Tables 14-16.

food, male adult clothing and education. A higher sex ratio leads to a significant increase on the expenditure of education, while reducing the share spent on food and adult male clothing. Given the generally accepted proposition that women have stronger preferences for collective consumption needs, this result provides evidence that a higher sex ratio might be related to a higher decision power of women, rather than the opposite. These results may be biased in the case the sex ratio is correlated with unobservable characteristics of the household that are related to current expenditure allocation. I will discuss this issue more in detail in Section 2.5.1.

When turning the attention to total expenditure, we can observe that results are in line with the literature, resulting in a higher total expenditure linked to a higher share on education and a lower share on food. Additionally, by comparing the two estimation methods, we can observe that instrumenting for total expenditure has the effect of pushing the coefficients towards zero, excluding female clothing, for which the coefficient becomes significantly positive.

Table 2.6 showed that using the sex ratio as a distribution factor allows us claiming that local variation in the marriage market has a significant effect on the expenditure allocation within households. However it is not clear if the sex ratio is affecting the decision power from a wife or from a husband perception, but it only identifies a measure of their relative power. In order to identify each partner's bargaining power and in order to test for the collective rationality hypothesis, I estimate the same demand system using two measures of the re-marriage market as two distribution factors following the strategy explained in Section 2.4. Results are reported in Table 2.7 and distribution factors are standardised around the mean. Columns (1) reports the estimated coefficients for the share of men and women that get married after a divorce and for total expenditure, while column (2) reports the standard errors.

The test for unitary rationality is rejected again at 1%, meaning that the selected distribution factors jointly have a significant effect on the allocation of consumption. We can note that the effect on expenditure allocation is significant mainly for food, child clothing and utilities, with a smaller effect on education and tobacco and alcohol. A higher wedding share of husbands attributed to previous divorces leads to a smaller share spent on food and a higher share of child clothing. Moreover, we cannot reject the collective rationality hypothesis (the p-value

for the proportionality hypothesis is equal to 0.7219) providing evidence that the effect of the distribution factors is proportional across goods. However, the result of the test might be due to the fact that the estimated coefficients are not all estimated with sufficient precision. The coefficients on the share of weddings in which new wives faced a previous divorce are significant for food and child clothing, with a smaller effect on tobacco and alcohol. In addition, we can observe that the effect of each distribution factor has the negative sign of the effect of the other, showing that they could be correct measures of the bargaining power of each spouse. The estimated coefficients for total expenditure are in line with the estimated coefficients reported in Table 2.6.

2.5.1 Robustness checks

This Section discusses some validity tests to support the evidence presents previously. Firstly, in order to consistently estimate the model and to estimate the unbiased responsiveness of budget shares to total expenditure, it is required that IV robustness and exclusion restrictions are satisfied. One problem with such strategy, which is common in literature, is that the exclusion restriction assumed for total income may not be satisfied in the case labor incomes are endogenous in the household's decision to allocate consumption to different goods and there is no separability among labour and consumption decisions. In fact, if individual preferences for leisure are not separable from consumption goods, total household income (which includes labour income) doesn't satisfy the exclusion restriction. In order to support the exclusion restriction, I estimate the demand system including labour supply dummy variables for wife and husband. Specifically I control whether they worked for a wage in last 12 months. Table 2.8 presents the results when controlling for husband and wife labour supply for wage and provides evidence that controlling for labour supply has little effect on the coefficients for the sex ratio and for expenditure. We need to note however that since the sample is composed by recipients of financial assistance and the condition to be part of it is to be unemployed, the proportion of individuals who worked for wages in the previous 12 months was very low. In fact, only 15 percent of husbands and 4 percent of wives worked for wages in the previous 12 months. It would be necessary to observe labour supply in the informal sector, which is by definition, at least in part, unobservable.

Following the strategy implemented byAttanasio and Lechene (2002), consisting in instrumenting total expenditure using local variation in agricultural wage. I

proceed by instrumenting total expenditure using variables that are reasonably correlated with local variation in wages. Specifically, total expenditure is instrumented with regional unemployment and with main characteristics of migration toward and from each municipality in the period 2006-2009, excluding the year in which the expenditure data is collected (instruments included are migration change in the period 2007 and 2009[15] and female share working in the agricultural sector). First stage regressions are reported in Table 2.10. Results are robust to alternative estimation strategies.

A second order of problems arises from the absence of data on sex ratios and on divorce rates divided across different ethnic groups. In fact, due to religious differences, mixed weddings are rare and therefore a better measure of the distribution factors would be ethnic-specific sex ratios and wedding shares. However, the distribution of ethnic groups is geographically concentrated in specific municipalities. This clearly reduces the problem of absence of the ethnic-specific data since data at municipality level is expected to be highly correlated with ethnic-specific sex ratios, wedding and divorce rates. Moreover, to deal with the effect of a wrong definition of sex ratio and age classes, I have experimented with different definitions using available data. Given that the average age difference between spouses in the sample is between 4 and 5 years and may be bigger then the considered age classes, I estimated the demand system using larger age groups sex ratios by including contiguous age groups (upwards and downwards). Moreover, even if the two measures are highly correlated, I have experimented by estimating the model using sex ratios at wife's age, instead of husband's age (Figure 2.4 shows the sample distribution of these two variables). Finally, in order to solve the problem related to the geographically limited definition of sex ratio, I use regional-level sex ratios instead of municipality level. In all these cases, results are not dissimilar.

2.5.2 Endogeneity of sex ratio and divorce rates

This section is discussing whether sex ratios and divorce rates are picking up differences in the distribution of power within the household or whether it is just reflecting unobserved characteristics. This is particularly important since I find that sex ratios with a higher percentage of women affect intra-household allocation of expenditure through an increase in the expenditure share of education

[15]Migration change is defined as $migr_m = \Delta migr_{m,2009} - (\Delta migr_{m,2008} + \Delta migr_{m,2007})$.

and health and through a decrease of the share allocated to food.

I previously showed evidence that the sex ratio might be influenced by the labour market, I need to consider the possibility that it is correlated through the error term with unobservable terms that relate to labour market characteristics. Firstly, one possibility is that prices varies across regions depending on sex ratios (for instance, we may expect regions with prevalence of agriculture to have lower prices for food and lower sex ratios) and in this case our strategy would be biased by the correlation of the sex ratio with the error term. To control for this issue I compute the median price at municipality level of the most costumed item in the sample (bread) and I estimate the demand system by controlling for the variation in prices. Results are presented in Column (3) of Table 2.8 and show little effect on the coefficients of expenditure and of sex ratio. It is additionally reasonable to assume that prices are affected only by a small variation given the small dimension of the whole Macedonian market. Secondly, I control for internal migration to check whether households migrate in relationship to changes in the labour market and I find that the migration of children between 0 and 14 years old account only for the 1 percent of the total migration, meaning that households with children are not the individuals migrating. This suggests that the majority of the migration flux is characterised by single workers (or by households without children), which in turn influence the local marriage market by affecting sex ratios. While the effect of the sex ratio on the bargaining position of spouses is a possible explanation for its correlation with intra-household allocation of consumption, it is by no means the only possible explanation. Another possible explanation supported by Grossbard-Shechtman and Treble (1993) tend to favour the idea that sex ratios reflect characteristics of the labour market and especially features of labour demand. For instance, if we assume that certain regions have geographical characteristics that make them more feasible for "male" sectors, such as agriculture, we would expect these regions to attract more men and sex ratio would form endogenously with a higher proportion of men. In the opposite direction, if certain regions would favour "female" sectors, we would expect sex ratios to endogenously form with a higher proportion of women. Note that this effect goes in the same direction as the marriage market hypothesis, since regions with a higher share of men would have a higher number of hours worked by husbands and then a higher share of income.

Thirdly, I proceed to estimate the demand system by instrumenting the sex ratio using different strategies. In order to look at the precise effect of agriculture,

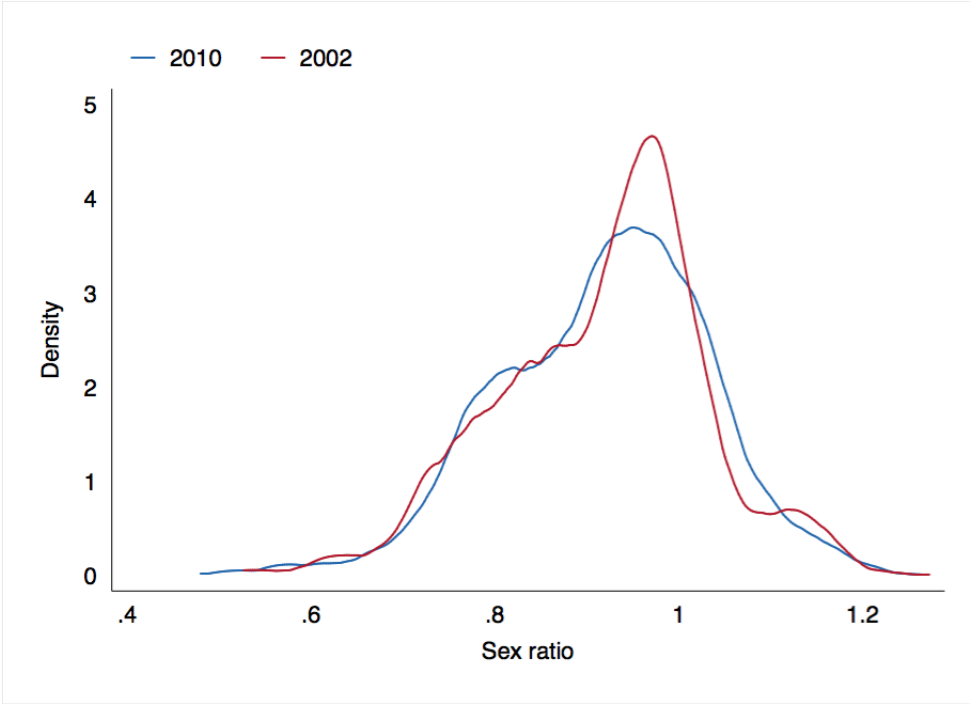
which typically capture more male workers and influence migration (which in turn influence the sex ratio), the sex ratio is instrumented with the lagged number of women working in agricultural in the municipality (normalised by the working age female population). The information is provided by the Census of Agriculture for the year 2007 (State Statistical Office, 2007). In a second specification, I instrument the actual sex ratio with the lagged sex ratio computed using the 2002 Census data; the distribution of lagged sex ratio is compared to the actual in Figure 2.4 (for completeness I report the sex ratios used, the one at husband age, and the one at wife age). The rationale to use this strategy is that Macedonia was hit by an armed conflict between the ethnic Albanian National Liberation Army (NLA) militant group and the army of the Republic of Macedonia that throughout 2001 stroked the north-western part of the country. One of the conflict consequences was a massive displacement of an estimated 74.000 people in 2001[16], which is a relevant number given an estimated population of around 2 million people. We can then claim that the sex ratio in 2002 might have been at least partly exposed to an exogenous shock caused by the war. Results are reported in Table 2.9, while the OLS first stage regressions of the sex ratio over the instruments are reported in Table 2.10. The table shows that instrumenting for sex ratio, the coefficients on sex ratio tend to become smaller and insignificant. This support the idea that sex ratios might actually be picking up as well labour market characteristics, which would drive the interpretation of our results toward the results in literature. However instrumenting for sex ratio leads to the same conclusion in terms of rejecting the unitary model, even if the significance of the test is weaker.

Similarly divorce rates and shares of new weddings where one partner has previously divorced present similar issues of endogeneity. In fact, local variation in divorce rates might be picking up the effect of (unobservable) differences in social norms, which in turn affect the relative divorce rates of men and women. We don't have a credible instrument picking up exogenous variation separately for both husband's and wive's divorce rates and lagged data on divorce rates are currently not available. Therefore estimates cannot be produced using an Instrumental Variable estimation strategy. For this case, we rely on the intra-municipality variation of ethnic composition and we need to assume that conditional on individual ethnic group, which should be capturing ethnic-related social norms, recent divorce rates are exogenous to unobservable determinants of household expenditure decisions. However, this would lead to biased estimates in case the divorce rates

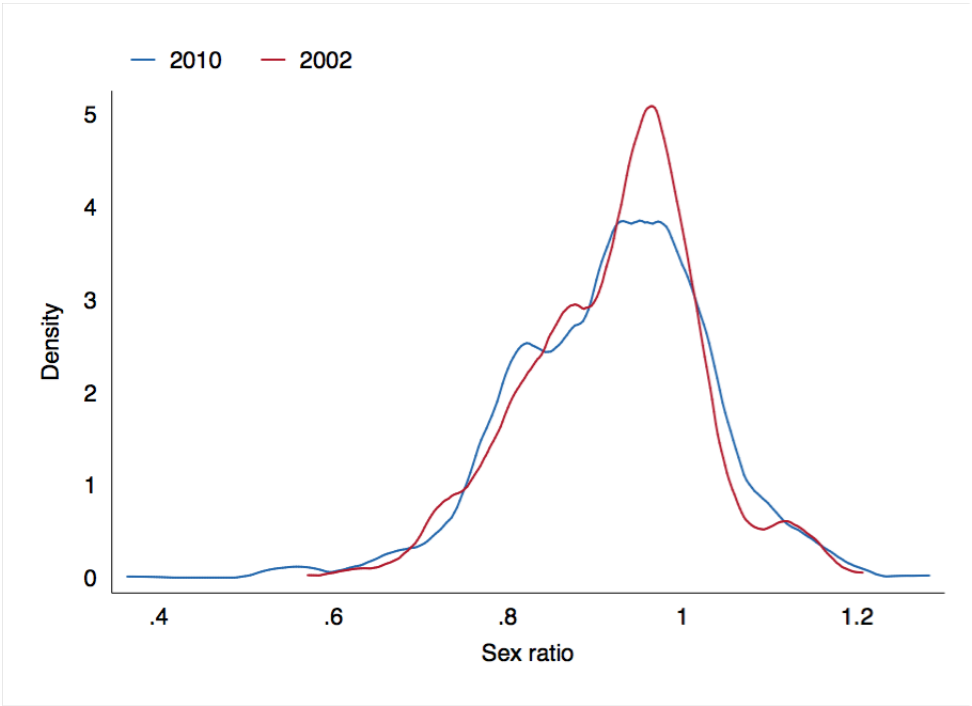
[16]Figure for the Republic of Macedonia is provided by the Internal Displacement Monitoring Centre (www.internal-displacement.org).

Figure 2.4: Distribution of sex ratio (2002 versus 2010)

5a. Husband age



5b. Wife age



Note. The graphs show density estimations of the sample distribution of the sex ratio at wife and husband age using two different sources, the 2002 census and the 2010 Macedonian State Statistical Office population statistics. Sex ratios are defined as the ratio between the number of women and the total number of individuals in the municipality of residence of the household and in the corresponding age group.

are capturing unobservable characteristics that could determine the decision to allocate resources to different goods. One possible case of endogeneity, which has already been discussed for the sex ratios, could be that divorce rates might be linked to the labour market variation, which in turn might be linked to the relative income shares of husbands and wives.

2.6 Conclusion

Understanding the "black box" called household is central in any policy initiative since it is fundamental to understand how individuals behave while being part of a group. To do so it is important to study and understand the role of intra-household distribution of decision power and how this impacts household decisions. This chapter provided evidence on the role of the marriage market in explaining how the bargaining position of each partner in expenditure choices can depend on the relative position in the wedding market. This complements evidence in the literature, that found empirical evidence on the role of the local sex ratio on labour supply and on the intra-household behaviour, and contributes with newly collected data from a developing country in which the empowerment of women is at the core of its development policy. By using a measure of the bargaining position of each partner in the re-marriage market, this chapter provided evidence in favour of the collective model. Recent studies provided mixed evidence about the efficiency assumption for household decision making, but it is important to note that the specific environment in which the household is living is central in drawing conclusions on the efficient or inefficient behaviour of households.

In this chapter I used measures of both the marriage and the re-marriage market to capture differences about these two variables. I find that sex ratios with a higher percentage of women affect intra-household allocation of expenditure through an increase in the expenditure share of education and health and through a decrease of the share allocated to food. This provides new evidence on the relationship of sex ratio with labour market characteristics since we would expect rarity of women to have a positive effect on the decision power of women within the household. Additionally, by focusing on the municipality-level share of weddings in which the spouse had previously divorced from another partner, I find evidence against the unitary rationality hypothesis and I provide evidence that is consistent with the effect of the sex ratio and in favour of the efficiency assump-

tion.

Much work has to be done in order to understand individual behaviour within households. The marriage market may not only explain the allocation of expenditures within households, but it may be related to other outcomes, such as domestic violence or educational choices (for a specific discussion on the role of sex ratios on education choices in Macedonia, refer to Section 5.7.2). To study the relationship between the marriage market and other household outputs is a subject that need to be deepened.

In conclusion, one important point to be noted is that testing for the different versions of the collective model requires an exogenous source of variation in the intra-household distribution of power. As this chapter shows, while we consider measures of income, we might lead to incorrect conclusions due to the fact that the estimates can be biased. For this reason, potential exogenous shocks have to be researched and studied, in order to improve our understanding of how families take decisions.

Table 2.1: Descriptive statistics on household structure, by type of municipality

	Rural (1)	Urban (2)	Total (3)
Household head characteristics			
Age	44.891	44.865	44.877
	(5.645)	(6.138)	(5.914)
Male	0.907	0.813	0.856
	(0.291)	(0.390)	(0.351)
Household members	4.816	4.588	4.692
	(1.133)	(1.176)	(1.161)
Number of children			
0-6 y.o.	0.117	0.168	0.145
	(0.352)	(0.454)	(0.411)
7-12 y.o.	0.475	0.418	0.444
	(0.634)	(0.618)	(0.626)
13-18 y.o.	1.645	1.519	1.577
	(0.673)	(0.641)	(0.659)
Ethnicity			
Macedonian and others	0.399	0.504	0.456
	(0.490)	(0.500)	(0.498)
Albanian	0.420	0.207	0.305
	(0.494)	(0.405)	(0.460)
Turkish	0.144	0.082	0.110
	(0.351)	(0.274)	(0.313)
Roma	0.037	0.207	0.129
	(0.189)	(0.405)	(0.335)
Area of residence			
East	0.225	0.351	0.293
	(0.418)	(0.477)	(0.455)
Center	0.296	0.371	0.336
	(0.457)	(0.483)	(0.473)
West	0.479	0.279	0.371
	(0.500)	(0.449)	(0.483)
Living in Skopje	0.000	0.247	0.134
	(0.000)	(0.432)	(0.340)

Note. Standard deviations in parenthesis. Urban and Rural areas are defined at Municipality level using the Macedonian State Statistical Office definition based on urbanisation and predominant type of economic activity. "Male" gender corresponds to code 1, while "Female" to code 0. Children are reported as total number. Skopje is the main urban agglomeration and capital city of Macedonia.

Table 2.2: Descriptive statistics on household structure, by ethnicity

	Macedonian (1)	Albanian (2)	Roma (3)	Turk (4)	Total (5)
Household head characteristics					
Age	44.962	45.949	41.918	45.025	44.877
	(5.813)	(5.525)	(6.121)	(6.046)	(5.914)
Male	0.772	0.949	0.855	0.951	0.856
	(0.420)	(0.220)	(0.353)	(0.217)	(0.351)
Household members	4.317	5.199	4.880	4.627	4.692
	(0.895)	(1.204)	(1.361)	(1.201)	(1.161)
Number of children					
0-6 y.o.	0.087	0.153	0.343	0.127	0.145
	(0.305)	(0.394)	(0.610)	(0.474)	(0.411)
7-12 y.o.	0.334	0.577	0.542	0.423	0.444
	(0.520)	(0.718)	(0.702)	(0.575)	(0.626)
13-18 y.o.	1.487	1.750	1.530	1.521	1.577
	(0.611)	(0.700)	(0.676)	(0.627)	(0.659)
Area of residence					
East	0.414	0.071	0.319	0.373	0.293
	(0.493)	(0.258)	(0.468)	(0.485)	(0.455)
Center	0.250	0.434	0.518	0.211	0.336
	(0.434)	(0.496)	(0.501)	(0.410)	(0.473)
West	0.336	0.495	0.163	0.415	0.371
	(0.473)	(0.501)	(0.370)	(0.495)	(0.483)
Living in Skopje	0.037	0.207	0.373	0.049	0.134
	(0.190)	(0.405)	(0.485)	(0.217)	(0.340)

Note. Standard deviations in parenthesis. "Male" gender corresponds to code 1, while "Female" to code 0. Children are reported as total number. Skopje is the main urban agglomeration and capital city of Macedonia.

Table 2.3: Descriptive statistics on mothers and fathers, by type of municipality

	Rural (1)	Urban (2)	Total (3)
MOTHER			
Age	41.730	42.085	41.922
	(5.750)	(6.421)	(6.122)
Education:			
No degree	0.005	0.017	0.012
	(0.071)	(0.130)	(0.107)
Lower primary	0.215	0.162	0.186
	(0.411)	(0.369)	(0.390)
Upper primary	0.535	0.461	0.495
	(0.499)	(0.499)	(0.500)
Secondary school or more	0.147	0.260	0.208
	(0.355)	(0.439)	(0.406)
FATHER			
Age	45.317	45.497	45.414
	(5.550)	(5.823)	(5.698)
Education:			
No degree	0.046	0.050	0.048
	(0.209)	(0.219)	(0.214)
Lower primary or less	0.124	0.134	0.129
	(0.329)	(0.340)	(0.335)
Upper primary	0.538	0.445	0.488
	(0.499)	(0.497)	(0.500)
Secondary school or more	0.288	0.366	0.330
	(0.453)	(0.482)	(0.470)

Note. Standard deviations in parenthesis. Urban and Rural areas are defined at Municipality level using the Macedonian State Statistical Office definition based on urbanisation and predominant type of economic activity. "Male" gender corresponds to code 1, while "Female" to code 0. Children are reported as total number. Skopje is the main urban agglomeration and capital city of Macedonia.

Table 2.4: Descriptive statistics on household total expenditure and expenditure shares

	Mean	Std.Dev.	Min	Max	% non-zeros
Monthly expenditure	22125.7	22742.9	2090.5	243742.6	.
Food	0.472	0.185	0.000	0.993	0.999
Cereals	0.152	0.105	0.000	0.787	0.967
Vegetables and Fruit	0.077	0.061	0.000	0.430	0.909
Meat	0.059	0.065	0.000	0.550	0.702
Fish	0.008	0.019	0.000	0.136	0.200
Oil and fats	0.044	0.033	0.000	0.248	0.932
Dairy	0.060	0.051	0.000	0.491	0.850
Turkish tea	0.004	0.015	0.000	0.121	0.114
Sugar, honey and salt	0.019	0.019	0.000	0.144	0.807
Chocolate and sweets	0.006	0.014	0.000	0.144	0.237
Coffe and tea	0.014	0.014	0.000	0.123	0.791
Beverages	0.015	0.024	0.000	0.221	0.473
Meals outside the household	0.003	0.014	0.000	0.263	0.050
Other food	0.012	0.018	0.000	0.174	0.479
Tobacco and alcohol	0.031	0.059	0.000	0.732	0.328
Tobacco	0.027	0.056	0.000	0.732	0.299
Alcohol	0.003	0.015	0.000	0.166	0.083
Clothing	0.047	0.043	0.000	0.256	0.814
Female	0.010	0.017	0.000	0.152	0.512
Male	0.010	0.017	0.000	0.190	0.504
Children	0.026	0.035	0.000	0.238	0.572
Education	0.130	0.223	0.000	0.934	0.852
Male	0.061	0.155	0.000	0.934	0.616
Female	0.069	0.168	0.000	0.925	0.573
Health	0.086	0.121	0.000	0.887	0.771
Wife	0.030	0.069	0.000	0.806	0.532
Husband	0.027	0.072	0.000	0.887	0.472
Children	0.023	0.063	0.000	0.662	0.358
Other	0.006	0.039	0.000	0.748	0.091
Utilities	0.141	0.086	0.000	0.837	0.991
Other	0.093	0.070	0.000	0.646	0.977
Home	0.074	0.053	0.000	0.443	0.972
Entertainment	0.009	0.025	0.000	0.294	0.240
Other expenses	0.011	0.042	0.000	0.628	0.213

Note. Monthly expenditure is expressed in Macedonian Denars (1 MKD = 0.015 GBP). Budget shares are computed as expenditure on the category divided by total household expenditure over the same time period. Budget shares for sub-categories are expressed as percentage of total expenditure rather than percentage of macro-category expenditure.

Table 2.5: Descriptive statistics on household total expenditure, consumption and shares

	Urban		*Rural*	
	Expenditure	Consumption	Expenditure	Consumption
Monthly value	20860.4	21212.6	23603.1	23528.9
	(20095.4)	(19924.9)	(25426.6)	(25555.5)
Food	0.476	0.478	0.468	0.460
	(0.177)	(0.169)	(0.193)	(0.184)
Tobacco and alcohol	0.038	0.037	0.022	0.021
	(0.067)	(0.060)	(0.047)	(0.046)
Clothing	0.045	0.049	0.049	0.054
	(0.043)	(0.045)	(0.044)	(0.046)
Education	0.118	0.115	0.144	0.144
	(0.207)	(0.202)	(0.240)	(0.240)
Health	0.082	0.081	0.090	0.089
	(0.111)	(0.108)	(0.132)	(0.129)
Utilities	0.148	0.145	0.134	0.135
	(0.091)	(0.090)	(0.080)	(0.082)
Other expenses	0.093	0.095	0.094	0.097
	(0.070)	(0.070)	(0.071)	(0.069)

Note. Standard deviations in parenthesis. Monthly value of expenditure and consumption is expressed in Macedonian Denars (1 MKD = 0.015 GBP). Shares are computed as expenditure (consumption) on the category divided by total household expenditure (consumption) over the same time period.

Table 2.6: Test of Individual Rationality using sex ratio

	SUR (1)		*FIML* (2)	
	Coeff.	Std.error	Coeff.	Std.error
Food				
Sex ratio	-0.199**	0.090	-0.189**	0.082
Expenditure	-0.167***	0.013	-0.136*	0.073
Clothing (female)				
Sex ratio	-0.005	0.007	-0.002	0.007
Expenditure	-0.001	0.001	0.012***	0.004
Clothing (male)				
Sex ratio	-0.020**	0.008	-0.019**	0.008
Expenditure	-0.001	0.001	0.002	0.003
Clothing (children)				
Sex ratio	-0.032	0.037	-0.030	0.038
Expenditure	-0.009***	0.003	-0.000	0.015
Education				
Sex ratio	0.279***	0.081	0.231***	0.082
Expenditure	0.257***	0.013	0.098**	0.049
Health				
Sex ratio	0.066	0.084	0.043	0.084
Expenditure	0.006	0.009	-0.070*	0.039
Tobacco and alcohol				
Sex ratio	-0.001	0.026	0.001	0.025
Expenditure	0.006	0.004	0.011	0.016
Utilities				
Sex ratio	-0.025	0.049	0.010	0.058
Expenditure	-0.075***	0.006	0.042	0.041
Observations	1268		1268	
Unitary rationality test				
$\chi^2(7)$	25.89		20.85	
Prob $> \chi^2$	<0.001		0.004	

Note. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are clustered at municipality level. Model (1) is estimated by SUR, while model (2) is estimated by instrumenting total expenditure with total income, square of total income, unemployment and migration. The dependent variables are defined as budget share, e.g. amount spent on one category divided by total expenditure. Expenditure is reported in logarithm. The sex ratio is defined as the ratio of female to male population at husband's age in the municipality of residence.

Table 2.7: Test of Individual and Collective Rationality using share of weddings attributed to previous divorces

	SUR (1)		*FIML* (2)	
	Coeff.	Std.error	Coeff.	Std.error
Food				
Wedding share (wife)	0.026*	0.015	0.025*	0.015
Wedding share (husband)	-0.058***	0.017	-0.054***	0.016
Expenditure	-0.172***	0.012	-0.120*	0.066
Clothing (female)				
Wedding share (wife)	0.002	0.001	0.002	0.001
Wedding share (husband)	-0.002	0.001	-0.001	0.001
Expenditure	-0.001	0.001	0.011***	0.004
Clothing (male)				
Wedding share (wife)	0.000	0.001	0.000	0.001
Wedding share (husband)	-0.001	0.001	-0.001	0.001
Expenditure	-0.001	0.001	0.002	0.003
Clothing (children)				
Wedding share (wife)	-0.016***	0.005	-0.016***	0.005
Wedding share (husband)	0.014**	0.006	0.015***	0.006
Expenditure	-0.009***	0.002	0.002	0.011
Education				
Wedding share (wife)	-0.004	0.013	-0.001	0.016
Wedding share (husband)	0.025*	0.014	0.026*	0.015
Expenditure	0.259***	0.014	0.090*	0.047
Health				
Wedding share (wife)	-0.004	0.009	-0.003	0.009
Wedding share (husband)	0.003	0.010	-0.002	0.010
Expenditure	0.005	0.009	-0.064*	0.038
Tobacco and alcohol				
Wedding share (wife)	-0.005*	0.003	-0.006*	0.003
Wedding share (husband)	0.002	0.004	0.002	0.004
Expenditure	0.006	0.005	0.013	0.014
Utilities				
Wedding share (wife)	0.003	0.008	0.001	0.008
Wedding share (husband)	0.018***	0.006	0.026***	0.007
Expenditure	-0.072***	0.006	0.026	0.027
Observations	1268		1268	
Unitary rationality test				
$\chi^2(16)$ (Prob $> \chi^2$)	84.88 (<0.001)		69.58 (<0.001)	
Collective rationality test				
$\chi^2(7)$ (Prob $> \chi^2$)	5.85 (0.5568)		4.49 (0.7219)	

Note. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are clustered at municipality level. Model (1) is estimated by SUR, while model (2) is estimated by instrumenting total expenditure with total income, square of total income, unemployment and migration. The dependent variables are defined as budget share, e.g. amount spent on one category divided by total expenditure. Expenditure is reported in logarithm. Wedding shares for husband and wife are defined as the share of new weddings in which the male (or female) partner has been previously divorced and is defined for the period 2007-2009.

Table 2.8: Total Expenditure, Labour supply and Prices

	FIML (1)		*FIML* (2)		*FIML* (3)	
	Coeff.	Std.error	Coeff.	Std.error	Coeff.	Std.error
Food						
Sex ratio	-0.189**	0.082	-0.185**	0.083	-0.205**	0.083
Expenditure	-0.136*	0.073	-0.135*	0.075	-0.145**	0.071
Clothing (female)						
Sex ratio	-0.002	0.007	-0.001	0.007	-0.002	0.007
Expenditure	0.012***	0.004	0.012***	0.004	0.011***	0.004
Clothing (male)						
Sex ratio	-0.019**	0.008	-0.018**	0.008	-0.019**	0.008
Expenditure	0.002	0.003	0.002	0.003	0.002	0.003
Clothing (children)						
Sex ratio	-0.030	0.038	-0.027	0.038	-0.028	0.039
Expenditure	-0.000	0.015	-0.001	0.015	-0.002	0.014
Education						
Sex ratio	0.231***	0.082	0.245***	0.082	0.275***	0.078
Expenditure	0.098**	0.049	0.098**	0.047	0.113**	0.044
Health						
Sex ratio	0.043	0.084	0.023	0.081	0.027	0.084
Expenditure	-0.070*	0.039	-0.071*	0.040	-0.070*	0.037
Tobacco and alcohol						
Sex ratio	0.001	0.025	-0.003	0.025	0.001	0.025
Expenditure	0.011	0.016	0.010	0.015	0.013	0.015
Utilities						
Sex ratio	0.010	0.058	0.012	0.059	0.002	0.059
Expenditure	0.042	0.041	0.045	0.041	0.040	0.039
Observations	1268		1268		1268	
Labour supply			✓		✓	
Price of bread					✓	
Unitary rationality test						
$\chi^2(7)$	20.85		20.50		24.95	
Prob $> \chi^2$	0.004		0.005		<0.001	

Note. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are clustered at municipality level. Model (1)-(3) are estimated by FIML by instrumenting total expenditure with total income, square of total income, unemployment and migration. In Model (2), dummy variables indicating whether the husband and wife have worked for wage in the last 12 months. Model (3) adds a control for median price of bread at municipality level. Expenditure is reported in logarithms. The sex ratio is defined as the ratio of female to male population at husband's age in the municipality of residence.

Table 2.9: Test of Unitary Rationality and Endogenous Sex Ratio

	INSTRUMENT					
	No instrument		*Female share of agricultural population*		Lagged sex ratio	
	(1)		(2)		(3)	
	Coeff.	Std.error	Coeff.	Std.error	Coeff.	Std.error
Food						
Sex ratio	-0.189**	0.082	-0.366**	0.163	-0.210*	0.127
Clothing (female)						
Sex ratio	-0.002	0.007	-0.007	0.014	-0.004	0.008
Clothing (male)						
Sex ratio	-0.019**	0.008	-0.028**	0.012	-0.020**	0.008
Clothing (children)						
Sex ratio	-0.030	0.038	-0.002	0.059	-0.024	0.048
Education						
Sex ratio	0.231***	0.082	0.196	0.130	0.158	0.127
Health						
Sex ratio	0.043	0.084	-0.031	0.128	0.075	0.072
Tobacco and alcohol						
Sex ratio	0.001	0.025	-0.010	0.046	0.013	0.038
Utilities						
Sex ratio	0.010	0.058	0.280**	0.109	0.056	0.065
Observations	1268		1268		1268	
Unitary rationality test						
$\chi^2(7)$	20.58		20.02		13.95	
Prob $> \chi^2$	0.004		0.006		0.052	

Note. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are clustered at municipality level. Model (1) and (2) are estimated by FIML by instrumenting total expenditure with total income, square of total income, unemployment and migration. In Model (1) sex ratio is instrumented by the lagged share of female workers in the agricultural sector (compared to men). In Model (2) sex ratio is instrumented by the lagged sex ratio, computed using the 2002 census.

51

Table 2.10: First stage regression for total expenditure and sex ratio

	Expenditure (1)	Expenditure (2)	Expenditure (3)	Expenditure (4)	Sex ratio (5)	Sex ratio (6)	Sex ratio (7)	Sex ratio (8)
Total Income	0.255***	0.115*						
	(0.07)	(0.06)						
Total Income (squared)	-0.000***	-0.000*						
	(0.00)	(0.00)						
Asset index	0.192***	0.167***						
	(0.03)	(0.03)						
Unemployment rate (2010)	-0.004**	-0.006***						
	(0.00)	(0.00)						
Migration change			0.001***	0.001***				
			(0.00)	(0.00)				
Female share working in agriculture			1.723***	2.548***				
			(0.46)	(0.58)				
Female workers in agriculture					-0.317***	-0.393***		
					(0.03)	(0.03)		
Sex ratio (lagged)							0.830***	0.778***
							(0.03)	(0.03)
Controls		✓		✓		✓		✓
Observations	1268	1268	1268	1268	1268	1268	1268	1268
R^2	0.189	0.322	0.059	0.290	0.420	0.538	0.642	0.669
F(k,n-k-1)	28.07	20.90	20.04	13.28	91.06	217.90	581.32	612.75

Note. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Robust standard errors are clustered at municipality level. All specification are estimated using OLS. F statistic is computed on a joint-significance test of the instruments. The dependent variables are total expenditure for models (1)-(4) and sex ratio for models (5)-(8). Expenditure is reported in logarithm. Female share working in agriculture is defined as number of female workers in agriculture divided by total agricultural workers in the municipality. Female workers in agriculture is defined at municipality level and normalised by the female working-age population.

52

Table 2.11: Descriptive statistics on household total expenditure and expenditure shares

	Mean	Std.Dev.	Min	Max
Age of head	44.877	5.914	18.765	65.057
Male head	0.856	0.351	0.000	1.000
Household members	4.692	1.161	3.000	11.000
Members 0-6 y.o.	0.145	0.411	0.000	3.000
Members 7-12 y.o.	0.444	0.626	0.000	3.000
Members 13-18 y.o.	1.577	0.659	0.000	4.000
Members 19-29 y.o.	0.478	0.737	0.000	5.000
Members 30-39 y.o.	0.558	0.712	0.000	2.000
Members 40-49 y.o.	1.140	0.754	0.000	3.000
Members 50-59 y.o.	0.284	0.578	0.000	2.000
Members >= 60y.o.	0.068	0.289	0.000	2.000
Macedonian and others	0.456	0.498	0.000	1.000
Albanian	0.305	0.460	0.000	1.000
Turkish	0.110	0.313	0.000	1.000
Roma	0.129	0.335	0.000	1.000
East	0.293	0.455	0.000	1.000
Center	0.336	0.473	0.000	1.000
West	0.371	0.483	0.000	1.000
Living in Skopje	0.134	0.340	0.000	1.000
Region: Vardar	0.111	0.314	0.000	1.000
Region: Eastern	0.120	0.325	0.000	1.000
Region: Southwestern	0.131	0.338	0.000	1.000
Region: Southeastern	0.096	0.295	0.000	1.000
Region: Pelagonia	0.109	0.311	0.000	1.000
Region: Polog	0.131	0.337	0.000	1.000
Region: Northeastern	0.077	0.267	0.000	1.000
Region: Skopje	0.225	0.418	0.000	1.000
Worked (husband)	0.153	0.360	0.000	1.000
Worked (wife)	0.040	0.195	0.000	1.000
Asset index	-0.000	1.000	-3.816	2.799
Husband's wage share	0.619	0.237	0.000	1.000
Sex ratio	0.918	0.114	0.480	1.263
Sex ratio (lagged, 2002)	0.917	0.113	0.527	1.272
Total Income	0.512	0.778	-3.656	15.260
Total Income (squared)	8674.407	81557.685	0.000	2328676.0
Asset index	-0.000	1.000	-3.816	2.799
Unemployment rate	31.646	12.107	11.500	62.800
Migration change	7.420	59.404	-96.000	500.000
Female share working in agriculture	0.424	0.044	0.222	0.491
Female workers in agriculture	0.340	0.237	0.000	1.295

Note. Standard deviations in parenthesis. Urban and Rural areas are defined at Municipality level using the Macedonian State Statistical Office definition based on urbanisation and predominant type of economic activity. "Male" gender corresponds to code 1, while "Female" to code 0. Children are reported as total number. Skopje is the main urban agglomeration and capital city of Macedonia.

Chapter 3

Validation of Subjective expectations

3.1 Introduction

The availability of credible measures for factors entering the decision process (i.e. expectations) is fundamental to estimate models of individual behaviour under weaker assumptions (Manski, 2004). Collecting accurate and credible data about subjective expectations turns to be important in data collection activities that require the gathering of this kind of information. While the collection of data about subjective expectations has been increasingly important in developed countries, the topic gathered attention only recently in developing countries. This is mainly due to the fact that questions related to probabilities might not be fully understood or might require a higher burdensome when the average level of education of respondents is low. In the other direction, data collected in developing countries are central since they might partly relate to the presence of market failures related to information and knowledge.

There is still no consensus in literature about the best practice to collect subjective expectations in developing countries and there is need for improvements and understanding on how to elicit subjective expectation. However, a growing and recent research agenda (an introduction about the strategies and the experiences about collecting expectations in developing countries is provided by Attanasio, 2009 and Delavande et al., 2010) provides important insights on how to design and implement a strategy to collect information about subjective probabilities.

In literature, data about expectation have been mainly collected by using non-probabilistic methods (such as Likert scales) and more recently by using probabilistic methods with or without visual aids. Even if Likert scales have proven to be partially related to subjective probabilities (Delavande and Kohler, 2009), non-probabilistic method might be problematic for inter-personal comparisons since we can't ascertain which is the quantity reported by the respondent, while by eliciting subjective probabilities we might be able to recover the moments of a distribution of interest. However, eliciting subjective probabilities might be problematic in developing countries, since the concept of probability might not be understood by respondents with a lower levels of education. For this reason, asking directly about probabilities (for instance, in the case of a developed country like Dominitz and Manski (1997), who ask about the probability of being employed in a future date in the US, and for a developing country like McKenzie et al. (2007), who ask about income expectations conditional on migration versus non-migration among Tongan migrants in New Zealand and Tongan residents) might not return credible results.

This chapter analyses the validity of questions related to subjective expectations using the data collected among social financial recipients in the Republic of Macedonia during the 2010 and 2012 data collection waves of the Macedonian "Secondary School Conditional Cash Transfer" evaluation household survey. Data on subjective expectations have been collected using the method proposed by Guiso et al. (2002). Under distributional assumptions, this method allows eliciting the subjective expected earning distribution for each respondent. Given the low level of schooling of most respondents in the sample, it was important to select a methodology that would have allowed eliciting a credible measure of subjective expectations, while avoiding asking directly about probabilities. Following Attanasio et al. (2005), who collected income expectations in Colombia, and Attanasio and Kaufmann (2009), who elicited income expectations of junior high school students in Mexico, in the Republic of Macedonia, data about expectations on education returns from primary school and from secondary school have been collected using a method that elicit subjective expectations without directly using probabilistic terminology. Under both assumptions of having completed either primary or secondary school, it was asked about the minimum and the maximum of what the child could have earned, and after computing a mid-point, it was asked to report on a scale from 0 to 100 "how likely" the child would have earned less (or more) than the mid-point. The questionnaire asks to the household head (or its spouse in case of absence during the interview) informa-

tion over the expected salary conditional on completion of primary or secondary school for at least one adolescent child in the household (in the case that are present two adolescents of different gender the information is collected for both). This chapter analyses the validity of the data by testing whether respondents understand the question and use sufficient mental effort to report their answers. Using randomisation in the order of questions, I find that when asked about the probability to earn less than a certain threshold and more than the same threshold, respondents tend to reply with sufficient mental effort only to the first question they are asked.

The interesting feature of this data is that it has been collected in a developing country, where evidence is still growing, and among its poorest quintiles (social financial assistance recipients), where respondents have a low level of schooling and have a lower experience in the formal labour market. In order to elicit subjective probabilities, it was used a visual aid strategy composed by a 0-100 ruler, that was initially presented with an example linking the probability of rain with the chosen scale (examples of more complex visual aids, such as the use of stones and coffee beans are provided by Luseno John and Winnie, 2003; Lybbert et al., 2007; Hill, 2006). The decision was linked to the fact that in the sample a large part of households live in rural areas and that rain is often a problem related to availability of utilities in urban areas. Section 3.2 describes the data and the methodology chosen to elicit subjective expectations. Section 3.3 provides a series of tests to validate the reported expectations.

3.2 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[1]. 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[2]. 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

[1]We aimed for a sample size of 17 households eligible for the CCT (recipients of social and financial assistance with children of secondary school age) per municipality, although in practice there was some variation in this number due to the fact that in some municipalities the eligible population was smaller then 17. For power calculations, we considered a power of 0.8 and a significance level of 0.05. With 42 clusters per arm and an inter-cluster correlation of 0.25, using 17 households per municipality it would be possible to detect a difference in expenditures in children's education (or in any other item) of 0.33 of a standard deviation and an increase in the proportion of students attending 85% or more of the classes of roughly 10% points.

[2]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.

were not found or refused to answer at follow-up, resulting in an attrition rate of 11.7%. 5.8 presents some robustness checks related to attrition and provides evidence that attrition at follow-up didn't changed significantly the composition of the sample for each treatment modality.

3.2.1 Subjective expectation module

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; 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[3]. 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[4], we can construct the distribution of the expected income and calculate its first moments[5] (Guiso et al., 2002). Specifically, assuming 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:

$$E[Y|E_i = 1] = \int_{y^L}^{y^U} y\, f_{Y|E}(y|E_i = 1)\, \mathrm{d}y \equiv \bar{y}_E \tag{3.1}$$

$$Var[Y|E_i = 1] = \int_{y^L}^{y^U} (y - \bar{y}_E)^2\, f_{Y|E}(y|E_i = 1)\, \mathrm{d}y \tag{3.2}$$

Table 3.1 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).

[3]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?"

[4]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.

[5]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.

Table 3.1: 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.

3.3 Testing the validity of subjective expectations

In order to understand whether collected answers are valid, it is important to answer the following question: does individuals understand probability questions? Are answers related to observable variables? How accurate are the answers? In order to answer the first question, during the 2010 and 2012 data collection, we followed a strategy similar to Attanasio et al. (2005). In each municipality, households were randomly allocated into two groups: one group faced first the question about how likely is to earn less (or equal) than the calculated mid-point (we defined this group "X") and the other who faces first the question about how likely is to earn more than the calculated mid-point (group "Y"). We can then test whether the sum of the sample means for each group sums up to one. Tables 3.2 and 3.3 compare the means of the reported probability in the two groups and the sum of the means for the first answer and for the second answer reported by the respondent. As we can note the first answer perform much better than the second answer. When considering the first answer, for both boys and girls and for primary and secondary school expectations, we cannot reject the null hypothesis of the sum being different than 1 for most of the cases. The test performs much better for expectations collected at baseline. However, if we consider the second question answered, we reject the null hypothesis in all cases and the sum is always significantly lower than 1. To support these results, I present a joint test checking that for boys and girls, for the two rounds of data collection and for first and second reported answer, the sum of the means of group "X" and "Y" is equal to 1. In other words, I run Romano and Wolf (2005) procedure to jointly test that each of the sum of the means presented in Table 3.2 and Table 3.3 is equal

to one, summing up to 48 hypothesis[6]. The individual and multiple hypothesis tests show that while the first reported answer conform to probability theory, the second answer tends to be significantly lower than one. A possible explanation is that, during the second question, respondents tend to reduce their mental effort and, for this reason, the answers tend not to conform to probability theory.

Unlike many applications, the questionnaire ask both questions to each respondent and allows then to test whether individual answers conform to probability theory by summing up to one. In other words, given the event A "earning an income between the minimum and the mid-point (included)", its complement $\bar{A}$ "earning an income between the mid-point and the maximum" and the event B "being employed", it is important to test whether the following condition is respected:

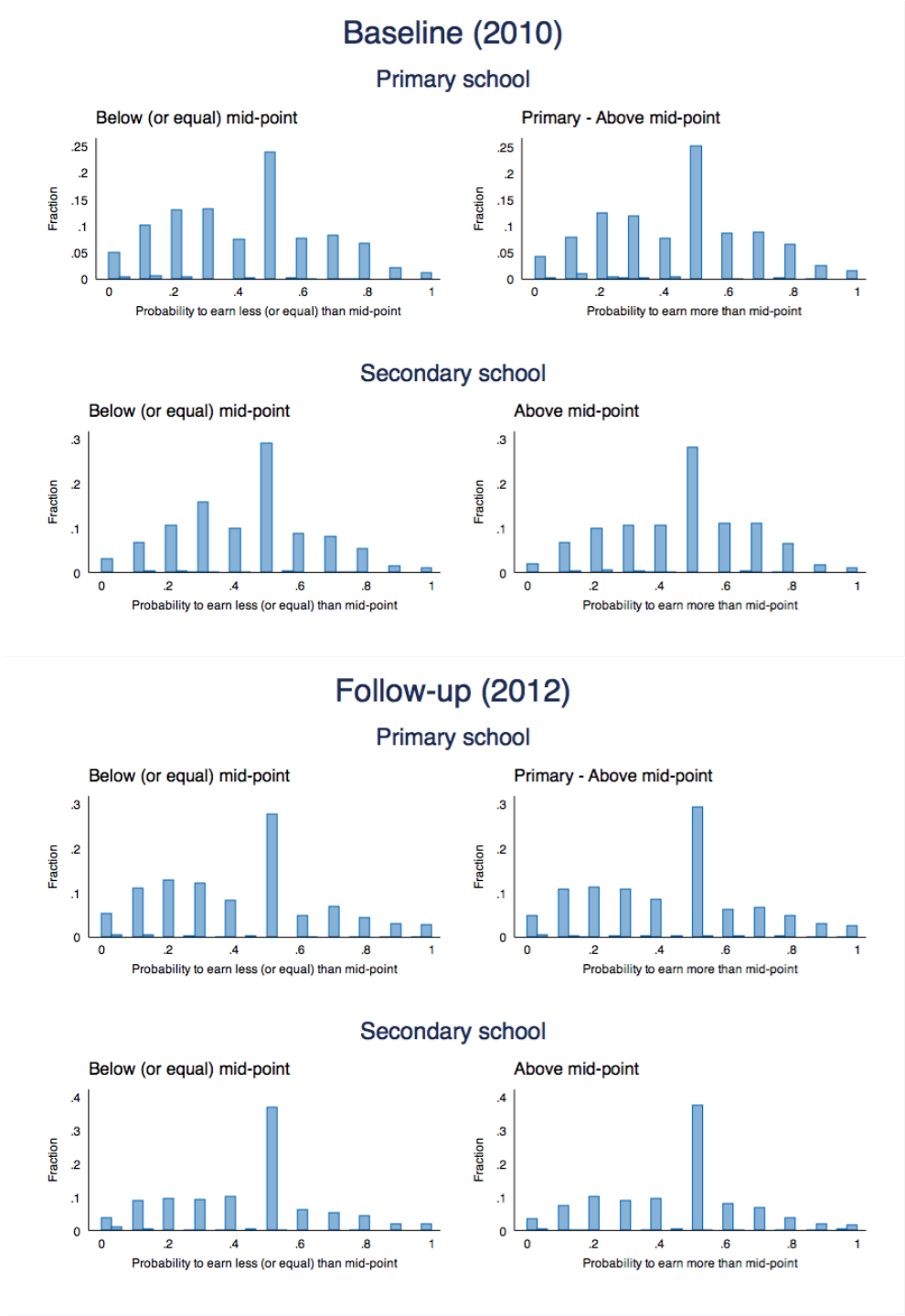
$$P(A \cup \bar{A}|B) = P(A|B) + P(\bar{A}|B) = P(A|B) + [1 - P(A|B)] = 1$$

Figure **3.1** shows the distribution of answers for event A (left panel) and for its complement $\bar{A}$ (right panels). In order to test whether answers conform to probability law, I first computed the sum of reported probabilities ($P(A|B)+P(\bar{A}|B)$) and then tested whether they differ from one[7]. Table 3.4 presents the distribution of the sum of reported probabilities for both primary school and secondary school expectations. We can observe that the cases in which probabilities sum up to one are ranging across the two data collection waves from a minimum of 44 percent to a maximum of 59 percent, depending on the precise category for which expectations are collected (achieved educational level and gender of the child). There is a strong tendency from respondents to report sums that are smaller than one, with roughly only 10 percent of observations being larger than 1.

[6]The test is implemented using Romano and Wolf (2005) procedure with 5000 bootstrap repetitions. The critical values generated through the procedure are the 90th, 95th and 99th percentile of the empirical distribution of $\tilde{t}$, where $\tilde{t}$ is the maximum t-statistic in each iteration testing, for each sum of the means, whether the coefficient estimated through bootstrap is equal to the coefficient estimated in the main sample. Table 3.2 and Table 3.3 presents whether each hypothesis is rejected and at which percentile.

[7]Mahajan et al. (2008) run a similar test by collecting subjective expectations about income in India using 10 stones for the questions about likelihood to earn less than the mid-point and other 10 stones to indicate the likelihood to earn more than the mid-point; they find that only 513 out of 1945 individuals answers with probabilities summing up to one, but the mean sum of probabilities is equal to 1.13 (including all observations) or 1.06 (excluding answers equal to 0 and 1), that is encouraging given the fact that answers were limited to multiples of 0.1.

Figure 3.1: Distribution of expected income after under different distributional assumptions



Note. Panels titled "Primary school" presents probabilities related to the completion of primary school only, while panels titled "Secondary school" presents probabilities related to the completion of secondary school.

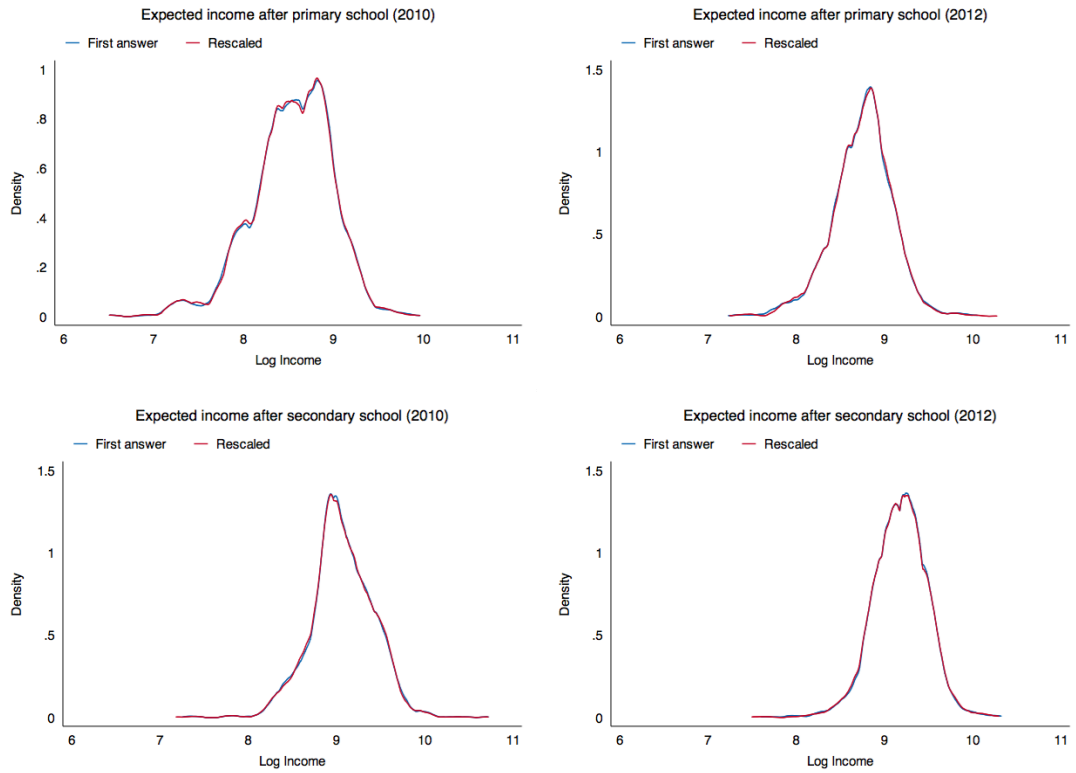
In order to understand how strong is the tendency of report sums smaller or larger than one, I look at the mean sum of probabilities for the observations that don't conform to probability law and compare them to the mean for all observations (Table 3.5). We can observe that the sum ranges from a minimum of 0.61 (0.56) to a maximum of 0.70 (0.60) for the baseline (follow-up) in the cases in which the sum is smaller than one. If we consider instead the few cases in which the sum is larger than one, means ranges from 1.16 (1.17) to 1.21 (1.22) for the baseline (follow-up). We find that in all cases, even if roughly half of the respondents provides answers that sum up to one, the mean probability is always significantly different than one, suggesting that it might be reasonable to use different specifications to construct the probability distribution, such as using rescaling of probabilities or using only the first answer. In Section "First answer versus rescaling" I discuss how these two methods interact with distributional assumptions, which are first introduced in the following section.

As showed in Section 3.3, when the respondent is asked about the probability to earn below (or equal) and above a certain threshold, there is a possibility that the respondent reduce its mental effort during the second answer, such that the reported probabilities do not sum up to one. It is therefore important to understand whether using only the first answer or use both answers would lead to fundamental differences to the subjective income distribution.

In order to test for differences, I use compare the effect of choosing rescaling versus using the first answer only by looking at the distribution of expected income and its variance. Assuming that each respondent answers both questions about earning below (p_A) and above (p_B) the mid-point, the first strategy is to consider only the first question answered by the respondent ($\bar{p} = p_A$) and then compute the complement ($1 - \bar{p}$) as $1 - p_A$. The second strategy is to use both answers and rescale them in case the sum $p_A + p_B$ is different than one. In this case, $\bar{p}$ is simply determined by $p_A/(p_A+p_B)$.

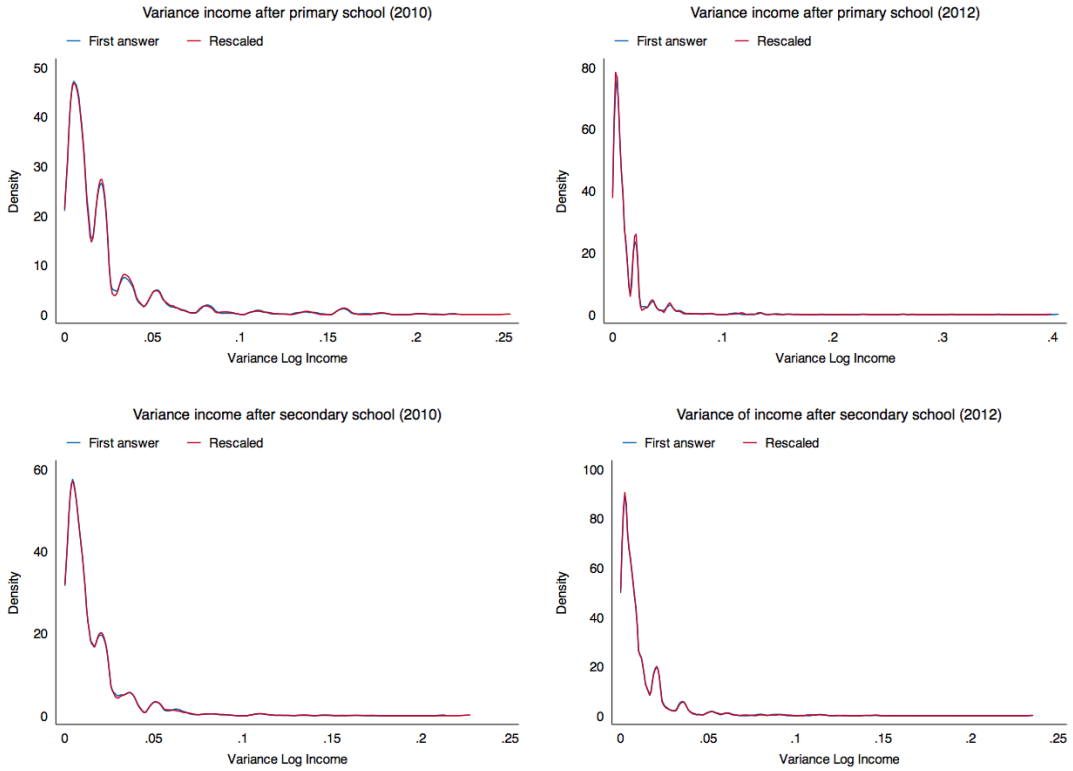
Table 3.6 reports expected income for boys and girls at baseline and follow-up computed under different distributional assumption and using either a rescaled probability or a first answer probability. If we look at the mean difference between these two methods, we can note that differences are never significantly different than zero at baseline and rarely significant at follow-up. This provides evidence that using different methods do not lead to significant differences in computed

Figure 3.2: Distribution of expected income comparing first answer and rescaling



Note. Expected income is computed using log-income. "Rescaled" expectations are computed by rescaling the sum of reported probabilities to be equal to one, while "first answer" expectations are computed using the first reported answer only. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). The two top panels show the expected income after completion of primary school only at baseline and at follow-up, while the lower panels show the expected income after completion of secondary school at baseline and follow-up.

Figure 3.3: Distribution of variance of income comparing first answer and rescaling



Note. Variance of income is computed using log-income. "Rescaled" expectations are computed by rescaling the sum of reported probabilities to be equal to one, while "first answer" expectations are computed using the first reported answer only. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). The two top panels show the variance of income after completion of primary school only at baseline and at follow-up, while the lower panels show the variance of income after completion of secondary school at baseline and follow-up.

expected incomes, even when considering different distributional assumptions. Additionally, we can compare the distributions of expected income and its variance for each different method. Figure 3.2 and Figure 3.3 show that there are no significant differences even when we consider the whole sample distribution.

3.4 Testing distributional assumptions

Given the structure of the collected information, assuming a specific class of distribution functions allows constructing the distribution of the expected salary and calculating its first moments. Specifically, assuming that a is the reported salary in the worst case, b is the reported salary in the best case and $f_Y(y)$ is the assumed continuous density function of the expected salary for one respondent, we can calculate the expected value and the variance using standard statistical formulas:

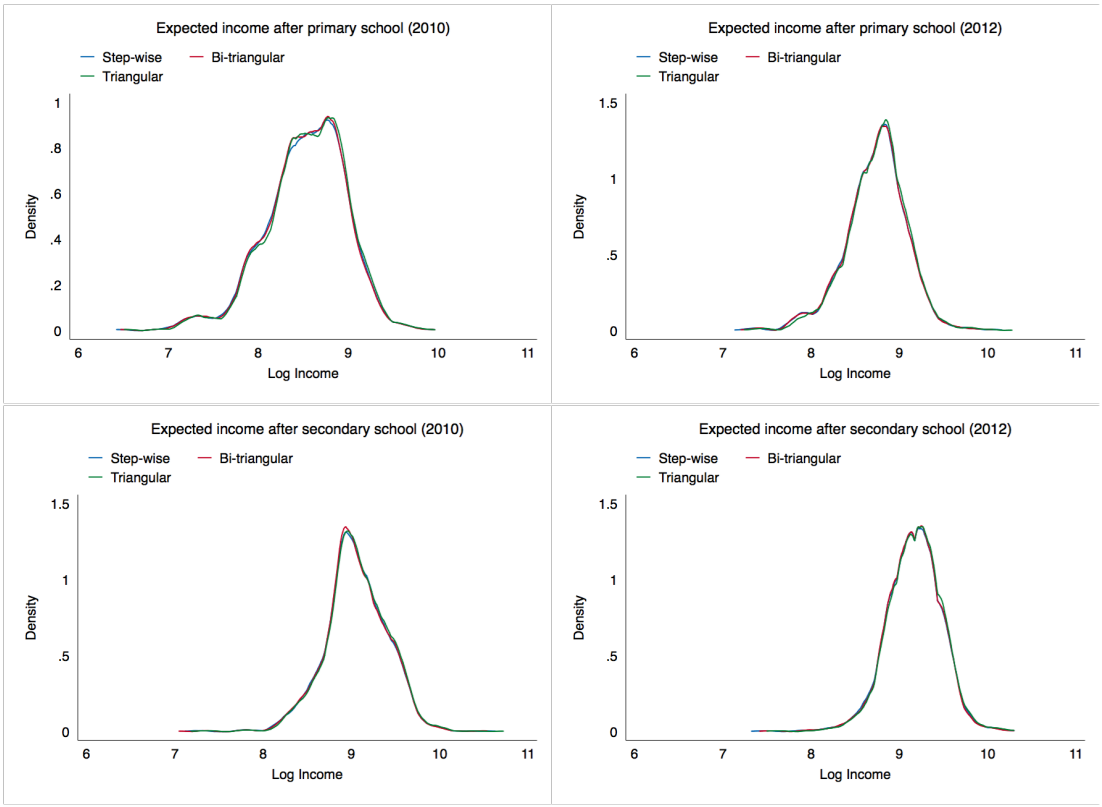
$$E[Y] = \int_a^b y\, f_Y(y)\, \mathrm{d}y = \mu \tag{3.3}$$

$$Var[Y] = \int_a^b (y-\mu)^2\, f_Y(y)\, \mathrm{d}y \tag{3.4}$$

In order to reconstruct the probability density function, it is necessary to consider distribution that can be identified using the a, b and the reported mass probability between a and the midpoint $(a+b)/2$. Distribution functions that are consistent with this setting are the step-wise uniform distribution, the triangular distribution and the bi-triangular distribution. Figure 3.4 reports the sample distribution of the expected income conditional on completing primary and secondary school at baseline and follow-up using different distributional assumptions. We can observe that there is no strong difference between distributional assumptions for what concerns the distribution of the expected income, both conditional on completing primary and secondary school.

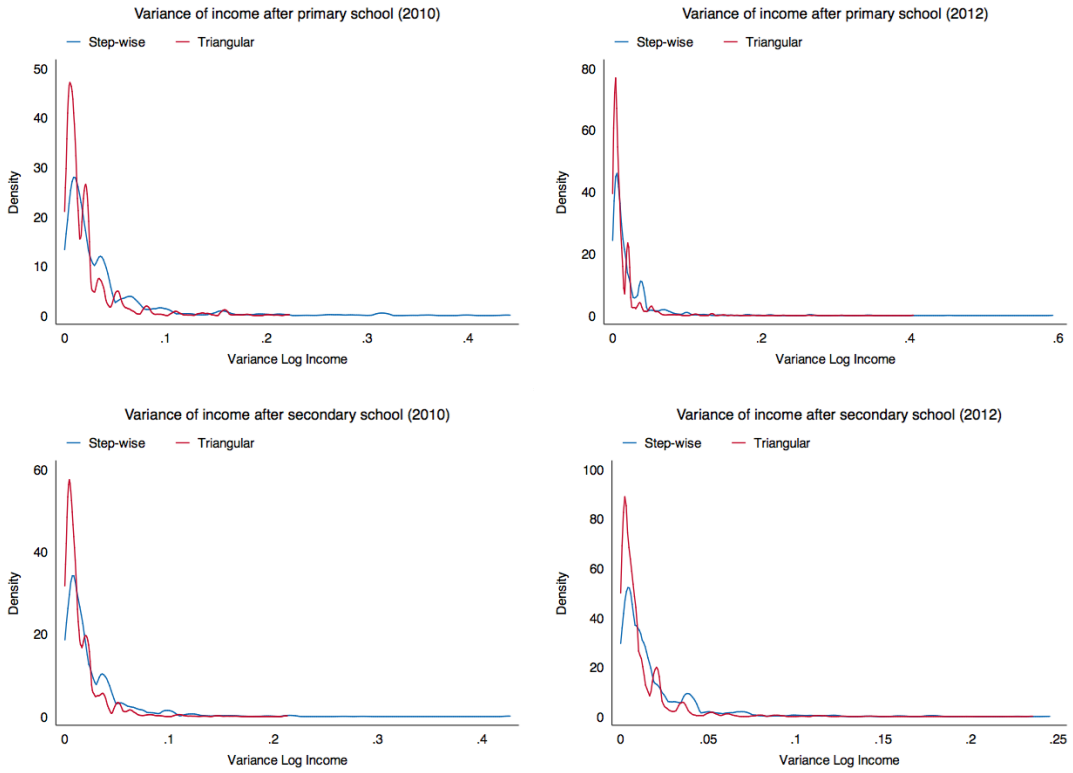
If we focus on the second moment of the distributions, we can note that differences become more significant. Figure 3.4 reports the sample distribution of the variance of income conditional on completing primary and secondary school at baseline and follow-up using different distributional assumptions (for clarity of the graph, I present only the distribution using only the step-wise uniform and the triangular distribution). An important characteristic to be researched among distribution functions that can be assumed for the purpose of constructing the

Figure 3.4: Distribution of expected income after under different distributional assumptions



Note. Expected income is computed using log-income and using different distributional assumptions (step-wise, bi-triangular and triangular). The two top panels show the expected income after completion of primary school only at baseline and at follow-up, while the lower panels show the expected income after completion of secondary school at baseline and follow-up.

Figure 3.5: Distribution of variance of income under different distributional assumptions



Note. Variance of income is computed using log-income and using different distributional assumptions (step-wise and triangular). The two top panels show the variance of income after completion of primary school only at baseline and at follow-up, while the lower panels show the variance of income after completion of secondary school at baseline and follow-up.

distribution of subjective income expectations is that the density is decreasing while moving towards the extremes. This is the case that led the triangular distribution to be used in all the application of the chapter. For this reason, we clearly expect the distribution of variances to differ between the two distributional assumptions.

3.5 Perceived returns and individual characteristics

In order to verify the validity of reported expectations, it is important to control the relationship between the answers about minimum and maximum reported income with observable characteristics of the household and the respondent. Table 3.7 and Table 3.8 present linear regressions of the minimum and maximum (log) income reported by the respondent upon primary and secondary school completion and its difference (defined here as delta) on a series of individual and household characteristics. Expected income for male children are significantly higher in both rounds of data collection and for both primary and secondary school outcomes, but there is no significant difference across female and male children when considering the difference. At the same time, education of the household head has a role in explaining reported income, but with a different pattern at baseline and follow-up. While at baseline higher education lead to a lower expected income after primary school, at follow-up lower education is correlated with lower incomes while higher education of the household head is correlated with higher incomes. To provide evidence that reported expectations are linked to monetary returns in the labour market, I control for unemployment rate at regional level, by dividing municipalities into (relatively) low and high unemployment. We can note that larger unemployment in the year before the interview affect negatively reported incomes for both educational levels, providing evidence that in areas with high unemployment respondents expect lower incomes.

Table 3.9 and Table 3.10 presents instead linear regressions of the expected (log) income and variance reported by the respondent upon primary and secondary school completion computed using triangular distribution and using the first reported probability. Expected income for male children is significantly higher in both rounds of data collection, but no significant difference is found for the variance of expected income. Similarly, education of the household head has a role

in explaining reported income, but only marginally its variance. The variance of expected income is correlated with ethnicity of the respondent and on whether the area of residence in a rural area. For the baseline data, respondents of Albanian, Roma and Turk ethnic groups tend to report a lower variance, while at follow-up the difference is weaker. At the same time, respondents who are resident in rural areas tend to report a lower variance. Similarly for the minimum and maximum expected income, we can note that, at baseline, larger unemployment in the year before the interview affect negatively reported incomes for both educational level, providing evidence that in areas with high unemployment respondents expect lower incomes. However, unemployment rates don't look to significantly affect the variance of the expected income. In conclusion, we can note that having spent a larger amount of time as recipient of Social Financial Assistance is correlated with lower expected incomes, but again no significant correlation with its variance.

3.6 Conclusion

A growing branch of literature is focusing on the use of subjective expectations in choice models. This requires advancements and a better understanding on how to elicit subjective expectations, especially when focusing on developing countries and when respondents have low levels of education. This is mainly due to the fact that questions related to probabilities might not be fully understood or might require a higher burdensome when the average level of education of respondents is low. At the same time, subjective expectations in developing countries might be central in individual and household choices since they might partly relate to the presence of market failures related to information and knowledge.

There is still no consensus in literature about the best practice to collect subjective expectations in developing countries and there is need for improvements and understanding on how to elicit subjective expectation. For this chapter, data on subjective expectations have been collected using the method proposed by Guiso et al. (2002) which, under distributional assumptions, allows eliciting the subjective expected earning distribution for each respondent. This chapter analysed the validity of answers using the data collected among social financial recipients in the Republic of Macedonia during the 2010 and 2012 data collection waves of the Macedonian "Secondary School Conditional Cash Transfer" evaluation household survey. By using the randomisation in the order of questions and testing

whether respondents understand the question and use sufficient mental effort to report their answers, this chapter showed that when asked about the probability to earn less than a certain threshold and more than the same threshold, respondents tend to reply with sufficient mental effort only to the first question they are asked. This suggests it would be sensible to use only the first information reported by the respondent, specifically because complementary answers related to probabilities tend to fail in their second component. At the same time, the chapter provided evidence on the robustness of the method in relation to different distributional assumptions.

Table 3.2: Testing for mean probability among groups to sum up to 1: first answer

	Baseline (2010)				Follow-up (2012)			
	Obs	Mean X	Mean Y	Sum	Obs	Mean X	Mean Y	Sum
Expectations after primary school								
Girls								
All observations	676	.504	.518	1.021	861	.456	.478	.935**
				(.024)				(.032)†
Exclude 0 and 1	648	.506	.529	1.035	809	.455	.473	.928**
				(.023)				(.030)††
Exclude .5, 0 and 1	436	.507	.527	1.033	497	.437	.455	.893**
				(.032)				(.042)††
Boys								
All observations	718	.513	.538	1.051**	962	.487	.488	.976
				(.023)†				(.029)
Exclude 0 and 1	704	.511	.540	1.052**	915	.483	.483	.965
				(.023)†				(.028)
Exclude .5, 0 and 1	493	.514	.537	1.052	543	.462	.451	.913**
				(.032)				(.041)†
Expectations after secondary school								
Girls								
All observations	674	.506	.530	1.036	863	.465	.477	.942**
				(.024)				(.027)††
Exclude 0 and 1	646	.509	.534	1.044*	810	.470	.474	.945**
				(.023)				(.025)††
Exclude .5, 0 and 1	434	.515	.548	1.063*	498	.451	.459	.910**
				(.033)				(.039)††
Boys								
All observations	715	.502	.538	1.040*	962	.485	.488	.973
				(.023)				(.024)
Exclude 0 and 1	701	.502	.535	1.038	915	.479	.489	.968
				(.023)				(.023)
Exclude .5, 0 and 1	489	.504	.550	1.053	543	.465	.482	.947
				(.033)				(.038)

Note. Standard errors in parenthesis. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Expectations 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. Standard errors are clustered at municipality level and * 0.10 ** 0.05 *** 0.01 represent the statistical significance of a test of equality to one of the sum of probabilities. † 0.10 †† 0.05 ††† 0.01 represent at which significance level the hypothesis of equality to one is rejected when each hypothesis presented in Table 3.2 and Table 3.3 are jointly tested. The test is implemented using Romano and Wolf (2005) procedure with 5000 bootstrap repetitions. The critical values generated through the procedure are the 90th, 95th and 99th percentile of the empirical distribution of $\tilde{t}$, where $\tilde{t}$ is the maximum t-statistic in each iteration testing, for each sum of the means, whether the coefficient estimated through bootstrap is equal to the coefficient estimated in the main sample.

Table 3.3: Testing for mean probability among groups to sum up to 1: second answer

	Baseline (2010)				Follow-up (2012)			
	Obs	Mean X	Mean Y	Sum	Obs	Mean X	Mean Y	Sum
Expectations after primary school								
Girls								
All observations	676	.354	.309	.663*** (.024)†††	861	.372	.346	.718*** (.020)†††
Exclude 0 and 1	644	.359	.323	.682*** (.024)†††	813	.385	.350	.735*** (.020)†††
Exclude .5, 0 and 1	463	.333	.299	.633*** (.027)†††	494	.360	.315	.675*** (.024)†††
Boys								
All observations	718	.349	.331	.680*** (.025)†††	962	.365	.349	.714*** (.021)†††
Exclude 0 and 1	696	.352	.338	.689*** (.025)†††	921	.371	.355	.726*** (.021)†††
Exclude .5, 0 and 1	509	.331	.315	.647*** (.028)†††	565	.341	.328	.669*** (.028)†††
Expectations after secondary school								
Girls								
All observations	674	.391	.347	.738*** (.023)†††	863	.399	.378	.777*** (.019)†††
Exclude 0 and 1	642	.399	.366	.765*** (.021)†††	815	.419	.389	.808*** (.020)†††
Exclude .5, 0 and 1	461	.356	.319	.674*** (.025)†††	496	.363	.323	.686*** (.025)†††
Boys								
All observations	715	.407	.369	.776*** (.023)†††	962	.407	.383	.790*** (.021)†††
Exclude 0 and 1	693	.413	.379	.792*** (.022)†††	920	.417	.396	.813*** (.020)†††
Exclude .5, 0 and 1	505	.378	.339	.716*** (.026)†††	564	.366	.328	.694*** (.024)†††

Note. Standard errors in parenthesis. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Expectations 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. Standard errors are clustered at municipality level and * 0.10 ** 0.05 *** 0.01 represent the statistical significance of a test of equality to one of the sum of probabilities. † 0.10 †† 0.05 ††† 0.01 represent at which significance level the hypothesis of equality to one is rejected when each hypothesis presented in Table 3.2 and Table 3.3 are jointly tested. The test is implemented using Romano and Wolf (2005) procedure with 5000 bootstrap repetitions. The critical values generated through the procedure are the 90th, 95th and 99th percentile of the empirical distribution of $\tilde{t}$, where $\tilde{t}$ is the maximum t-statistic in each iteration testing, for each sum of the means, whether the coefficient estimated through bootstrap is equal to the coefficient estimated in the main sample.

Table 3.4: Distribution and shares for the sum of reported probabilities

	Baseline (2010)		Follow-up (2012)	
	Female	Male	Female	Male
Expectations after primary school				
Sum smaller than 1	314	302	366	376
	(0.464)	(0.421)	(0.423)	(0.389)
Sum equal to to 1	297	349	444	512
	(0.439)	(0.486)	(0.513)	(0.530)
Sum larger than 1	65	67	55	78
	(0.096)	(0.093)	(0.064)	(0.081)
Expectations after secondary school				
Sum smaller than 1	255	268	312	321
	(0.378)	(0.375)	(0.360)	(0.332)
Sum equal to to 1	331	355	488	569
	(0.491)	(0.497)	(0.563)	(0.589)
Sum larger than 1	88	92	67	76
	(0.131)	(0.129)	(0.077)	(0.079)

Note. In parenthesis, I report the share of total observation in the category. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Expectations 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.

Table 3.5: Average sum of reported probabilities

	Baseline (2010)		Follow-up (2012)	
	Female	Male	Female	Male
Expectations after primary school				
Sum smaller than 1	0.618	0.641	0.560	0.566
	(0.274)	(0.254)	(0.268)	(0.260)
Sum larger than 1	1.208	1.175	1.200	1.172
	(0.152)	(0.100)	(0.139)	(0.120)
All observations	0.843	0.865	0.826	0.845
	(0.291)	(0.259)	(0.293)	(0.281)
Expectations after secondary school				
Sum smaller than 1	0.644	0.697	0.565	0.593
	(0.248)	(0.222)	(0.272)	(0.259)
Sum larger than 1	1.168	1.170	1.209	1.216
	(0.128)	(0.099)	(0.140)	(0.152)
All observations	0.887	0.908	0.860	0.882
	(0.254)	(0.222)	(0.283)	(0.262)

Note. Standard deviations in parenthesis. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Expectations 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.

Table 3.6: Comparison of expected income with first answer versus rescaling under different distributional assumptions

	Step-wise uniform		*Bi-triangular*		*Triangular*	
	Female	Male	Female	Male	Female	Male
Baseline (2010)						
Expectations after primary school						
Rescaled	8.482	8.541	8.480	8.541	8.502	8.563
	[0.448]	[0.464]	[0.444]	[0.459]	[0.444]	[0.456]
First answer	8.478	8.541	8.477	8.541	8.498	8.561
	[0.450]	[0.467]	[0.446]	[0.461]	[0.446]	[0.459]
Difference	0.002	0.000	0.001	0.000	0.003	0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Expectations after secondary school						
Rescaled	8.993	9.072	8.991	9.071	9.007	9.085
	[0.349]	[0.356]	[0.346]	[0.352]	[0.346]	[0.352]
First answer	8.992	9.070	8.990	9.070	9.006	9.084
	[0.348]	[0.355]	[0.346]	[0.351]	[0.347]	[0.352]
Difference	0.001	0.002	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Follow-up (2012)						
Expectations after primary school						
Rescaled	8.681	8.762	8.681	8.761	8.698	8.778
	[0.336]	[0.358]	[0.332]	[0.356]	[0.329]	[0.352]
First answer	8.680	8.762	8.679	8.761	8.699	8.780
	[0.338]	[0.358]	[0.333]	[0.355]	[0.328]	[0.350]
Difference	0.001	-0.000	0.000	-0.000	-0.002*	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Expectations after secondary school						
Rescaled	9.122	9.191	9.121	9.190	9.135	9.204
	[0.301]	[0.312]	[0.298]	[0.307]	[0.297]	[0.309]
First answer	9.118	9.189	9.118	9.188	9.133	9.202
	[0.304]	[0.317]	[0.300]	[0.311]	[0.297]	[0.311]
Difference	0.003*	0.002	0.002*	0.001	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)

Note. Standard deviations in brackets, standard errors in parenthesis. "Rescaled" expectations are computed by rescaling the sum of reported probabilities to be equal to one, while "first answer" expectations are computed using the first reported answer only. The difference is computed for each observation as the difference between the "rescaled" and the "first answer" values. Standard errors are clustered at municipality level and * 0.10 ** 0.05 *** 0.01 represent the statistical significance of a t-test for equality to zero of the difference. Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Expectations 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.

Table 3.7: Maximum and minimum income and individual characteristic (Baseline 2010)

	Primary school			Secondary school		
	Min	Max	Delta	Min	Max	Delta
Male child	0.054**	0.068***	0.002	0.093***	0.090***	-0.001
	(0.024)	(0.025)	(0.018)	(0.019)	(0.020)	(0.013)
Child age (years)	0.006	-0.000	-0.005	0.015***	0.005	-0.009***
	(0.007)	(0.006)	(0.004)	(0.005)	(0.005)	(0.003)
Male (head)	0.019	0.035	0.010	0.047	-0.038	-0.062
	(0.069)	(0.080)	(0.044)	(0.050)	(0.059)	(0.040)
Age head (less than 40 y.o.)	-0.013	0.007	0.019	0.040	0.005	-0.032
	(0.045)	(0.031)	(0.034)	(0.034)	(0.025)	(0.025)
Lower primary (head)	-0.024	-0.035	-0.004	-0.028	-0.079**	-0.037
	(0.053)	(0.042)	(0.030)	(0.034)	(0.033)	(0.025)
Secondary school (head)	-0.103**	-0.079**	0.021	-0.048	-0.052*	-0.001
	(0.043)	(0.040)	(0.027)	(0.030)	(0.029)	(0.020)
Albanian	0.111	-0.006	-0.112***	0.009	-0.087	-0.096***
	(0.079)	(0.066)	(0.039)	(0.055)	(0.059)	(0.030)
Roma	0.046	-0.021	-0.086*	0.001	-0.092*	-0.087**
	(0.063)	(0.054)	(0.045)	(0.048)	(0.048)	(0.037)
Turk	0.011	-0.024	-0.087*	0.064	0.048	-0.035
	(0.074)	(0.056)	(0.045)	(0.054)	(0.049)	(0.038)
Household members	-0.002	0.009	0.009	0.002	0.018	0.016*
	(0.023)	(0.013)	(0.016)	(0.015)	(0.013)	(0.009)
Boys 13-18 y.o.	0.039	-0.005	-0.052**	-0.005	-0.025	-0.026
	(0.032)	(0.029)	(0.023)	(0.021)	(0.022)	(0.016)
Girls 13-18 y.o.	0.026	0.011	-0.041**	0.009	-0.002	-0.019
	(0.034)	(0.031)	(0.018)	(0.023)	(0.024)	(0.015)
Father is present	-0.131	-0.091	0.075	-0.134*	-0.011	0.111**
	(0.085)	(0.083)	(0.050)	(0.072)	(0.068)	(0.050)
Mother is present	0.006	0.029	0.046	0.062	0.099	0.045
	(0.093)	(0.061)	(0.057)	(0.065)	(0.061)	(0.044)
Rural	0.070	-0.035	-0.112***	0.040	-0.052	-0.091***
	(0.054)	(0.045)	(0.032)	(0.049)	(0.047)	(0.030)
Part of City of Skopje	0.043	-0.100	-0.135	0.018	-0.082	-0.074
	(0.155)	(0.143)	(0.108)	(0.098)	(0.103)	(0.081)
Wealth (low)	-0.113**	-0.077**	0.025	-0.047	-0.037	0.012
	(0.048)	(0.034)	(0.038)	(0.033)	(0.031)	(0.025)
Wealth (high)	0.044	-0.027	-0.068*	0.055*	0.006	-0.053**
	(0.054)	(0.046)	(0.035)	(0.033)	(0.034)	(0.026)
Time in Fin.Ass. (1-6 years)	-0.017	0.032	0.048	-0.064*	-0.039	0.022
	(0.053)	(0.047)	(0.035)	(0.037)	(0.037)	(0.023)
Time in Fin.Ass. (> 6 years)	-0.084	-0.038	0.035	-0.065*	-0.059*	0.001
	(0.054)	(0.048)	(0.034)	(0.036)	(0.035)	(0.025)
Unemployment (<=0.30)	0.220**	0.237***	0.029	0.188**	0.265***	0.089*
	(0.109)	(0.082)	(0.058)	(0.072)	(0.068)	(0.047)
Constant	8.239***	8.790***	0.554***	8.630***	9.160***	0.515***
	(0.199)	(0.153)	(0.130)	(0.124)	(0.127)	(0.105)
Observations	1328	1354	1328	1364	1367	1364

Note. Standard errors in parenthesis (* 0.10 ** 0.05 *** 0.01). Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Minimum and maximum income are reported in logarithms. Unemployment is computed at regional level for the year before the interview and ranges from 14.4 percent to 64.8 percent in 2009 and from 9.3 percent to 42.8 percent in 2011. Months in Financial Assistance are reported by the respondent. Omitted categories include: Female child, Macedonian and other ethnicities, Female head, Time in Financial Assistance (less than 1 year), Unemployment (larger than 30 percent). All specifications include regional dummies.

Table 3.8: Maximum and minimum income and individual characteristic (Follow-up 2012)

	Primary school			Secondary school		
	Min	Max	Delta	Min	Max	Delta
Male child	0.078***	0.080***	0.000	0.078***	0.077***	-0.002
	(0.020)	(0.017)	(0.012)	(0.017)	(0.014)	(0.010)
Child age (years)	0.006	0.005*	0.000	0.001	0.003	0.002
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)
Male (head)	0.043	0.034	-0.011	0.069*	0.010	-0.057*
	(0.050)	(0.040)	(0.039)	(0.039)	(0.027)	(0.034)
Age head (less than 40 y.o.)	-0.066	0.045	0.102**	0.056*	0.059*	0.003
	(0.055)	(0.034)	(0.043)	(0.033)	(0.031)	(0.026)
Lower primary (head)	-0.161***	-0.066**	0.097***	-0.070**	-0.073**	-0.005
	(0.045)	(0.031)	(0.035)	(0.032)	(0.031)	(0.023)
Secondary school (head)	0.035	0.048*	0.006	0.060**	0.061**	-0.000
	(0.032)	(0.025)	(0.023)	(0.025)	(0.024)	(0.029)
Albanian	0.047	0.031	-0.034	-0.005	-0.003	0.006
	(0.051)	(0.041)	(0.042)	(0.036)	(0.037)	(0.044)
Roma	0.065	0.003	-0.070*	0.004	-0.031	-0.035
	(0.058)	(0.041)	(0.039)	(0.039)	(0.040)	(0.027)
Turk	0.064	-0.028	-0.091**	0.022	-0.038	-0.056*
	(0.061)	(0.043)	(0.041)	(0.041)	(0.034)	(0.031)
Household members	0.004	0.003	-0.004	0.000	0.002	0.001
	(0.014)	(0.011)	(0.009)	(0.011)	(0.011)	(0.007)
Boys 13-18 y.o.	0.013	-0.005	-0.015	-0.003	-0.018	-0.014
	(0.024)	(0.019)	(0.015)	(0.018)	(0.017)	(0.014)
Girls 13-18 y.o.	-0.011	-0.002	0.009	-0.025	-0.007	0.019
	(0.023)	(0.017)	(0.016)	(0.017)	(0.016)	(0.013)
Father is present	-0.161**	-0.046	0.125***	-0.064	-0.019	0.042
	(0.061)	(0.051)	(0.047)	(0.050)	(0.039)	(0.046)
Mother is present	-0.030	-0.036	-0.003	-0.075	-0.036	0.039
	(0.072)	(0.059)	(0.063)	(0.076)	(0.070)	(0.035)
Rural	-0.028	-0.056	-0.031	-0.071	-0.088*	-0.018
	(0.055)	(0.042)	(0.043)	(0.044)	(0.045)	(0.038)
Part of City of Skopje	0.063	-0.022	-0.058	-0.044	-0.033	-0.000
	(0.112)	(0.107)	(0.080)	(0.067)	(0.061)	(0.060)
Wealth (low)	0.025	-0.030	-0.042*	0.020	0.016	-0.002
	(0.038)	(0.031)	(0.025)	(0.029)	(0.027)	(0.022)
Wealth (high)	0.073**	0.043	-0.031	0.017	0.037	0.021
	(0.031)	(0.030)	(0.029)	(0.030)	(0.030)	(0.030)
Time in Fin.Ass. (1-6 years)	-0.045	0.017	0.060**	-0.015	-0.012	0.002
	(0.038)	(0.032)	(0.029)	(0.028)	(0.029)	(0.023)
Time in Fin.Ass. (> 6 years)	-0.096***	-0.072**	0.029	-0.075***	-0.090***	-0.016
	(0.035)	(0.029)	(0.028)	(0.026)	(0.025)	(0.026)
Unemployment (<=0.30)	-0.026	0.115*	0.039	0.160***	-0.032	-0.123***
	(0.120)	(0.060)	(0.093)	(0.053)	(0.069)	(0.046)
Constant	8.431***	8.799***	0.364***	8.923***	9.299***	0.380***
	(0.141)	(0.097)	(0.111)	(0.123)	(0.104)	(0.088)
Observations	1415	1434	1415	1437	1442	1437

Note. Standard errors in parenthesis (* 0.10 ** 0.05 *** 0.01). Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Minimum and maximum income are reported in logarithms. Unemployment is computed at regional level for the year before the interview and ranges from 14.4 percent to 64.8 percent in 2009 and from 9.3 percent to 42.8 percent in 2011. Months in Financial Assistance are reported by the respondent. Omitted categories include: Female child, Macedonian and other ethnicities, Female head, Time in Financial Assistance (less than 1 year), Unemployment (larger than 30 percent). All specifications include regional dummies.

Table 3.9: Expected income moments and individual characteristic (Baseline 2010)

	Primary school		Secondary school	
	Mean	Variance	Mean	Variance
Male child	0.055**	-0.000	0.092***	0.001
	(0.023)	(0.003)	(0.018)	(0.001)
Child age (years)	0.004	-0.001	0.009**	-0.001**
	(0.006)	(0.001)	(0.005)	(0.000)
Male (head)	0.040	0.002	0.016	-0.004
	(0.061)	(0.007)	(0.043)	(0.004)
Age head (less than 40 y.o.)	-0.002	0.000	0.024	-0.003
	(0.034)	(0.005)	(0.027)	(0.003)
Lower primary (head)	-0.016	-0.002	-0.050	-0.004*
	(0.048)	(0.005)	(0.033)	(0.002)
Secondary school (head)	-0.100**	0.001	-0.049*	-0.000
	(0.039)	(0.004)	(0.028)	(0.002)
Albanian	0.043	-0.015***	-0.039	-0.009***
	(0.065)	(0.005)	(0.054)	(0.003)
Roma	-0.004	-0.012*	-0.043	-0.009**
	(0.054)	(0.007)	(0.045)	(0.004)
Turk	-0.029	-0.012**	0.048	-0.004
	(0.066)	(0.006)	(0.050)	(0.004)
Household members	0.004	0.001	0.011	0.001
	(0.017)	(0.002)	(0.013)	(0.001)
Boys 13-18 y.o.	0.010	-0.007**	-0.024	-0.002
	(0.027)	(0.003)	(0.019)	(0.002)
Girls 13-18 y.o.	0.005	-0.006*	-0.004	-0.001
	(0.031)	(0.003)	(0.022)	(0.002)
Father is present	-0.105	0.007	-0.074	0.008
	(0.073)	(0.007)	(0.063)	(0.006)
Mother is present	0.036	0.009	0.106*	0.006
	(0.071)	(0.007)	(0.058)	(0.004)
Rural	0.006	-0.013***	-0.020	-0.008***
	(0.047)	(0.004)	(0.046)	(0.003)
Part of City of Skopje	-0.035	-0.013	-0.023	-0.004
	(0.127)	(0.019)	(0.086)	(0.010)
Wealth (low)	-0.151***	0.005	-0.053*	0.002
	(0.035)	(0.005)	(0.028)	(0.002)
Wealth (high)	-0.038	-0.009*	-0.002	-0.003
	(0.040)	(0.005)	(0.029)	(0.003)
Time in Fin.Ass. (1-6 years)	0.019	0.005	-0.049	0.002
	(0.046)	(0.005)	(0.035)	(0.002)
Time in Fin.Ass. (> 6 years)	-0.055	0.005	-0.065*	0.001
	(0.047)	(0.004)	(0.034)	(0.003)
Unemployment <=30	(0.096)	(0.007)	(0.068)	(0.004)
Constant	8.547***	0.033**	8.911***	0.027**
	(0.168)	(0.015)	(0.117)	(0.010)
Observations	1328	1328	1364	1364

Note. Standard errors in parenthesis (* 0.10 ** 0.05 *** 0.01). Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Mean and variance of expected (log-)income are computed assuming a triangular distribution and using the first reported probability. Unemployment is computed at regional level for the year before the interview and ranges from 14.4 percent to 64.8 percent in 2009 and from 9.3 percent to 42.8 percent in 2011. Months in Financial Assistance are reported by the respondent. Omitted categories include: Female child, Macedonian and other ethnicities, Female head, Time in Financial Assistance (less than 1 year), Unemployment (larger than 30 percent). All specifications include regional dummies.

Table 3.10: Expected income moments and individual characteristic (Baseline 2012)

	Primary school		Secondary school	
	Mean	Variance	Mean	Variance
Male child	0.079***	-0.000	0.078***	-0.000
	(0.017)	(0.002)	(0.015)	(0.001)
Child age (years)	0.006*	-0.000	0.001	0.000
	(0.003)	(0.000)	(0.003)	(0.000)
Male (head)	0.019	0.001	0.019	-0.003
	(0.040)	(0.004)	(0.031)	(0.003)
Age head (less than 40 y.o.)	-0.000	0.010*	0.053*	-0.000
	(0.038)	(0.005)	(0.030)	(0.002)
Lower primary (head)	-0.115***	0.010**	-0.074**	0.001
	(0.035)	(0.004)	(0.030)	(0.002)
Secondary school (head)	0.049*	-0.001	0.064***	0.001
	(0.026)	(0.002)	(0.021)	(0.003)
Albanian	0.030	-0.005	-0.003	0.000
	(0.037)	(0.004)	(0.030)	(0.004)
Roma	0.030	-0.008	-0.013	-0.005*
	(0.044)	(0.005)	(0.037)	(0.003)
Turk	0.014	-0.010**	-0.014	-0.004
	(0.045)	(0.004)	(0.033)	(0.003)
Household members	0.000	-0.001	0.000	-0.000
	(0.012)	(0.001)	(0.010)	(0.001)
Boys 13-18 y.o.	0.002	-0.001	-0.013	-0.001
	(0.020)	(0.001)	(0.016)	(0.001)
Girls 13-18 y.o.	-0.010	0.001	-0.014	0.001
	(0.019)	(0.002)	(0.016)	(0.001)
Father is present	-0.065	0.010**	-0.022	0.004
	(0.049)	(0.004)	(0.037)	(0.004)
Mother is present	-0.018	-0.003	-0.050	0.005*
	(0.055)	(0.007)	(0.070)	(0.003)
Rural	-0.039	-0.002	-0.079**	-0.002
	(0.042)	(0.004)	(0.039)	(0.003)
Part of City of Skopje	0.030	-0.005	-0.053	-0.002
	(0.085)	(0.008)	(0.060)	(0.005)
Wealth (low)	0.033	-0.004	0.038	0.000
	(0.029)	(0.003)	(0.024)	(0.002)
Wealth (high)	0.047*	-0.001	0.020	0.002
	(0.027)	(0.002)	(0.027)	(0.003)
Time in Fin.Ass. (1-6 years)	0.004	0.003	0.001	0.000
	(0.032)	(0.003)	(0.028)	(0.002)
Time in Fin.Ass. (> 6 years)	-0.078***	0.001	-0.080***	-0.001
	(0.028)	(0.003)	(0.023)	(0.002)
Unemployment <=30	(0.075)	(0.004)	(0.052)	(0.003)
Constant	8.597***	0.016	9.128***	0.007
	(0.102)	(0.013)	(0.106)	(0.007)
Observations	1415	1415	1437	1437

Note. Standard errors in parenthesis (* 0.10 ** 0.05 *** 0.01). Expectations are restricted to recipients of Social Financial Assistance and include all respondents (including resampled households at follow-up). Mean and variance of expected (log-)income are computed assuming a triangular distribution and using the first reported probability. Unemployment is computed at regional level for the year before the interview and ranges from 14.4 percent to 64.8 percent in 2009 and from 9.3 percent to 42.8 percent in 2011. Months in Financial Assistance are reported by the respondent. Omitted categories include: Female child, Macedonian and other ethnicities, Female head, Time in Financial Assistance (less than 1 year), Unemployment (larger than 30 percent). All specifications include regional dummies.

Chapter 4

Parental perceived returns to schooling and human capital investment

4.1 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 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 4.2 presents a theoretical framework to describe how parental expectations affect investment in children's human capital. Section 4.3 describes the empirical strategy and Section 4.4 presents the data used in the chapter. Section 4.5 shows the main results and presents the robustness checks.

4.2 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:

$$U(y_j) = \frac{y_j^{1-\gamma}}{1-\gamma} \tag{4.1}$$

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 by the following maximisation problem

$$\max_{\delta_i \in [0,1]} \delta_i [\beta_i \frac{U(y_s)}{c_i \epsilon_i}] + (1 - \delta_i)(1 + \beta_i) U(y_p) \tag{4.2}$$

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

[1] A detailed derivation of the results is presented in 4.6.

(discounted) expected utility from the completion of secondary school is larger than the (discounted) expected utility from the completion on primary school only:

$$\beta_i \frac{1}{c_i \epsilon_i} \int U_i(y_s)\, f^i_{Y_s}(y_s)\, dy_s > (1+\beta_i) \int U_i(y_p)\, f^i_{Y_p}(y_p)\, dy_p \tag{4.3}$$

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

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

By substituting for the expected utilities in equation (4.3) and taking logs, we can write the condition for enrolling the child in secondary school as

$$\Phi \equiv \ln \frac{\beta_i}{1+\beta_i} - \tilde{c} + (1-\gamma)\,\mu_s + \frac{(1-\gamma)^2}{2}\sigma_s^2 - (1-\gamma)\,\mu_p - \frac{(1-\gamma)^2}{2}\sigma_p^2 > \tilde{\epsilon}_i \tag{4.4}$$

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

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

where $F_{\tilde{\epsilon}}\left(\tilde{\epsilon}\right)$ 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, $Pr\left[\delta_i = 1 \,|\, \Phi\right]$, 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:

$$\frac{\partial Pr\left[\delta_i = 1 \mid \Phi\right]}{\partial \mu_p} = -g_\epsilon\left[-\Phi\right]\left(1-\gamma\right) < 0 \tag{4.5}$$

$$\frac{\partial Pr\left[\delta_i = 1 \mid \Phi\right]}{\partial \mu_s} = g_\epsilon\left[-\Phi\right]\left(1-\gamma\right) > 0 \tag{4.6}$$

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.

$$\frac{\partial Pr\left[\delta_i = 1 \mid \Phi\right]}{\partial \sigma_p^2} = -g_\epsilon\left[-\Phi\right]\frac{\left(1-\gamma\right)^2}{2} < 0 \tag{4.7}$$

$$\frac{\partial Pr\left[\delta_i = 1 \mid \Phi\right]}{\partial \sigma_s^2} = g_\epsilon\left[-\Phi\right]\frac{\left(1-\gamma\right)^2}{2} > 0 \tag{4.8}$$

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.

4.3 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

[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

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

$$\begin{aligned} \delta_{im,2012} &= 1 \\ &\Leftrightarrow \\ \delta^*_{im,2012} &= \alpha + \beta_0 \cdot ExpIncPrim_{i,2010} + \beta_2 \cdot ExpIncSec_{i,2010} + \\ &\quad + \sum_{j=1}^{2} \tau_j \cdot VarInc_{ij,2010} + X_i'\gamma + M_m'\eta + \epsilon_{im} > 0 \end{aligned} \tag{4.9}$$

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 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

et al. (2012).

probability. Therefore I extend equation 4.9 estimating the following model

$$\begin{aligned}\delta^*_{im,2012} &= \alpha + \beta_0 \cdot ExpIncPrim_{i,2010} + \beta_2 \cdot ExpIncSec_{i,2010} + \\ &+ \sum_{j=1}^{2} \tau_j \cdot VarInc_{ij,2010} + \sum_{j=1}^{2} \gamma_j \cdot PrWork_{ij,2010} + \\ &+ X_i'\gamma + M_m'\eta + \epsilon_{im} > 0 \end{aligned} \tag{4.10}$$

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.

4.4 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 4.1 presents the main descriptive statistics on child and household characteristics.

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

[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.

without mentioning directly the term "probability" (Attanasio et al., 2005; 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. Section 3.2.1 presents in details the methodology selected for eliciting subjective expectations. Chapter 3 presents instead a series of validity tests.

Table 4.2 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.

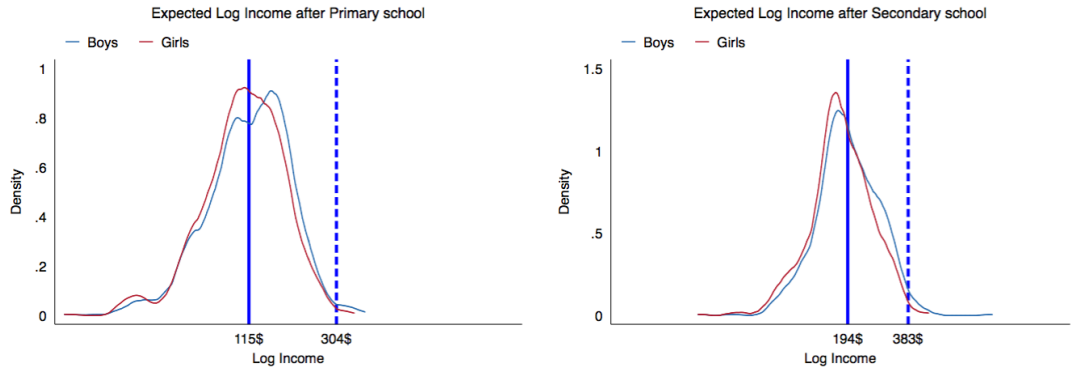
In order to compare parental expected income and market returns, Figure 4.1 presents a comparison between the sample distribution of expected income conditional on completion of primary or secondary school with the Macedonian national average net wage for the correspondent education group. For both boys 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.

4.5 Results

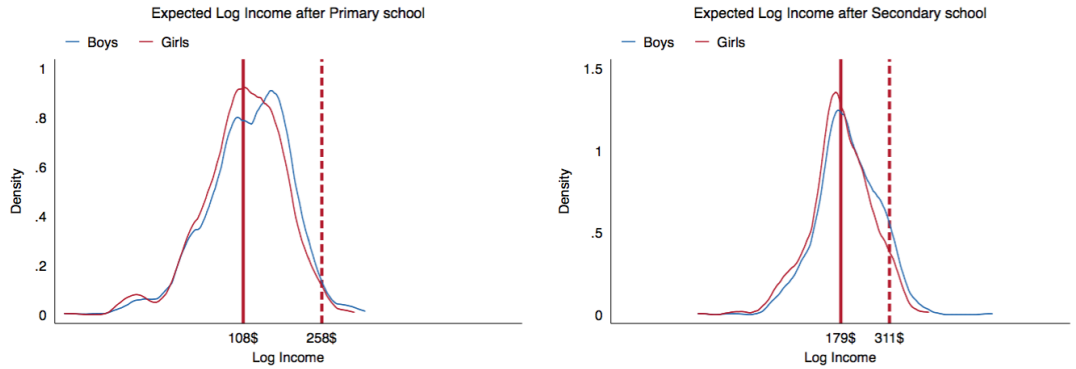
This section presents the estimates of Equations 4.9 and 4.10 for the sample of children in all Social Financial Assistance households born from 1993 to 1998, for which data about subjective expectations are available at 2010. In all specifica-

Figure 4.1: 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.

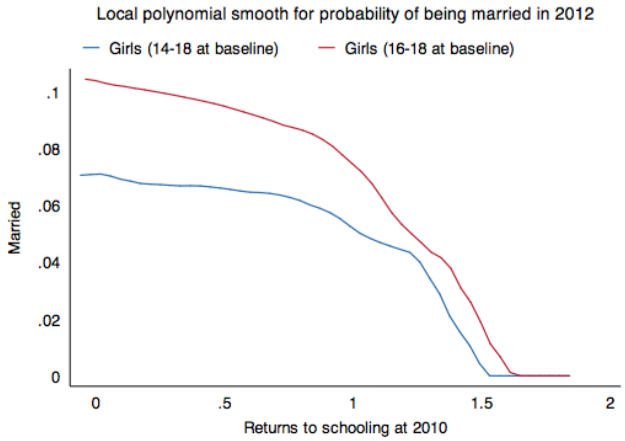
tions, the dependent variable is an indicator variable that is equal to 1 if the child is enrolled or has completed any secondary school at the beginning of the school year 2012/2013 and is equal to 0 otherwise. Controls include gender and age of the child, education, gender and age of the household head, ethnicity, religion, household size and number of children, household asset group and distance from the closest school[4] 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 4.4 shows the estimates of Equations 4.9 and 4.10 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 4.10 is estimated separately for boys and girls in urban and rural areas. Table 4.5 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 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

[4]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.

Figure 4.2: 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.

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 4.2).

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, recip-

ients 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 4.10 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[5]. 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 4.6 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 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.

4.5.1 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[6]

[5]See 5.8 for details on how the index was built and for robustness checks.

[6]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

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[7], 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

$$\begin{aligned} E[y_j] &= p\,E[y_j|E_i=1] + (1-p)\,E[y_j|E_i=0] \\ &= p\,\bar{y}_{j,E} + (1-p)\,y_{UN} \end{aligned} \tag{4.11}$$

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 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:

$$\begin{aligned} Var[y_j] &= E[y_j^2] - E[y_j]^2 \\ &= p\,E[y_j^2|E_i=1] + (1-p)\,E[y_j^2|E_i=0] - E[y_j]^2 \\ &= p\left[E[y_j^2|E_i=1] + \bar{y}_{j,E}^2\right] + (1-p)\,\bar{y}_{UN}^2 - E[y_j]^2 \\ &= pVar[y_j|E_i=1] + p\,(1-p)\left(\bar{y}_{j,E} - y_{UN}\right)^2 \end{aligned} \tag{4.12}$$

Table 4.3 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 4.4, Table 4.7 presents the estimates of Equations 4.9 and 4.10 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

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).

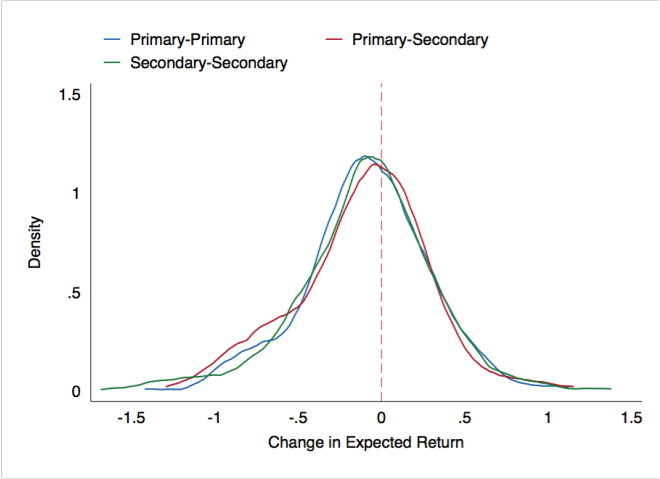
[7]This assumption might be restrictive since the expected income in case of unemployment might vary by level of education in case it is generated by activities in the informal sector rather than state assistance. However, we don't have enough information to develop further this difference.

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.

4.5.2 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. These issues are treated in details in Section 5.7.2. 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 4.8 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 4.9 compares the coefficient by estimating the model using a linear probability model versus a probit model. Results are

Figure 4.3: 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.

robust to the two estimation methods, before and after controlling for individual characteristics.

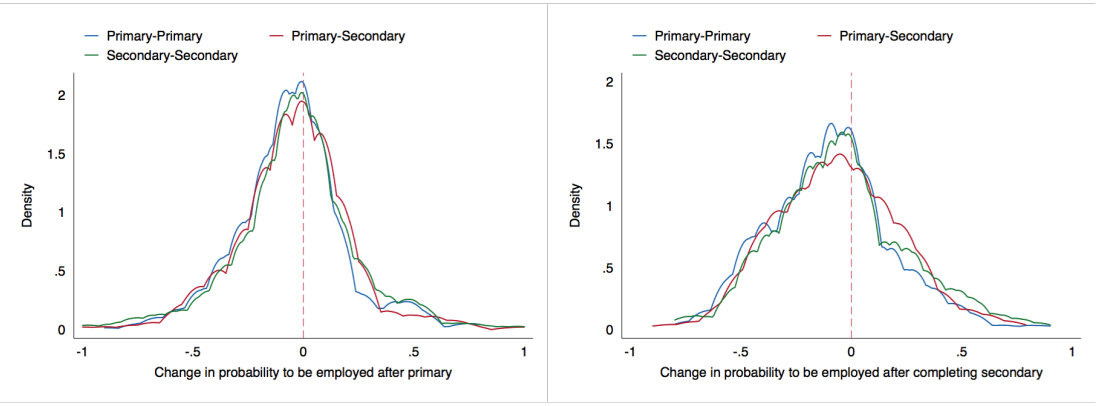
4.5.3 Cognitive dissonance bias

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:

$$E[Y|E_i = 1, J = j] > E^*[Y|E_i = 1, J = j] \tag{4.13}$$

Using subjective expectations affected by cognitive dissonance in choice models would therefore upward bias our estimates. In order to test for cognitive dissonance, I make use of the panel dimension of the dataset and I compare the

Figure 4.4: 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.

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 4.3 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 4.4 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 5.11). 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 (see Section 5.7.2).

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 4.5 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.

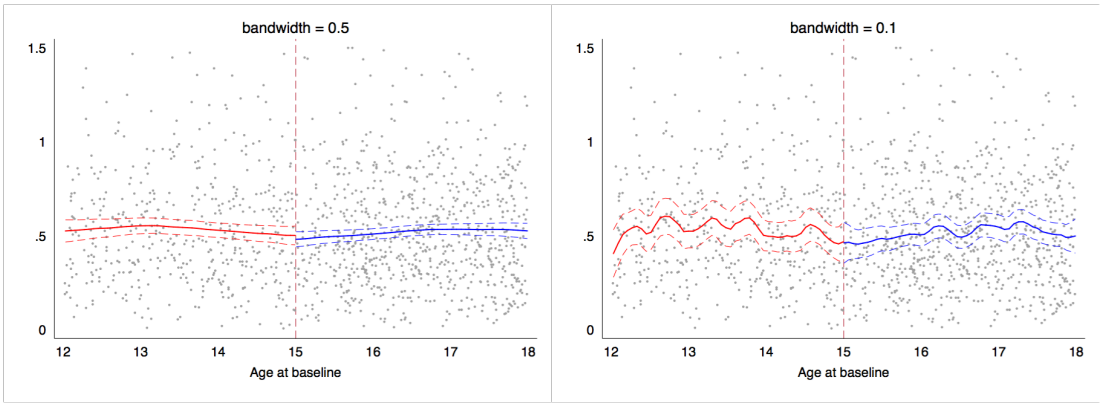
4.6 Conclusion

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).

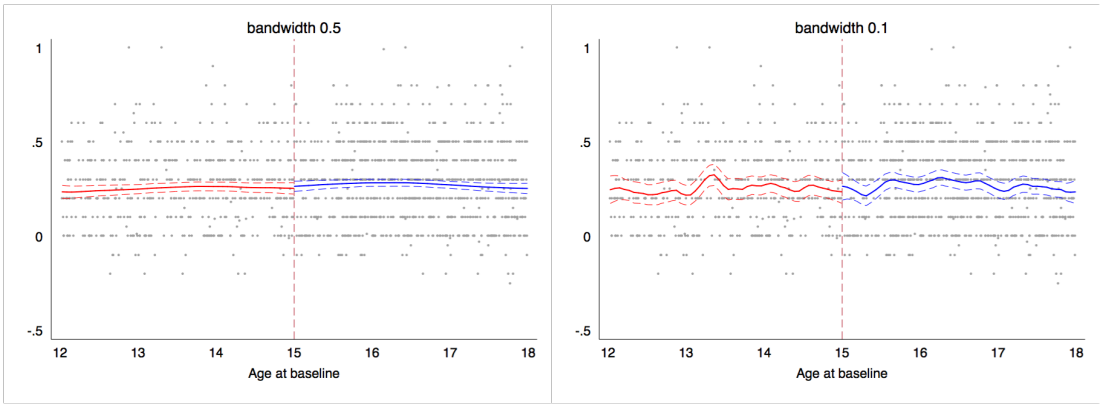
I provide evidence that ex-ante parental expectations are important in explaining secondary schooling decisions for children. Additionally, important differences exists across gender. This chapter shows that expected income conditional on completing secondary school is particularly important for girls' 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.

Figure 4.5: 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.

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.

Table 4.1: 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 4.2: 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 4.3: 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 4.4: Enrolment regression and parental perceived returns

	Dep.var.: Enrolled or completed secondary school					
	Probit (1)	Probit (2)	Probit (3)	Probit (4)	Probit (5)	Probit (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 4.5: 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 4.6: 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 4.7: Enrolment regression and parental perceived unconditional returns

	Dep.var.: Enrolled or completed secondary school			
Unemployment income hypothesis	1000 MKD		3000 MKD	
	Probit (1)	Probit (2)	Probit (3)	Probit (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 4.8: Enrolment regression, parental perceived returns and distance to school

	Dep.var.: Enrolled or completed secondary school		
	Probit	Probit	Probit
	(1)	(2)	(3)
Expected income (prim.)	-0.027	-0.027	-0.040
	(0.058)	(0.057)	(0.057)
Expected income (sec.)	0.190**	0.190**	0.161**
	(0.076)	(0.076)	(0.075)
Var. income (prim.)	-0.178	-0.172	-0.179
	(0.533)	(0.533)	(0.529)
Var. income (sec.)	0.308	0.308	0.440
	(0.695)	(0.697)	(0.691)
Prob. of employment (prim.)	-0.225*	-0.225	-0.177
	(0.137)	(0.137)	(0.110)
Prob. of employment (sec.)	0.269**	0.269**	0.223**
	(0.107)	(0.107)	(0.108)
Distance to school		-0.004	-0.126
		(0.018)	(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 4.9: 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 4.A
Proof for Section 4.2

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

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

Chapter 5

Who wears the trousers in the family?

5.1 Introduction

A large body of research shows that Conditional Cash Transfer (CCT) programs in the developing world have been beneficial for human capital investments in children among poor households[1]. However, the vast majority of these programs transfer money to women in the household, keeping uncertainty on whether the delivered effect is due to an increase of resources or to a behavioural change related to their use. The understanding of these mechanisms is particularly important in countries where the empowerment of women may have important indirect implications on children's human capital investment, especially through behavioural changes within households.

At the same time, decisions to invest in human capital are likely to depend on expected costs and benefits of schooling. It is therefore reasonable to believe that the effect of CCT programs transferring money to women rather than men should vary with parental expected returns to schooling[2]. This is particularly important in environments where limited or imperfect information about future income possibilities are a deeper issue. Still, it is unclear how the link between the identity of the recipient receiving a cash transfer and the perceived returns to schooling in the family might influence human capital investments in their

[1] Fiszbein and Schady (2009) provide a comprehensive review of the literature on the effects of Conditional Cash Transfers.

[2] I focus only on benefits, since the role of costs should be symmetric.

children.

This chapter studies the effect on children's human capital investment of channelling cash transfers to women versus household heads when parental[3] perceived returns to schooling are heterogeneous across households. I study this effect in the context of the CCT for Secondary School education in the Republic of Macedonia, the first CCT program to be implemented in the Balkan region. This is a national program providing cash transfers to poor households conditional on secondary school attendance of their children. What makes this program unique is that this is the first nationally implemented study to randomise cash transfers to the household head versus the mother of the child. The CCT provides an exogenous variation in the share of resources controlled by mothers that can be studied to determine whether channelling resources to women has an impact on children's human capital investment. To understand whether this effect is varying depending on perceived returns to schooling, I collected a unique dataset on parental expectations of children's future income possibilities and employment probability under different educational achievements[4]. This allowed uncovering the large heterogeneity characterising parental perceived returns to schooling.

I find that targeting CCT payments to mothers had a small and not statistically significant effect on child enrolment in secondary school. However, when ex-ante parental perceived returns to schooling are sufficiently large, channelling resources to women led to an increase in enrolment and achievement rates. This effect is associated with an increase in individual expenditure shares on education for the children in this group. For children in the higher tercile of the parental perceived returns, individual expenditure shares increased by roughly 4 percent. These findings support the idea that households tend to invest more in children when the payment is transferred to mothers, but only when the perceived returns to schooling are large enough to justify the investment. In order to understand the joint effect of channelling transfers to mothers and high parental perceived returns to schooling on human capital investment, we need to disentangle the relationship between each of the two components with the decision to invest in children.

[3]Throughout the chapter I will consider parental expectations as the shared expectations of both parents. A test about about this hypothesis is discussed in the Appendix.

[4]Since the aim of the Macedonian CCT program is to increase the low secondary school enrolment of children in poor households, in the chapter I focus on the return of completing secondary school versus completing only primary school.

The majority of educational and health-related social programs have targeted women in the past using the justification that mothers have stronger preferences for child education and health, but evidence about this mechanism is still lacking[5]. Policy interventions shifting the relative income of women versus men within households have proven to have an effect on different family decisions (Lundberg et al. (1997) and Ward-Batts (2008) use the 1979 UK Reform of Child benefits, Attanasio and Lechene (2002) and Bobonis (2009) use Mexican Progresa). However, there is little experimental evidence on the differential effect of targeting a payment to mothers or fathers when the objective is to subsidise education. Recently Benhassine et al. (2013) studied an unconditional (labeled) cash transfer for primary school attendance in Morocco using an experimental design. They compare payments made to fathers versus a more standard modality of payments made to mothers and they find very little effect of targeting fathers versus mothers. However, the program object of the study is based in Morocco, where the vast majority of the population is muslim. Similarly, Akresh et al. (2012) studied the effect of different CCT modalities on preventative health visits in rural Burkina Faso and they find no effect of targeting payments to women. It is reasonable to believe that targeting women might generate different outcomes depending on the social norms that characterise the household, which are strong determinants of the relative power distribution across household members. In Macedonia, this is particularly important since the heterogeneity in our sample, in terms of religions and ethnicities[6], allows studying the effect of targeting payment to mothers in a very diverse environments.

Results suggest that one possible channel through which targeting payments to mothers affect human capital investment is through a differential allocation of resources within the household. This is supported by the literature, which provides evidence that targeting payments to mothers could lead to a distinct impact through a shift in relative decision power within the household, resulting in a differential allocation of expenditures. A large body of research reports evidence that the amount of resources that each household member contributes to the family affects its allocation of expenditures (for a literature review, see Duflo, 2005). However, there is no clear consensus on the precise mechanism through which households take decisions and allocate consumption when receiving a cash transfer. Evidence on collective models suggests that targeting payments to a

[5]The other reason is that women have a lower participation in the labour market and therefore have more available time to collect the payment and fulfil administrative procedures.

[6]The sample includes households from Macedonian, Albanian, Turkish and Roma ethnicities. At the same time across ethnicities, I observe variation in religion (christian orthodox and muslim).

precise household member might result in different outcomes[7] if individuals have different preferences and the targeted payment affects the relative decision power within the household. At the same time, evidence suggests that women have different preferences over consumption than men, favouring public goods rather than private goods (Thomas, 1990; Hoddinott and Haddad, 1995; Lundberg et al., 1997; Doss, 2006; Ward-Batts, 2008). This chapter contributes to this branch of literature by providing evidence that channelling resources to women might indeed change the allocation of expenditures.

In CCTs the presence of a conditionality gives parents incentives, such as monitoring of school attendance, that might have additional indirect effect on education and that might depend on who is entitled to receive the cash transfer. Different studies provide evidence that conditionality is beneficial, since it might generate the incentive to improve performance in order to achieve the conditionality (De Brauw and Hoddinott, 2011; Barrera-Osorio et al., 2011; Baird et al., 2012; Akresh et al., 2012). This might translate in a differential incentive on monitoring of child enrolment and attendance and on a differential incentive to invest more time (or less) with the child. However, this chapter provides evidence against differential effects related to the conditionality of targeting mothers. I find no impact on the frequency in which parents monitor their children and on the amount of time they spend with them. One possible reason is that for children of secondary school age, this mechanism might not be relevant since parents have smaller control or influence over them.

Human capital investments in children are likely to be influenced not only by whether who control resources in the household is more prone to invest on it, but also by how parents value its costs and benefits. It is realistic to believe that parents make schooling decisions for their children based on subjective expectations rather than actual schooling returns (Manski, 2004), 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

[7] I will discuss in the chapter the alternative of targeting different adults in the household. However, an alternative strategy would be to target directly children, especially beyond a certain age, but there is little or no evidence on the effect of such targeting in developing countries. Ashworth et al. (2002) studied the effect of targeting children for the Education Maintenance Allowance (EMA) in the United Kingdom, which provides a cash subsidy to young people aged 16–19 from poor households to avoid drop-outs before the end of compulsory education. At the piloting stage of the program, different variants of the program were tested and among other one was comparing the targeting of parents versus the targeting of children. The authors find that the effect on participation was twice as large when the subsidy was paid to young individuals.

is little reason to believe that individuals with similar information form their expectations in the same way. Instead, the availability of subjective expectations allows eliciting ex-ante beliefs given information that parents have at the time of decision making, allows controlling for the heterogeneity in people's expectations and on ex-ante perceptions of employment risk. This chapter allows testing for effect heterogeneity of targeting payments to women by directly eliciting subjective returns to schooling at the beginning of the program.

Most of the early chapters linking perceived returns to schooling to educational choices focused on developed countries, while only recently the attention turned to developing countries. This is particularly important since, among poor households, the decision of investing in human capital might be strictly related to expectations, due to the fact that budget constraints might be more binding. Evidence shows that in developed countries individuals have fairly correct expectations of returns to schooling[8], while there is still little evidence on how returns to schooling are perceived in developing countries. In such environment, it is reasonable to believe that students and parents are not well-informed about future returns to schooling. This might be related to a scarce availability of information about earnings, especially when informal labour markets are large.

Perceived returns are particularly important for developing countries since measured returns are high, but schooling tend to remain low. As noted by Jensen (2010), in the Dominican Republic around 80-90% of youths complete primary school, but only 25% to 30% complete secondary school, compared to a secondary school return of over 40%. In his study, he finds that 8th grade pupils underestimate the returns to schooling, while informing a random set of children about the average returns to schooling significantly increase their perceived returns and the attained years of schooling. Similarly, Nguyen (2008) finds that informing a random subset of children in Madagascar about their returns to schooling increased attendance and test scores. Attanasio and Kaufmann (2009) for Mexico finds that higher expected returns are related to higher college attendance. These studies provide evidence that perceived expected returns are heterogeneous even across comparably similar groups. If learning about future income is happening

[8]Freeman (1971) and Betts (1996) were among the first to collect individual information among college undergraduates about earnings for different categories of jobs. Smith and Powell (1990) collected college seniors' income expectations for the first year of their job and after 10 years. Similarly, Blau (1990) collected college seniors' information about initial, after 10 years and after 20 years if they were to stay in the same occupation after leaving school. Dominitz and Manski (1996) provided the first computer-assisted interview to collect information among high school and college students.

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. Consistently with this literature, I show that parental perceived returns are particularly important determinants of secondary schooling decisions in Macedonia.

Whether information matters for educational choices might depend as well on whose information we are considering. This chapter focuses on parental expectations without distinguishing between mother and father since information is collected when both parents are present. There is no evidence in literature that justifies differences in perceived returns to schooling across partners and I believe it is reasonable to assume partners are sharing the same information set. In contrast, it would not be reasonable to assume that parents have the same expectations of their children. Attanasio and Kaufmann (2009) provide evidence for college enrolment in Mexico by using responses about schooling returns from mothers and from children and find that there are significant differences between male and female children. Mother's expectations are important for female enrolment into college, while they don't matter for male children. Giustinelli (2011) provides evidence instead on whether major choice for high school students depends on parental versus child expectations.

The chapter is organised as follow. In Section 2, I present the theoretical framework. In Section 3, I present the Macedonian CCT for Secondary School Education and the research design. In Section 4, I present the data used in the chapter and the way different measures are constructed. In Section 5 I present the empirical strategy and the results of the chapter, while in Section 6 I present the robustness checks.

5.2 Theoretical framework

Targeting payments to different household members has indirect implications for the welfare analysis of human capital investment on children. This section illustrates how the effect of channelling resources to women within a household can

be studied in a collective household framework[9]. I consider a static version of a collective model for the decision to either consume or invest in child education for a household composed by two decision-makers (mother and father, indicated by the subscript m and f) and their child. The household decides how to allocate income (y) between consumption (c), which includes private and public consumption, and human capital investment for their child (h). Individual d preferences are represented by a twice continuously differentiable utility function $U_d(c, h)$, which I assume is separable in consumption and human capital investment. I assume therefore that the utility function for each decision maker is defined by

$$U_d(c, h) = u(c) + v_d(r \cdot h) \tag{5.1}$$

where r is the return (in terms of utility) of the human capital investment. I am therefore assuming that both parents have the same preferences for consumption, while they have different tastes for human capital investment. In addition, I assume the household faces uncertainty on the return to human capital investment and, for simplicity, I assume that the rate of return can be either low or high, $r = (r_L, r_H)$. The probability assigned to the higher return is π_H, while the probability assigned to the lower return is $\pi_L = 1 - \pi_H$. Information is shared among parents, so that mother and father in the household share the same expectation for the rate of return to human capital investment. We can therefore define the expected utility derived by each parent from consumption and human capital investment by

$$EU_j(c, h) = u(c) + \pi_L \cdot v_j(r_L \cdot h) + \pi_H \cdot v_j(r_H \cdot h) \tag{5.2}$$

Following the literature on collective households, I assume that the decisions made by the household are Pareto-efficient (Chiappori, 1992). The household decision to allocate income to either consumption or human capital investment is therefore defined by the following maximisation problem:

$$\begin{aligned} \max_{c,h} \quad & (1-\lambda)\left[u(c) + E\left[v_f(r \cdot h)\right]\right] + \lambda\left[u(c) + E\left[v_m(r \cdot h)\right]\right] \\ \text{subject to} \quad & y \geq c + p \cdot h \\ & c \geq \bar{c} \end{aligned} \tag{5.3}$$

[9]The main reference is Blundell et al. (2005), who extend a general collective model with labour supply to allow for the presence of public goods expenditures.

where the Pareto weight $\lambda \in [0,1]$ reflects the weight of the mother in the household, d $\bar{c}$ is a minimum consumption which is necessary for the household before investing in human capital and p is the (relative) price for human capital investment. Since the program object of the study is a Conditional Cash Transfer, we would expect non-linearities in the budget set due to the presence of a conditionality linked to the reception of the educational subsidy. In fact, the introduction of a subsidy for all households would lowers the price of education p, but in the presence of a conditionality the reduction would only be faced by households whose children attend a number of classes lager than the program's minimum requirement (85 percent). However, baseline analysis of attendance showed that the conditionality imposed by the program was only partially binding; in fact, at baseline, only 8 percent of children attending school was attending less than 85 percent of classes and 60 percent was attending 95 percent. or more classes (Armand and Carneiro, 2013). Since the minimum attendance is generally non-binding in the data, modelling the programme as a decrease in the price of education services is therefore considered an adequate approximation.

With an interior solution**,** the maximisation problem lead to the following first order condition:

$$\begin{aligned}(1-\lambda)\left[\pi_L \cdot r_L \cdot v_f'(r_L \cdot h) + \pi_H \cdot r_H \cdot v_f'(r_H \cdot h)\right] + \\ \lambda\left[\pi_L \cdot r_L \cdot v_m'(r_L \cdot h) + \pi_H \cdot r_H \cdot v_m'(r_H \cdot h)\right] &= p \cdot u'\left(y - p \cdot h\right) \quad (5.4)\end{aligned}$$

If we define $\Phi_d = {}^{\left(\pi_L r_L \cdot v_d'(r_L \cdot h) + \pi_H r_H \cdot v_d'(r_H \cdot h)\right)}/_{u'(c)}$ as the marginal willingness to pay for the human capital investment for each parent, we can rewrite the optimality condition (5.4) as:

$$(1-\lambda) \cdot \Phi^f + \lambda \cdot \Phi^m = p \quad (5.5)$$

The efficiency condition for human capital investment takes the standard Bowen-Lindahl-Samuelson form for public good expenditures. Parents will invest in human capital up to the point in which the weighted sum of the (expected) marginal willingness to pay for human capital investment of father and mother is equal to the price of education.

How does the Macedonian CCT program relate to this setting? The targeting of mothers versus household heads changes parental relative income and therefore

provides an exogenous change in the Pareto weight λ. If we indicate w_f and w_m as the contribution to household income attributed to the mother and the father in the family, the CCT program generates an exogenous change in the relative income in the household, $^{w_f}/_{w_m}$. We are implicitly assuming that the direction of the derivative is positive in municipalities where the payments are made to mothers (since λ indicates the weight associated to mother's utility function) and negative in municipalities where the payments are made to household heads. We are therefore assuming that

$$\frac{\partial \lambda}{\partial \left(\frac{w_f}{w_m}\right)} > 0 \tag{5.6}$$

in the municipalities where the payments is targeted to mothers.

Since the CCT program provides an exogenous shift in the Pareto weight, we are therefore interested in understanding how such a change towards one household member or the other would affect the decision to invest on the child. Using Implicit Function theorem, we can derive the change in h induced by a change in λ[10]:

$$\frac{\partial h}{\partial \lambda} = \frac{\pi_L \cdot r_L \cdot \left(v'_m(r_L \cdot h) - v'_f(r_L \cdot h)\right) + \pi_H \cdot r_H \cdot \left(v'_m(r_H \cdot h) - v'_f(r_H \cdot h)\right)}{D} \tag{5.7}$$

where $D > 0$. Targeting women is beneficial for human capital investment if they have stronger preferences for child education, e.g. $\frac{\partial h}{\partial \lambda} > 0$ if $v'_f(h) < v'_m(h)$ for any h. If we observe a positive increase in investment in human capital in municipalities where the payments are made to women, we would expect that this is driven by a change in the Pareto weight induced by the program. Additionally, the model indicates that this would attributable to a different sensitivity of the marginal propensity to pay for child education compare with respect to consumption among decision makers. The size of the effect depends on preference differences among parents and on the cost/benefit of human capital investment (the relative price for education and the expected returns). The intuition behind this result is that when expected returns are small both parents have small incentives to invest. Once the return becomes larger the incentives to invest on education become stronger for the parent who has stronger preferences for human capital investment. When expected returns are sufficiently large both parents have strong

[10]A detailed derivation is provided in the 5.8.

incentives to invest in human capital. We would therefore expect to observe a differential effect on human capital investment only when subjective expectations are sufficiently large to compensate for reductions in consumption.

5.3 The Macedonian CCT for Secondary School Education

5.3.1 Background

The Republic of Macedonia is a country of roughly 2 million inhabitants at the centre of the Balkan region, in South-Eastern Europe. It is classified as an upper middle income country, registering in 2012 a GDP per capita of 4,568$ in current USD and 11,700$ in PPP[11]. From the education perspective, overall, the country is achieving good levels of primary school completion, with a gross (adjusted) enrolment rate in primary school equal to 98 percent in 2010[12]. At the same time, it is slightly under-performing on secondary school completion rates in comparison to the average among developing countries in the Europe and Central Asia (ECA) region. In 2010, Macedonia registered a gross secondary school enrolment of 90 percent and a net rate of 84 percent (compared to 92 and 85 percent in the ECA region). However, if we focus on the poorest share of the population, secondary school achievement is significantly lower, putting at risk the skill formation of the children with higher probabilities to face the consequences of child poverty in their lives. Net enrolment rates in secondary school age among Social Financial Assistance (SFA) recipients were 67% for girls and 65% for boys at the beginning of the school year 2009/2010, roughly 20 percentage points smaller than the country average.Figure 5.2 shows the share of children in SFA households who are either enrolled or have completed any secondary school program (two, three or four-year) by age of the child. This shows that among poor households, children tend to enter secondary school later (the curve has a positive slope from 15 to 16 years old) and the achievement rate for higher ages is considerably smaller, providing evidence of a high rate of drop-outs.

Before introducing the program, it is important to summarise the Macedonian education system (see Figure 5.3 for a graphical summary). Students access

[11]Based on The World Bank classification and databank.

[12]Source: United Nations Educational, Scientific, and Cultural Organisation (UNESCO) Institute for Statistics.

secondary school after the completion of primary education, which is offered from 6 to 15 years old on a three three-year cycles with classroom teaching in grades 1-5 and subject teaching in grades 6-9. Secondary education, which is compulsory and free of charge in public schools, is divided into specialised (languages and science-mathematics) preparatory schools offering four-year programs, general education and art secondary school offering three or four-year programs, and vocational education schools offer two-, three-, or four-year programs. At the end of preparatory schools and general secondary education, students sit the matura exam, but there are also final exams at the end of three- and four-year vocational education programs.

5.3.2 The program

The Macedonian "Conditional Cash Transfer (CCT) for Secondary School Education" is a social protection program aiming at increasing secondary school enrolment and completion rate among children in the poorest households of the population. It was first implemented by the Macedonian Ministry of Labour and Social Policy in Fall 2010 and provided cash transfers to poor households conditional on having children in school-age attending secondary school at least 85% of the time. In order to target poor households, the Ministry of Labour and Social Policy decided to offer the program to the beneficiaries of the Social Financial Assistance (SFA) benefit, which is the most significant income support program, accounting for around 0.5 percent of GDP and 50 percent of total spending on social assistance (Verme, 2008). SFA is a mean-tested monetary transfer granted to people who are fit for work, are socially not provided for and cannot support themselves. The amount paid for SFA is equivalent to the difference between household income and the social assistance amount determined for the household, depending on household size and time spent in SFA, varying from 1 825 MKD (around 40 USD) for one-member household to 4 500 MKD (around 98 USD) for households with 5 or more members. It is considered as the benefit of last resort, meaning it is provided after other benefits if the household income is still below a certain living standards threshold. It is mainly collected by households in the poorest tail of the income distribution; in 2009, the World Bank[13] reports that total SFA benefits are collected for 55 percent by the poorest quintile, 22 percent by the second poorest quintile and 11 percent to the middle quintile.

[13]I make reference to the "Project Appraisal Document - Report No: 47195-MK" between the Former Yugoslav Republic of Macedonia and the World Bank.

The total annual amount of the subsidy provided by the CCT if all conditions are met is 12 000 MKD (roughly 240 USD) to be paid in quarterly instalments. Cash transfers refer to the school quarters that constitute a school year, which follows the following division in quarters: from September to October, from November to December, from mid-January to March and from April to mid-June. CCT payments are made immediately after the school quarter is completed and data about attendance is checked. Therefore, the payments are scheduled at the following times: December, February, May and July.

The management of the CCT program was integrated within the social protection system and conditionality is controlled using a national software. Secondary schools enters attendance data at the end of each period and Social Welfare Centres (SWC), which are the administrative bodies managing payments for all benefits of financial assistance, issue the payment if the conditionality is met. Compliance with local guidelines governing the gender of the recipient is therefore easy to ensure, given that the full CCT management is computerised and the payments are processed depending on the family composition originally entered in the social protection system. The payment is processed via nominal cheques, which can be cashed in at banks or post offices[14].

5.3.3 Research design

The objective of the program was to increase the enrolment and attendance of children in secondary school age among poor households. In a first impact evaluation of the program, Armand and Carneiro (2013) compared the evolution in outcomes of SFA recipients from the pre-CCT to the post-CCT period to the evolution of outcomes for recipients of a similar type of Social Assistance and ineligible for CCT during the same period and found that after two years of its implementation the CCT had led to an increase in secondary school attendance among 15-19 years old children by roughly 6.5 percentage points. This is a substantial impact, given that average enrolment is close to 65% for this group. At the same time, they find that the effect of the program is driven only by enrolment, while the program had no effect on attendance. Was such an increase in enrolment equal among payment modalities or was there a beneficial effect of targeting mothers versus household head? Even if on average the program had

[14]Starting from the third year of the CCT, which is not considered in this chapter, payments have been processed using transactional accounts only, which allow a stronger enforcement of the payment modality.

no impact on attendance, would targeting mothers increase the outcome in that dimension too?

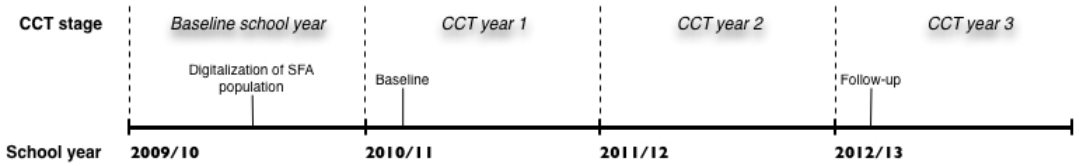
When first considering the implementation of a program to fight the low Macedonian enrolment rates among poor households, the government of Macedonia faced very specific design challenges and considered that gender-targeted transfers could have played a central role in educational choices since women empowerment is an important issue in the country. In 2011, the World Bank reported that the ratio of female-to-male labour force participation was 62 percent, smaller than the average of developing countries in the ECA region (68 percent). The ratio is presumably much more significant when considering the poorest share of the Macedonian population. For this reason, changing the identity of who controls the resources within the household was thought as having potentially dramatic consequences for household decision-making.

An experiment was then designed to test whether gender-targeted transfers could generate differential results. For this purpose, recipients of the cash transfer were randomised allowing payments to be received by either the the mother of the child or the household head, who is generally male[15]. The CCT program defines "Household Head" the person in the household that is registered at the Social Welfare Centre (SWC) for Social Financial Assistance. According to the rulebook for acquiring the right to Financial Assistance, the Household Head is determined by the following ordered rules: if there is an employed person in the household, the household head would be the employed person; if there is a pensioner, the household head would be the pensioner; if no employed person or pensioner exist in the household, the household head is the unemployed person representing the household; for all other households, the SWC selects the Household Head as the person representing the household.

Randomisation of the payment modality was done at municipality level using stratification by population size. The Republic of Macedonia is divided into 84 municipalities, which were first divided into 7 groups depending on population size and then randomised into two groups, one of which has 42 municipalities

[15]Looking at the Baseline data, among Social Financial Assistance recipients, in non-single parent households, in 90% of households the household head is the male partner (and father of children eligible for the CCT). In single parent households, the household head is the male partner only in 32% of households. Non-single parent households represent 88% of Social Financial Assistance households.

Figure 5.1: Program timescale and data collection



and where the payment of the transfer is done to the mother of the child, and the other which also has 42 municipalities and where the payment is transferred to the household head, regardless of gender. Panel A of Figure 5.4 presents the randomisation of treatment modalities across Macedonian municipalities.

5.4 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. Figure 5.1 shows the timescale of the program implementation and of the data collections.

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[16].

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[17]. 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%. 5.8 presents some robustness checks related to attrition and provides evidence that attrition at follow-up didn't changed significantly the composition of the sample for each treatment modality.

Table 5.2 presents the main characteristics at baseline of the household and the children in the sample, comparing the two different treatment groups. Households

[16]We aimed for a sample size of 17 households eligible for the CCT (recipients of social and financial assistance with children of secondary school age) per municipality, although in practice there was some variation in this number due to the fact that in some municipalities the eligible population was smaller then 17. For power calculations, we considered a power of 0.8 and a significance level of 0.05. With 42 clusters per arm and an inter-cluster correlation of 0.25, using 17 households per municipality it would be possible to detect a difference in expenditures in children's education (or in any other item) of 0.33 of a standard deviation and an increase in the proportion of students attending 85% or more of the classes of roughly 10% points.

[17]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.

are composed by 4.7 members and have on average 0.7 boys and 0.78 girls in the age category 13-18 years old. Household heads are male in 90 percent of households and have a low level of education with half having completed upper primary only and 20 percent having completed lower primary or not having a degree. Almost half of the sample lives in rural municipalities, while 14 percent of sampled households lives in the main urban area[18], which is the capital city of Skopje. If we look at living conditions, we can note that almost all households have access to a private toilet, but for only 45 percent the toilet is connected to sewerage and in only 76 percent of cases households have access to a connection to public water. If we turn our attentions at the structure of consumption, we can notice that households consume roughly 12 percent in education, compared to a 60 percent share in food. As an indicator of the disadvantaged situation of these households, we can underline that the consumption share for tobacco and alcohol is roughly 6 percent. In a comparison of the mothers and fathers, we can note that mothers are on average 42 years old and relatively younger than fathers, which are on average 45 years old. In addition, mothers are relatively less educated than fathers, with roughly 30 percent having achieved only lower primary or less, compared to a 20 percent for fathers. If we look instead at children, we can note that average age at baseline was 15.30 years old and among sampled children 57.4 percent was enrolled in secondary school and 34.1 percent was already enrolled in secondary school.

On most dimensions, the samples in municipalities where the payments are made to mothers and where the payments are made to household heads are balanced at baseline. While we find some unbalanced variables, normalised differences never exceeds 0.25, which is the limit suggested by Imbens and Wooldridge (2009) beyond which a linear specification is not appropriate. In order to control for such imbalances, in addition to individual and municipality controls, I include in any specification the baseline value of the dependent variable and the mean dependent variable at baseline for the correspondent age group.

[18]I use the Macedonian administrative definition of rural area as defined by the Law on Territorial Organisation of the Local Self-Government (OG 55/2004, 12/2005). According to the Law, municipalities are defined as "rural" if they have a seat in a village and "urban" if they have a seat in town. Towns are defined as "compactly built up residential areas with a population exceeding 3000, has a developed structure of various economic activities, above 51% of the employees are working in the secondary and tertiary sector, has an urban physiognomy of zones for residence, recreation and green area (parks), town square, street infrastructure, communal services and acts as a functional centre for the surrounding populated places". Villages are defined as "mono-functional populated areas, in which one business activity is prevalent and whereas the area has agricultural physiognomy and function".

5.4.1 Subjective expectations of schooling returns

To the purpose of collecting information on the heterogeneity in subjective expectations, during both baseline and follow-up data collection, a specific section was filled for each male and female youngest adolescent (aged 10 up to 17) in the household. This resulted in a total of 1455 children (750 boys and 705 girls) selected to answer the section, of which 136 didn't complete the section (characterising non-response rate of 9.35 percent), 111 where attrited at follow-up[19] and 120 were too young to be enrolled in secondary school at follow-up. Among those, for the purpose of the study, I select the children that are in secondary school age at the time of the follow-up (14-19 years old, including age 14 and 19 to allow for children to enter secondary school earlier or stay longer due to not passing a grade) and I restrict the sample to only households where both parents are present. These results in 920 children that will be object of the study.

A description of the module used to collect information on subjective expectations of schooling returns is provided in Section 3.2.1. In the chapter, in order to build expected income and variance, I assume a triangular distribution and I make use of only the first probability reported by the respondent. Chapter 3 analyses the differences between distributional assumptions and the choice of using different reported probabilities. All the results are robust to these assumptions.

The sample provides evidence that expected income and returns are greatly heterogeneous across individuals. Figure 5.5 reports the sample distribution of expected income for different levels of completed education, divided by gender. As we can notice both expected income after primary and secondary school are similar for boys and girls, but the main characteristics is that in the sample households have largely heterogeneous expectations. Table 5.3 presents the descriptive statistics for subjective expectations and a comparison among different treatment groups. I cannot identify any significant difference across groups, providing evidence that at baseline, expectations were balanced among different treatment modalities. Additionally, in relation to the complexity of this section of the questionnaire, a possible issue is the presence of missing values. Panel B of Table H1 reports the estimated difference in the probability of having a missing value for expectations at baseline. For each child in the household that was selected in order to collect subjective expectations, the dependent variable is a dummy

[19] Attrition didn't generate significant differences driven by the treatment modality. 5.8 discusses the checks related to attrition bias in detail.

variable equal to one if data is missing due to incomplete reporting or to refusal and is equal to zero if the data is complete. Columns 1-2 and 4-5 are estimated using OLS, while Column 3 and 6 are estimated using a Probit model. The coefficient is very close to zero, stable across specifications and never significant. The treatment doesn't seem to influence the chance of having a missing data in the expectation section, even when considering a extended (14-19 years old) versus a reduced sample (15-18 years old).

5.5 Empirical strategy

The evaluation design for the comparison of alternative CCT modalities allows examining differences in outcomes by comparing households living in municipalities with different payment modalities. Since the municipalities were allocated at random to different payment modalities, they should be identical (on average) on all their other characteristics, observed or unobserved. Therefore, a simple comparison across municipalities will give us the impact on enrolment δ_{im} of implementing one versus another payment modality. Let M_{im} be an indicator that takes value 1 if household i lives in municipality m where payments are done to the mother of the child and equal to 0 if payments are instead done to the head of household. In order to estimate the effect of different modalities on the enrolment I estimate the following Probit model:

$$\begin{aligned} \delta_{im,2012} &= \alpha + \sigma_M M_{im} + \beta_0 \cdot ExpRet_{im,2010} + \beta_2 \cdot ExpIncPrim_{im,2010} \\ &\quad + \sum_{j=1}^{2} \tau_j \cdot VarInc_{ijm,2010} + \qquad (5.8) \\ &\quad + \sum_{j=1}^{2} \gamma_j \cdot PrWork_{ijm,2010} + X'_{im}\gamma + \delta_{im,2010} + \epsilon_{im} \qquad (5.9) \end{aligned}$$

where $ExpRet_{im,2010}$ is the expected return to secondary school, $ExpIncPrim_{im,2010}$ is the expected income when completing only primary school, $VarInc_{ijm,2010}$ are the variances of income when completing educational level j, $PrWork_{ijm,2010}$ are the probabilities to be employed at age 25 when completing educational level j, X_{im} is a vector of individual, household and municipality characteristics and ε_{im} is a residual. Educational levels considered are $j = 1$ if the only primary school is completed and $j = 2$ if secondary school is completed. The impact on enrolment of paying the mother of the child as opposed to paying the head of household is given by σ_M. In order to control for potential imbalances in the outcomes of

interest at baseline, I estimate Equation 5.9 by including the observed value of the dependent variable at baseline, $\delta^*_{im,2010}$ in the model.

In order to check for heterogeneity in the impact of the payment modality in the subjective returns to schooling, I compute indicator variables identifying the quantile of the sample distribution to which the individual return belong to and I estimate the following model by interacting the indicator variable with the payment modality indicator:

$$\begin{aligned}\delta_{im,2012} &= \alpha + \sigma_{M,L} M_{im} \cdot D^L_{im,2010} + \sigma_{M,M} M_{im} \cdot D^M_{im,2010} + \sigma_{M,H} M_{im} \cdot D^H_{im,2010} + \\ &+ \eta_L \cdot D^L_{im,2010} + \eta_H \cdot D^H_{im,2010} + \beta_2 \cdot ExpInc_{ij,2010} + \\ &+ \sum_{j=1}^{2} \tau_j \cdot VarInc_{ijm,2010} + \sum_{j=1}^{2} \gamma_j \cdot PrWork_{ijm,2010} + \\ &+ X'_{im}\gamma + \delta_{im,2010} + \epsilon_{im} \end{aligned} \tag{5.10}$$

where $D^L_{im,2010}$, $D^M_{im,2010}$ and $D^H_{im,2010}$ are indicator variables for the return being in the lowest, middle and higher tercile.

A similar specification is used to analyse the effect heterogeneity in the subjective expectations of returns to schooling in terms of employment. In this case, I look at the interaction between the treatment indicator variable and the subjective gain in the probability of being employed after completing secondary school. This is defined as the difference between the subjective probability of being employed at age 25 after having completed secondary school and the subjective probability of being employed at age 25 after having completed primary school only, e.g. $PrWork_{i2m,2010} - PrWork_{i1m,2010}$.

5.6 Results

This section presents the results of the chapter. In all specifications I include controls for gender, age and education of mothers and fathers, ethnicity and religion of the household, household size, number of female and male children in age 14-19 (extended secondary school age) and age 6-13 (primary school age) and municipality controls (rural and capital city dummies). Year and semester of birth dummies and Regional dummies are included in all specifications. For

clarity, all tables related to the effect of paying mothers versus household heads omit all estimated coefficients for controls different than subjective expectations.

The large set of outcomes studied in the chapter raises concerns about multiple inference, i.e. the probability of erroneously rejecting at least one null hypothesis of no impact naturally increases with the number of outcomes considered. To deal with multiple inference, all significance levels are adjusted following Romano and Wolf (2005).

5.6.1 Resource ownership and schooling outcomes

In order to understand how targeting (conditional) cash transfers to mothers versus household heads lead to differential outcomes, Table 5.4 shows the estimates of the enrolment regressions specified by Equations 5.9 and 5.10. The dependent variable is equal to one if the child is either enrolled or has completed any secondary school. The model is estimated using a linear index Probit model and allows controlling for baseline average of the dependent variable.

We can note that in municipalities where payments were targeted to women, at the beginning of the school year 2012/2013 there is no significant difference in term of enrolment/achievement in secondary school when considering the sample as a whole. Since we are interested in analysing the heterogeneity in the effect on ex-ante expectations, Panel B presents the results for the interactions with the subjective returns to schooling, while Panel C presents the interaction with the return in the probability of being employed. If we look at the effect for children with different expected returns, we can note that targeting mothers provides a significant positive effect for the highest tercile of the distribution of returns to schooling, where the probability of being enrolled or having completed secondary school is 9.8 percent higher. If we look instead at the heterogeneity in the return in terms of employment, we find a significant positive effect for children in the middle tercile of the return distribution. For this children, targeting payments to mothers leads to an increase in the probability to enrol in secondary school of 10.5 percentage points. It is important to note that, when controlling for heterogeneity in ex-ante expectations, the coefficient in the lowest tercile is close to zero and not significant. This provides evidence that targeting mothers is beneficial, but only if parental perceived returns are sufficiently large. I don't find any significant difference of targeting mothers versus household head for

children whose pre-program expectations were low.

What mechanism is driving an increase in the probability to be enrolled in school when targeting women in households with high expected returns? In the next sub-sections, I will compare possible mechanisms through which targeting women can be beneficial for school achievement.

5.6.1.1 Expenditure shares

In order to understand the mechanism driving a larger school achievement in municipalities where the payments are made to women, I estimate the effect of the payment modality on individual expenditure shares[20]. Individual expenditure shares on education are defined as the ratio between monetary expenditure on education for the child and total household expenditure. Expenditure on education includes school fees, uniforms, school supplies, textbooks, additional courses and other expenses, transportation and meals at school. While up to secondary school, public education is free, cost such as transportation and living costs are still important to determine whether children go to school or drop-out.

Table 5.5 presents the estimates of a linear regression of individual shares on the payment modality indicator and its interactions with ex-ante returns to schooling and ex-ante returns in terms of employment. Results show that while we cannot identify a significant effect for the whole sample, in municipalities were payments are made to mothers, a significant difference in individual shares is found, but only for children whose ex-ante expectations presented larger returns to schooling, both in monetary and employment terms. For children in the highest tercile of the distribution of expected returns, targeting mothers increase individual shares by 3.8 percent compared to targeting household heads. Similarly, an increase of 3.9 percent in found for children in the highest tercile of the distribution of returns to schooling in terms of employment. This is consistent with the idea that targeting women would improve educational achievement by switching expenditures towards public goods, such as education, if women have stronger preferences for this good compared to men. These results are consistent with estimation using correction for attrition (see 5.8 for a discussion about attrition).

[20] **??** describes how expenditure data have been collected and how it is structured.

If we look at boys and girls differently, we can observe that this effect is mainly driven by higher monetary returns to schooling for girls and higher returns in terms of employment for boys. Table 5.6 provides estimates of the effect on individual expenditure shares by estimating the model separately for boys and girls. The results are consistent with the idea that boys and girls have different issues related to schooling and the entrance in the labour market. Results might be explained by the fact that girls tend to go to school more often if their parents perceive they can obtain a higher monetary return, while for boys targeting of resources is mainly driven by gains in the probability of being employed at age 25.

5.6.1.2 Monitoring and time use

The conditionality introduced by the CCT might interact with parental incentives related to child attendance. Parental behaviour might change in response to who is receiving the CCT transfer if parents increase their monitoring of schooling activities in a different way. In other words, are payment delivered to mothers providing better outcome in terms of secondary school achievement because mothers tend to monitor better their children and control whether they are attending to school[21]?

To this purpose, we collected at baseline and at follow-up information on the frequency in which parents talk to children about school. We collected this information for the youngest adolescent enrolled in primary school and for the youngest adolescent enrolled in secondary school during the year previous to the interview (the two years for the follow-up). Table 5.7 presents the estimates of the effect of payment modalities on the probability for the parents to talk to children about school on a daily basis[22]. Results show no effect on parental monitoring. This is consistent with the findings on the overall impact of the CCT

[21] Additionally, I find no difference among treatment modalities related to whether parents are more informed or have a better knowledge of the program where payments are made to the mother versus the household head. I don't test this hypothesis jointly with the others. In order to answer this question, we asked the respondent to answer some specific questions about the program characteristics, and specifically whether they heard about the program name, whether they know the conditionality, whether they are aware of the total amount of the CCT transfer, whether the know which groups are eligible, which school level is targeted and whether they know how many instalments are paid. I cannot identify a precise pattern of difference across the two groups, providing evidence that the program modality didn't generate substantial differences in the way people understand and know about the program.

[22] Similar results are obtained if we control for the probability to talk to children monthly, yearly or never.

on attendance, showing a zero impact and providing evidence that monitoring of school attendance might not be central for the Macedonian case (Armand and Carneiro, 2013). In fact, at baseline only 8 percent of children attending school was attending less than 85 percent of classes and 60 percent was attending 95 percent or more classes.

Another mechanism through which parents might invest differentially on children is through time spent with children. To this purpose, we collected information on the amount of time spent by both parents the day before the interview on different activities. Table 5.8 presents the results for the total time (expressed as share of the day) spent by both mothers and fathers with their children[23]. Results show very little effect on the way parents allocate their time in municipalities with different payment modalities. This provides additional evidence that targeting mothers do not change significantly the way parents monitor the schooling or the way parents spend time with their children. This might due to the fact that the program is targeting children older than 15 years old, an age category in which human capital investment through time spent with them might not be relevant.

5.7 Robustness checks

5.7.1 Expectations and enrolment in Secondary School

The first question we need to answer in this section is whether subjective expectations do correlate with schooling outcomes or in other words we need to control whether subjective returns to schooling matter in explaining education demand. Table 5.9 presents the coefficients on subjective expectations for the model (5.9). 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. 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 after completion of 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 and secondary school, shows that both

[23]Similar results are found if we analyse the share of the day spent helping children studying or on leisure activities with children.

variables explain secondary school enrolment. If parents expected a high probability of employment after primary school, the probability of having completed or being enrolled in secondary school two years after is lower, while the opposite is true for expectations of employment after secondary school. This result is consistent with the recent literature (Jensen, 2010; Attanasio and Kaufmann, 2009) which provides evidence that perceived returns are important to explain how individuals take educational choices.

5.7.2 Endogeneity of reported expectations

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. 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. One possible solution proposed by Filmer and Pritchett (2001) is to use principal-component approach and information collected on assets owned by the household to compute an asset index proxying wealth[24]. I make use of the rich information about household asset ownership collected at baseline to build a pre-program wealth index and divide households into three groups depending on the percentile position in distribution of the index.

While Column (1) in Table 5.10 present estimates of the subjective expectations coefficients in model 5.9 controlling only for individual and municipality characteristics, Column (2) reports the same estimates by controlling for household pre-program wealth. Results suggest that controlling for household wealth doesn't affect significantly the coefficients on expectations. This is supported by

[24]See 5.8 for details on how the index was built and for robustness checks.

the fact that expected returns are equally distributed across wealth groups (see Figure 5.6). In other words, children in households with low wealth tend not to enrol in secondary school, but among those the ones associated with higher returns have higher probability to go to school. This result is consistent with using expenditure data to rank households, rather than asset information.

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. 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. Secondary schools are built in main towns and cities in the largest municipalities, therefore for children living in smaller villages, the accessibility to a secondary school might be the main reason for early drop-outs and, perhaps, for low expected returns.

For this reason, I compute distance from secondary school as a proxy to capture costs associated with transportation and living costs. Figure 5.4 presents the geographical distribution of secondary schools (distinguishing among schools offering only courses in Macedonian language and school offering course in Albanian and Turkish) and of sampled households, along with the road network. 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. In addition, in order to check for the robustness of the measure, I perform the same method using the closest school providing a program taught in the same ethnic language of the household and by the type of the program offered[25]. Results are summarised in Table 5.1.

In order to check for school quality, information on the main characteristics of the schools were collected. In particular, information was collected about the number

[25] I differentiate schools offering preparatory high school programs versus any other programs.

Table 5.1: Distance from secondary school

Distance from	**Mean**	**Standard deviation**
...closest school	9.610	(9.615)
...closest school of same ethnicity	13.277	(16.179)
...closest preparatory high school	10.260	(9.657)

Note. Distances are reported in kilometres. Closest distance is computed as the minimum distance from the dwelling of the household to the closest available school using the available road network.

of students, the number of classes, the number of teachers and a series of supply-side indicators related to the school building (for example, the number of toilets available or whether the school has a gym) and to the learning offer (for example, the number of computers available or whether the school provides classes in a science lab). To control for local school quality, I construct the teacher-to-student ratio and the suspension rate for violent behaviour in the closest school to the household. Column (3) in Table 5.10 present estimates of the subjective expectations coefficients in model 5.9 adding controls for distance to the closest school and for the teacher-to-student ratio in the closest school. The coefficients are robust even after controlling for these measures, providing evidence that measures related to direct costs associated to schooling and to school quality do not affect reported expectations.

Another concern related to using subjective expectations in schooling models is that reported returns might be correlated with unobserved taste heterogeneity. In this case, in order to test for the robustness of the estimates, I check whether subjective expectations encompass monetary returns related to completion of primary or secondary school that are not directly related to the increase in productivity associated to schooling. For instance, reported returns to secondary school might include higher returns in the marriage market (Attanasio and Kaufmann, 2009) or might be correlated with returns in the crime market. Both are outside options that could have important consequences on schooling decisions and, in both cases, reported expectations might be different from the monetary returns linked to an increase in human capital. For this purpose, I construct measures of the marriage market and the crime market to control for this relationship. In Column (4) of Table 5.10 a Probit model for the enrolment decision is estimated by controlling for a marriage market indicator. I use male and female population at municipality level to build sex ratios in the age group 10-24 years old. In Column (5) I additionally control for the local characteristics of the juvenile crime market. I control for the local variation in the number of

convictions and reported crimes for juvenile perpetrators (younger than 18 years old) in the period before the data collection (2007-2010)[26] and by normalising it by the municipality population. In both cases, we can observe that controlling for measures related to the marriage and the crime market do no affect the estimates significantly.

5.7.3 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:

$$E[Y|E_i = 1, J = j] > E^*[Y|E_i = 1, J = j] \tag{5.11}$$

Using subjective expectations affected by cognitive dissonance in choice models would therefore upward bias our estimates. 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. Section 4.5.3 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. This provides evidence that cognitive dissonance bias is not an issue.

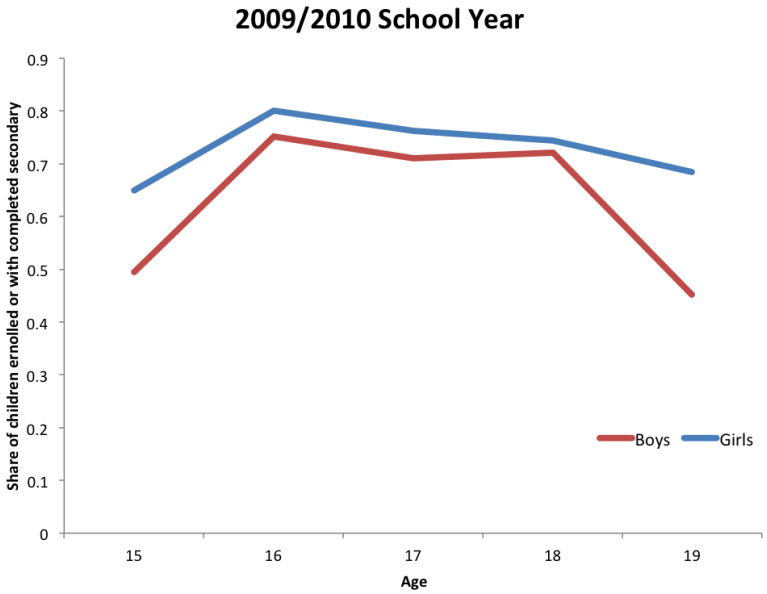
[26]I use the Macedonian State Statistical Office definitions for adult and juvenile perpetrator of a crime. Adult perpetrator of crime is a person who has committed crime and who at the time of committing the crime had reached 18 years of age, and has committed the crime as: executor, accomplice, initiator or assistant. Convicted person is an adult person recognised as responsible, against whom penal measures have been imposed. Juvenile perpetrator of crime is a perpetrator of crime who at the time of the execution of the crime had reached the age of 14, but not yet the age of 18 and has preformed the crime as: executor, accomplice, initiator or assistant. Reported juvenile is a juvenile against whom the legal procedure after the filed charges was not raised (the charge was rejected), against whom the proceeding has been stopped or a proposal has been applied for announcing a penalty or educational measure. Convicted person is a juvenile perpetrator of crime against whom with a Court decision a legal sanction has been pronounced-juvenile imprisonment or educational measures.

5.8 Conclusion

Decisions about human capital investment for children are particularly important in developing countries, where there is tendency to underinvest due to market failures related to information, such as incomplete or asymmetric information. At the same time there is still no consensus about the decision processes of households, which are central in determining human capital investments for the children. It is therefore important to study how intra-household resource control and subjective expectations for the returns to schooling interact to determine human capital investment decisions. In this chapter, in order to identify the causal effect of a change in intra-household resource ownership on household decisions, I made use of a randomised experiment linked to a secondary school conditional cash transfer in the Republic of Macedonia and of information on subjective returns to schooling. The conditional cash transfer provides an exogenous shock to intra-household resource ownership by targeting payments to mothers versus household heads. The chapter provides evidence that targeting cash transfers to mothers has a beneficial effect, but only for households were ex-ante expectations presented higher returns to schooling, both in monetary and employment terms. I provide evidence that for these children, individual expenditure shares for education are higher where payments are targeted to mothers.

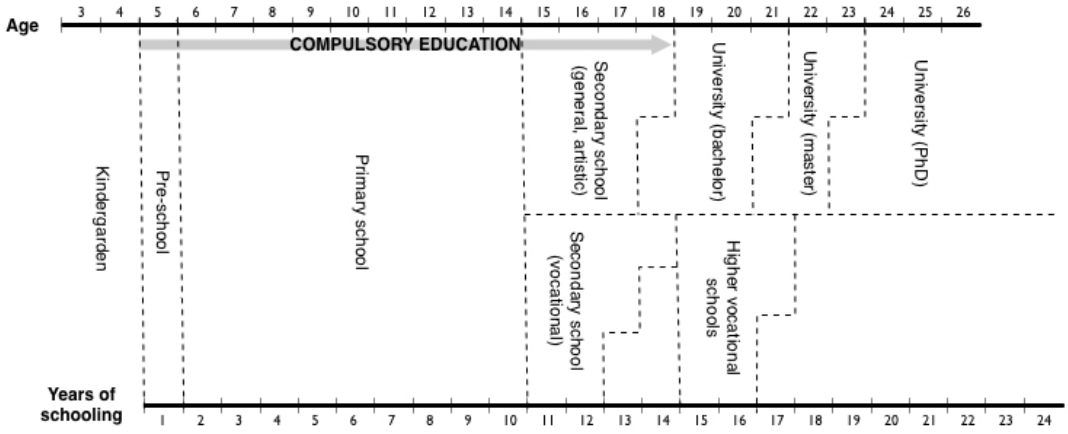
These findings suggest that in order to understand the role of each member in familial interactions is key to clarify how these relates to subjective expectations. This is particularly important in developing countries since perceived returns to education are often are below the market returns. Additionally, as I showed in the chapter, individuals have largely heterogeneous expectations related to education and the perceived returns do correlate with future choices and outcomes. If shifts in parental expectations have indirect impacts on the decision to invest on human capital, then much work need to be done in order to understand how subjective expectations form, how they evolve over time and how they interact with individual and collective choices.

Figure 5.2: Enrollment curve among Social Financial Assistance recipients (15-19 years old)



Note: The graph represents the share of children which are enrolled in 2, 3 or 4-year secondary school programs among Social Financial Assistance households. This is based on author's calculation using the baseline household survey and using sampling weights based on Social Financial Assistance population updated at summer 2010.

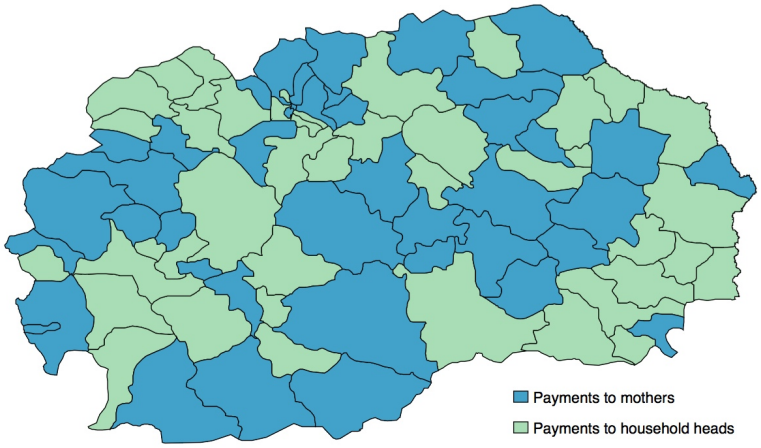
Figure 5.3: Educational system in Macedonia



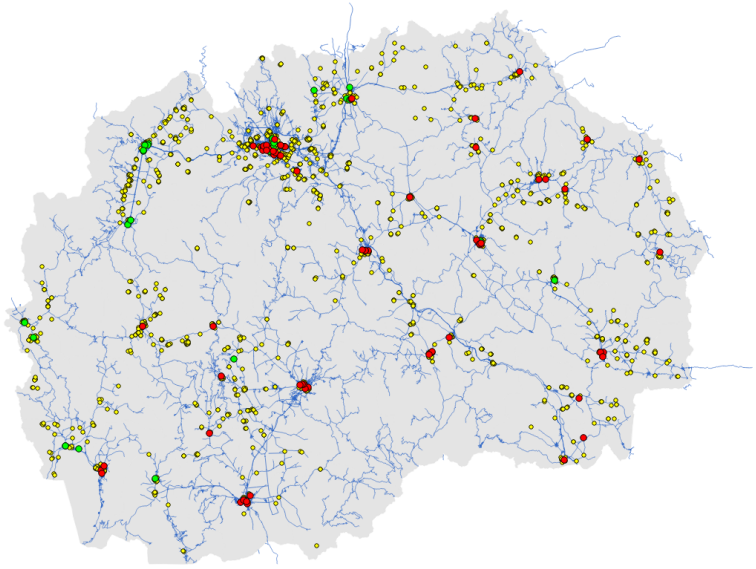
Note: Conditional Cash Transfer is targeting children who are in Social Financial Assistance families and who haven't completed secondary school up to age 23. Access to secondary school is provided upon completion of primary school. Access to university is possible only after completion of general and artistic secondary school. We don't consider here religious education.

Figure 5.4: Randomisation of treatment across municipalities and distribution of secondary schools and sampled households

Panel A. Randomisation of treatment across municipalities

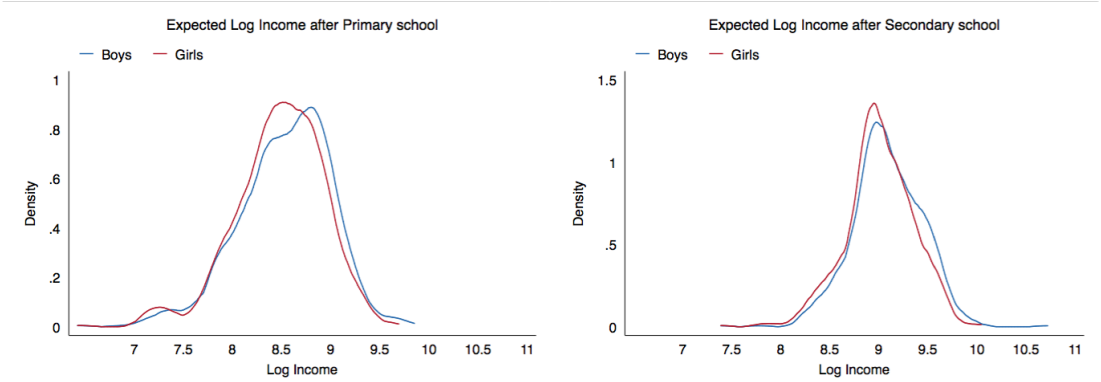


Panel B. Secondary schools and sampled households



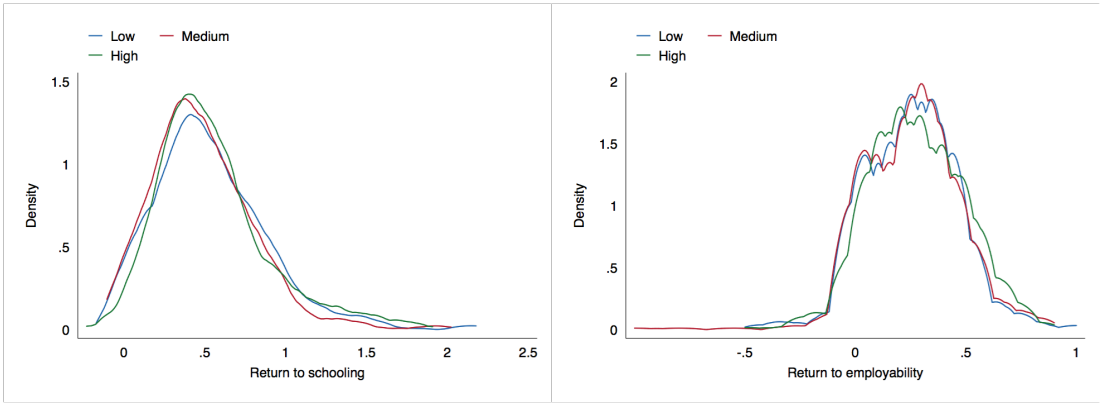
Note. The Macedonian administrative division includes 8 regions and 84 municipalities. Randomisation of the payment modality was done at municipality level and using stratification by population size, dividing municipalities into 7 groups. Households and secondary schools presented in the map are computed using geo-coding data collected at follow-up. Red dots represent secondary schools providing educational programs only in Macedonian language, while green dots show secondary schools providing programs in Albanian or Turkish (in addition to Macedonian). In blue, the main and secondary road network.

Figure 5.5: Sample distribution of expected income after primary and secondary school by gender



Note. Left panel shows the sample distribution of the expected income after primary school, while the panel on the right shows the expected income after secondary school. Income is reported in logarithms and expected income is computed using triangular distribution.

Figure 5.6: Expected return to education by asset group



Note. Left Panel 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). Right Panel 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. Asset groups (low, medium and high) are computed using principal component analysis on asset ownership at baseline and by dividing the distribution of the asset index derive for the first principal component into three terciles.

Table 5.2: Baseline descriptive statistics, by treatment status

	All	Household head	Mother	Diff.	Norm. diff.
	(1)	(2)	(3)	(4)	(5)
Household characteristics					
Household members	4.718	4.701	4.734	0.0334	0.0209
	[1.127]	[1.151]	[1.103]	(0.125)	
Mother characteristics					
Age	42.06	42.01	42.11	0.100	0.0125
	[5.689]	[5.873]	[5.502]	(0.487)	
Schooling (years)	7.105	7.117	7.092	-0.0252	-0.005
	[3.257]	[3.131]	[3.383]	(0.331)	
Father characteristics					
Age	45.43	45.24	45.64	0.399	0.0532
	[5.302]	[5.350]	[5.253]	(0.469)	
Schooling (years)	8.047	7.923	8.173	0.251	0.060
	[2.933]	[3.001]	[2.860]	(0.297)	
Expenditures					
P.c. monthly expenditure	8136.4	8150.4	8122.0	-28.45	-0.003
	[7888.4]	[7832.0]	[7957.2]	(797.4)	
Education share (girls)	0.0616	0.0613	0.0619	0.001	0.003
	[0.159]	[0.156]	[0.163]	(0.016)	
Education share (boys)	0.0599	0.0592	0.0607	0.00145	0.001
	[0.148]	[0.146]	[0.149]	(0.015)	
Food share	0.485	0.479	0.491	0.0126	0.0501
	[0.178]	[0.179]	[0.177]	(0.019)	
Child characteristics					
Age	15.30	15.32	15.28	-0.041	-0.0174
	[1.645]	[1.585]	[1.704]	(0.096)	
Male	0.531	0.531	0.531	0.001	0.001
	[0.499]	[0.500]	[0.500]	(0.024)	
Enrolled in primary school	0.574	0.587	0.561	-0.0260	-0.0372
	[0.495]	[0.493]	[0.497]	(0.036)	
Enrolled in secondary school	0.341	0.321	0.361	0.0403	0.0600
	[0.474]	[0.467]	[0.481]	(0.030)	
Ind. exp. share on education	0.0390	0.0410	0.0371	-0.004	-0.0279
	[0.0976]	[0.0999]	[0.0954]	(0.009)	

Note. Standard deviations in brackets, standard errors in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. Difference in Column (4) is computed as (3)-(2). The standard errors on the differences are estimated from running the corresponding least squares regression allowing for the errors to be clustered at municipality level and controlling for strata dummies. The normalised difference is computed following Imbens and Wooldridge (2009) and allowing clustering at municipality level. Per capita monthly expenditure is expressed in Macedonian Denars. Individual expenditure share on education is computed as expenditure on education for each child divided by total household expenditure.

Table 5.3: Baseline descriptive statistics of Expectations, by treatment status

	All	Household head	Mother	Diff.	Norm. diff.
	(1)	(2)	(3)	(4)	(5)
Primary school expectations					
Lower bound	8.197	8.209	8.185	-0.0249	-0.0328
	[0.536]	[0.528]	[0.544]	(0.0625)	
Upper bound	8.822	8.842	8.802	-0.0405	-0.0693
	[0.412]	[0.388]	[0.434]	(0.0507)	
Expected income	8.532	8.548	8.516	-0.0321	-0.0506
	[0.448]	[0.430]	[0.466]	(0.0531)	
Variance income	0.0222	0.0227	0.0216	-0.00106	-0.0252
	[0.0297]	[0.0308]	[0.0287]	(0.00373)	
Prob. to find a job	0.216	0.209	0.223	0.0147	0.0544
	[0.191]	[0.174]	[0.206]	(0.0252)	
Secondary school expectations					
Lower bound	8.784	8.783	8.785	0.00260	0.00478
	[0.385]	[0.371]	[0.399]	(0.0529)	
Upper bound	9.301	9.320	9.283	-0.0376	-0.0746
	[0.356]	[0.335]	[0.376]	(0.0526)	
Expected income	9.060	9.070	9.050	-0.0199	-0.0410
	[0.343]	[0.318]	[0.367]	(0.0495)	
Variance income	0.0150	0.0162	0.0139	-0.00231	-0.0796
	[0.0205]	[0.0224]	[0.0183]	(0.00249)	
Prob. to find a job	0.481	0.492	0.471	-0.0219	-0.0724
	[0.214]	[0.225]	[0.202]	(0.0262)	
Return to secondary school	0.528	0.522	0.534	0.0122	0.0250
	[0.344]	[0.338]	[0.351]	(0.0458)	

Note. Standard deviations in brackets, standard errors in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. Difference in Column (4) is computed as (3)-(2). The standard errors on the differences are estimated from running the corresponding least squares regression allowing for the errors to be clustered at municipality level and controlling for strata dummies. The normalised difference is computed following Imbens and Wooldridge (2009) and allowing clustering at municipality level. Returns to secondary school are computed assuming a triangular distribution.

Table 5.4: Enrolment regression and interaction with returns to schooling

Dep.var.: Enrolled or completed secondary school			
	Probit (1)	Probit (2)	Probit (3)
A. No interaction			
Payment to mother	0.061		
	(0.032)		
B. Interaction with return to schooling			
Payment to mother * Return (1st tercile)		0.017	
		(0.046)	
Payment to mother * Return (2nd tercile)		0.046	
		(0.056)	
Payment to mother * Return (3rd tercile)		0.098**	
		(0.035)	
C. Interaction with return to schooling in terms of employment			
Payment to mother * Return (1st tercile)			-0.006
			(0.055)
Payment to mother * Return (2nd tercile)			0.105**
			(0.033)
Payment to mother * Return (3rd tercile)			0.066
			(0.047)
Observations	920	920	920

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. Significance level adjusted for multiple hypothesis testing (Romano and Wolf, 2005). In Panel A, I estimate Equation 5.9, while in Panels B and C I estimate Equation 5.10 using interactions with returns to schooling in monetary terms and in terms of employment. 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. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included.

Table 5.5: Effect on Individual expenditure shares on Education

Dep.var.: Ind. expenditure share on education			
	OLS (1)	OLS (2)	OLS (3)
A. No interaction			
Payment to mother	0.012		
	(0.011)		
B. Interaction with return to schooling			
Payment to mother * Return (1st tercile)		-0.011	
		(0.018)	
Payment to mother * Return (2nd tercile)		0.015	
		(0.018)	
Payment to mother * Return (3rd tercile)		0.038*	
		(0.014)	
C. Interaction with return to schooling in terms of employment			
Payment to mother * Return (1st tercile)			-0.016
			(0.016)
Payment to mother * Return (2nd tercile)			0.019
			(0.014)
Payment to mother * Return (3rd tercile)			0.039*
			(0.015)
Observations	911	911	911

Note. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. Significance level adjusted for multiple hypothesis testing (Romano and Wolf, 2005). In Panel A, I estimate Equation 5.9, while in Panels B and C I estimate Equation 5.10 using interactions with returns to schooling in monetary terms and in terms of employment. The dependent variable is computed as the ratio between the education expenditure on the children and the total expenditure of the household. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included. Return in terms of employment is defined as the difference between the probability of being employed after secondary school and the probability for being employed after primary school at age 25.

Table 5.6: Effect on Individual expenditure shares on Education, by gender

	Dep.var.: Individual expenditure share on education					
		Girls			Boys	
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
A. No interaction						
Payment to mother	0.004			0.020		
	(0.012)			(0.016)		
B. Interaction with return to schooling						
Payment to mother * Return (1st tercile)		-0.011			-0.012	
		(0.019)			(0.025)	
Payment to mother * Return (2nd tercile)		-0.024			0.060	
		(0.018)			(0.030)	
Payment to mother * Return (3rd tercile)		0.046*			0.029	
		(0.018)			(0.021)	
C. Interaction with return to schooling in terms of employment						
Payment to mother * Return (1st tercile)			-0.003			-0.025
			(0.014)			(0.022)
Payment to mother * Return (2nd tercile)			-0.008			0.038
			(0.018)			(0.018)
Payment to mother * Return (3rd tercile)			0.023			0.057*
			(0.020)			(0.024)
Observations	426	426	426	485	485	485

Note. Standard errors clustered at municipality level in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. Significance level adjusted for multiple hypothesis testing (Romano and Wolf, 2005). In Panel A, I estimate Equation 5.9, while in Panels B and C I estimate Equation 5.10 using interactions with returns to schooling in monetary terms and in terms of employment. The dependent variable is computed as the ratio between the education expenditure on the children and the total expenditure of the household. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included. Return in terms of employment is defined as the difference between the probability of being employed after secondary school and the probability for being employed after primary school at age 25.

147

Table 5.7: Effect on the Probability to talk to children about school daily

Dep.var.: Talked to children about school (daily)			
	Probit (1)	Probit (2)	Probit (3)
A. No interaction			
Payment to mother	-0.006		
	(0.017)		
B. Interaction with return to schooling			
Payment to mother * Return (1st tercile)		-0.019	
		(0.039)	
Payment to mother * Return (2nd tercile)		-0.023	
		(0.025)	
Payment to mother * Return (3rd tercile)		0.016	
		(0.016)	
C. Interaction with return to schooling in terms of employment			
Payment to mother * Return (1st tercile)			-0.030
			(0.028)
Payment to mother * Return (2nd tercile)			-0.012
			(0.025)
Payment to mother * Return (3rd tercile)			0.020
			(0.025)
Observations	663	663	663

Note. Standard errors clustered at municipality level in parenthesis. In Panel A, I estimate Equation 5.9, while in Panels B and C I estimate Equation 5.10 using interactions with returns to schooling in monetary terms and in terms of employment. The dependent variable is computed as the ratio between the education expenditure on the children and the total expenditure of the household. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included. Return in terms of employment is defined as the difference between the probability of being employed after secondary school and the probability for being employed after primary school at age 25.

Table 5.8: Time use: share of the day spent with children

	Dep.var.: Share of the day spent with children					
	Father	Mother	Father	Mother	Father	Mother
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
A. No interaction						
Payment to mother	0.011	0.001				
	(0.013)	(0.013)				
B. Interaction with return to schooling						
Payment to mother * Return (1st tercile)			0.016	0.014		
			(0.022)	(0.019)		
Payment to mother * Return (2nd tercile)			0.012	-0.010		
			(0.019)	(0.015)		
Payment to mother * Return (3rd tercile)			0.004	-0.002		
			(0.019)	(0.019)		
C. Interaction with return to schooling in terms of employment						
Payment to mother * Return (1st tercile)					0.031	0.015
					(0.019)	(0.018)
Payment to mother * Return (2nd tercile)					-0.008	-0.008
					(0.018)	(0.016)
Payment to mother * Return (3rd tercile)					-0.005	-0.003
					(0.018)	(0.016)
Observations	743	743	743	743	743	743

Note. Standard errors clustered at municipality level in parenthesis. In Panel A, I estimate Equation 5.9, while in Panels B and C I estimate Equation 5.10 using interactions with returns to schooling in monetary terms and in terms of employment. The dependent variable is computed as the share of the day spent on the activity. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included. Return in terms of employment is defined as the difference between the probability of being employed after secondary school and the probability for being employed after primary school at age 25.

149

Table 5.9: Correlates of Secondary School enrolment at follow-up

Dep.var.: Enrolled or completed Sec. school			
	Probit (1)	Probit (2)	Probit (3)
Expected income (primary)	-0.055	-0.065	-0.017
	(0.048)	(0.054)	(0.058)
Expected income (secondary)	0.186***	0.192***	0.137*
	(0.068)	(0.069)	(0.078)
Variance of income (primary)		-0.315	-0.324
		(0.542)	(0.547)
Variance of income (secondary)		0.275	0.478
		(0.738)	(0.742)
Probability of employment (primary)			-0.260**
			(0.115)
Probability of employment (secondary)			0.281***
			(0.107)
Individual controls	✓	✓	✓
Municipality controls	✓	✓	✓
Observations	920	920	920

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. 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. Distance from School is computed as distance from the closest school providing educational program in the same ethnic language of the individual. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included.

Table 5.10: Secondary school enrolment and subjective expectations: robustness checks

	Dep.var.: Enrolled or completed Secondary school				
	(1) Probit	(2) Probit	(3) Probit	(4) Probit	(5) Probit
Expected income (primary)	-0.024	-0.014	-0.027	-0.023	-0.023
	(0.057)	(0.058)	(0.057)	(0.057)	(0.058)
Expected income (secondary)	0.152*	0.141*	0.144*	0.137*	0.137*
	(0.079)	(0.078)	(0.075)	(0.075)	(0.076)
Variance income (primary)	-0.583	-0.287	-0.336	-0.347	-0.349
	(0.556)	(0.551)	(0.524)	(0.520)	(0.527)
Variance income (secondary)	0.332	0.474	0.690	0.665	0.668
	(0.773)	(0.746)	(0.733)	(0.731)	(0.732)
Probability of employment (primary)	-0.262**	-0.264**	-0.251**	-0.259**	-0.259**
	(0.120)	(0.116)	(0.115)	(0.114)	(0.113)
Probability of employment (secondary)	0.277***	0.280***	0.268***	0.280***	0.280***
	(0.105)	(0.106)	(0.103)	(0.101)	(0.101)
Individual and municipality controls	✓	✓	✓	✓	✓
Wealth controls		✓	✓	✓	✓
Distance from school and quality			✓	✓	✓
Sex ratio				✓	✓
Crime market controls					✓
Observations	920	920	920	920	920

Note. Marginal effects. Standard errors clustered at municipality level in parenthesis. 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. Distance from School is computed as distance from the closest school providing educational program in the same ethnic language of the individual. I include controls for gender, age and education of mother and father, ethnicity, religion, household size, number of female and male children in age 14-19 and age 6-13, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included.

151

Table 5.11: Kolmogorov-Smirnov test for equality of the distribution if changes in subjective probabilities from baseline to follow-up

Smaller Group and K-S combined	Returns to schooling		Probability of employment after... primary school		secondary school		Returns in terms of employment	
	Difference (1)	p-value (2)	Difference (3)	p-value (4)	Difference (5)	p-value (6)	Difference (7)	p-value (8)
Compare primary-primary and primary-secondary groups								
- Primary-Primary	0.042	0.626	0.052	0.459	0.088	0.111	0.068	0.268
- Primary-Secondary	-0.060	0.380	-0.017	0.922	-0.013	0.955	-0.024	0.852
- Combined K-S	0.060	0.688	0.052	0.805	0.088	0.205	0.068	0.497
Compare primary-primary and secondary-secondary groups								
- Primary-Primary	0.027	0.731	0.066	0.148	0.075	0.083	0.063	0.171
- Secondary-Secondary	-0.045	0.424	-0.019	0.856	-0.008	0.972	-0.035	0.587
- Combined K-S	0.045	0.760	0.066	0.279	0.075	0.156	0.063	0.323
Compare primary-secondary and secondary-secondary groups								
- Primary-Secondary	0.056	0.370	0.050	0.437	0.046	0.494	0.077	0.141
- Secondary-Secondary	-0.022	0.857	-0.022	0.856	-0.023	0.842	-0.031	0.732
- Combined K-S	0.056	0.676	0.050	0.778	0.046	0.851	0.077	0.265
Observations	383 / 236 / 562							

Note. Kolmogorov-Smirnov test is performed on the difference between the expectation at follow-up and at baseline. The group "Primary-primary" is composed by children whose maximum educational level achieved at baseline and follow-up is primary school reported for the same child. The group "Primary-secondary" is composed by children whose maximum educational level achieved at baseline was primary school, while at follow-up was secondary school. The group "Secondary-secondary" is composed by children whose maximum educational level achieved at baseline and follow-up was higher than primary school. P-values on combined Kolmogorov-Smirnov test are the exact p-values. For each comparison, the first line test the hypothesis that the values for the first group are smaller than the values for the second group, while the second line tests whether the values of the first group are larger than the values for the second group. Column Difference reports the largest difference between the groups. Observations are reported as number of children in each category: Primary-primary, Primary-Secondary, Secondary-secondary. Delta probability is defined as the difference between the probability of being employed at age 25 after secondary school and after primary school.

152

Appendix 5.A
A collective model with human capital investment

The household decision to allocate income to either consumption (c) or human capital investment (h) is defined by the following maximisation problem:

$$\begin{aligned} \max_{c,h} \quad & (1-\lambda)\left[u(c)+E\left[v_f(r\cdot h)\right]\right]+\lambda\left[u(c)+E\left[v_m(r\cdot h)\right]\right] \\ \text{subject to} \quad & y \geq c+p\cdot h \\ & c \geq \bar{c} \end{aligned} \tag{5.12}$$

where the Pareto weight $\lambda \in [0,1]$ reflects the weight of the mother in the household, p is the (relative) price for human capital investment and where $\bar{c}$ is a minimum (subsistence) consumption which is required for the household before investing in human capital, r is the rate of return of the investment in human capital which is unknown to parents. I assume that the return can have either a low value (r_L) or a high value (r_H) and that both parents share the same expectations, attributing a probability π_H to the higher return and a probability $(1-\pi_H)$ to the lower return.

With an interior solution, the problem lead to the following first order condition:

$$\begin{aligned} (1-\lambda)\left[(1-\pi_H)\,r_L\cdot v_f'(r_L\cdot h)+\pi_H r_H\cdot v_f'(r_H\cdot h)\right]+ & \\ \lambda\left[(1-\pi_H)\,r_L\cdot v_m'(r_L\cdot h)+\pi_H r_H\cdot v_m'(r_H\cdot h)\right] &= \\ = \quad p\cdot u'\left(y-p\,(5.13)\right) & \end{aligned}$$

If we define $\Phi_d = \left((1-\pi_H)r_L\cdot v_d'(r_L\cdot h)+\pi_H r_H\cdot v_d'(r_H\cdot h)\right)/u'(c)$ as the marginal willingness to pay for the human capital investment for each parent, we can rewrite the optimality condition (5.13) as:

$$(1-\lambda)\cdot\Phi^f+\lambda\cdot\Phi^m = p \tag{5.14}$$

Using Implicit Function theorem we can derive the derivative of human capital investment with respect to a change in the Pareto weight, $\frac{\partial h}{\partial \lambda}$. This is equal to

the following expression:

$$\frac{\partial h}{\partial \lambda} = \frac{(1-\pi_H)\cdot r_L \cdot \left(v'_m(r_L \cdot h) - v'_f(r_L \cdot h)\right) + \pi_H \cdot r_H \cdot \left(v'_m(r_H \cdot h) - v'_f(r_H \cdot h)\right)}{D} \tag{5.15}$$

where

$$\begin{aligned} D \; = - \;& (1-\lambda)\left(p^2 \cdot U''(y-hp) + \pi_H \cdot r_H^2 \cdot v''_f(r_H \cdot h) + (1-\pi_H)\cdot r_L^2 \cdot (v''_f(r_L \cdot h)\right) + \\ & -\lambda\left(p^2 \cdot U''(y-hp) + \pi_H \cdot r_H^2 \cdot v''_m(r_H \cdot h) + (1-\pi_H)\cdot r_L^2 \cdot v''_m(r_L \cdot h)\right) = \\ = - \;& p^2 \cdot U''(y-hp) - (1-\lambda)\left((1-\pi_H)\cdot r_L^2 \cdot v''_f(r_L \cdot h) + \pi_H \cdot r_H^2 \cdot v''_f(r_H \cdot h)\right) \\ & -\lambda\left((1-\pi_H)\cdot r_L^2 \cdot v''_m(r_L \cdot h) + \pi_H \cdot r_H^2 \cdot v''_m(r_H \cdot h)\right) \end{aligned}$$

Since D is positive, $\frac{\partial h}{\partial \lambda}$ is larger than zero if $v'_f(r \cdot h) < v'_m(r \cdot h)$ for any $h \times r$.

Appendix 5.B
Attrition and missing values

In this Appendix, I present some robustness checks related to sample bias related to attrition and missing expectation values. Panel A of Table H1 reports the estimated difference in attrition rate for SFA households across treatment modalities. The dependent variable is a dummy variable equal to one if the household was interviewed only at baseline and equal to zero if the household was interviewed at baseline and follow-up. In Column 1 I estimate the difference across treatment modalities by controlling only for regional dummies, while in Columns 2 and 3 I control for household characteristics. Columns 1 and 2 are estimated using OLS, while Column 3 is using a Probit model. The coefficient is roughly equal to 2 percent and stable across specifications, but is never significant. This provides evidence that the 2-year attrition is not explained by being resident in municipalities where the payments are made to mothers rather than the other municipalities. We can observe that attrition is mainly driven by Roma households, households living in Skopje and families where the father has upper primary education. I observe lower levels of attrition among Macedonian households. Among most of dimensions there are no significant statistical differences between attrited households and non-attrited households. Similar conclusion can be drawn when we look at child-level characteristics.

In order to control for robustness of the results to attrition, Table H2 present the estimates for the effect on individual expenditure shares comparing the unweighted results with the weighted results using inverse probability weighting. In the latter case weights are the inverse on the estimated probability of being interviewed at baseline and follow-up (see Wooldridge 2002). This method allows increasing the weight of observations with a higher attrition at follow-up. I generate weights using estimates from Column 3 in Table H1. The weights are generated using a Probit regression of an indicator variable being equal to 1 if the observation was interviewed at baseline, but was missing at follow-up and 0 otherwise on a series of observable individual, household and municipality-level characteristics. From the regression, I compute predicted probabilities and I computed weights by taking the inverse of one minus the predicted probabilities. Panel B of Table H1 reports the estimated difference in the probability of having a missing value for expectations at baseline. For each child in the household that was selected in order to collect subjective expectations, the dependent variable is a dummy variable equal to one if data is missing due to incomplete reporting or to refusal and is equal to zero if the data is complete. Columns 1 and 2 are estimated using OLS, while Column 3 is using a Probit model. The coefficient is very close to zero and stable across specifications, but is never significant. The treatment doesn't seem to influence the chance of having a missing data in the expectation section.

Appendix 5.C
Construction of the wealth index

In order to control for pre-program household wealth, I construct an asset index using factor analysis (Filmer and Pritchett, 2001) by exploiting the information available at baseline. Following the Filmer-Pritchett (FP) procedure, I compute asset ownership or access to resources by using indicator variables for whether the household own the good (or has access to a resource) and I compute indicator weights by using Principal Component Analysis (PCA). As suggested by McKenzie (2005), I make use of only the first factors produced by PCA to represent the wealth index and I consider a wide range of asset variables to avoid issues related to clumping and truncation[27]. The variables used to build the index are the

[27]Clumping occurs when the wealth index has limited variation (e.g. it groups households into a limited number of groups), while truncation occurs when there is limited variation in asset ownership.

following: durable goods property[28], access to utilities (public water, sewerage, electricity, phone line), type of dwelling and type of property, use of shared toilet, land property and livestock ownership. I group variables in which ownership is limited to very few households (smaller than 30 households in the sample). Figure H1 presents the distribution of the resulting wealth index. The first component explains roughly 11% of the total variation and the distribution has limited issues related to high skewness and kurtosis. In order to show internal coherence of the wealth index, I split the measure into three groups depending on the tercile of the distribution. I indicate these groups by low, middle and high wealth. Table H3 shows the share of households owning a specific asset and compares them by using the three wealth groups computed through PCA. We can observe that as the wealth quantile increases, households do own better assets and do have better access to utilities, suggesting that PCA methodology provides a credible method for grouping households into wealth groups.

Appendix 5.D
Testing for male versus female expectations

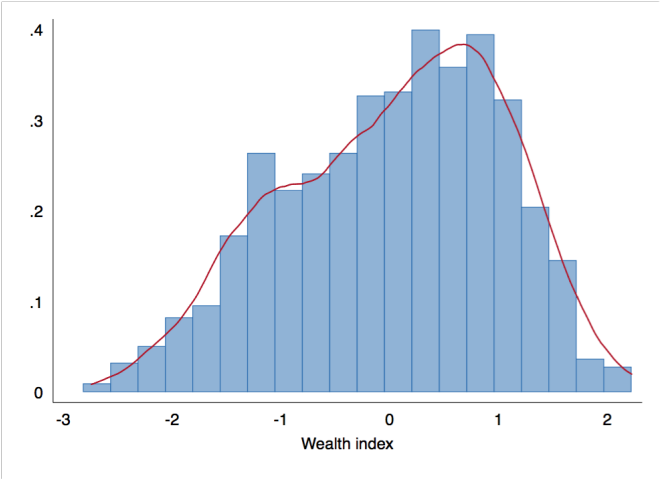
Throughout the paper we have assumed that parental expectations affect human capital investment, but we haven't been able to distinguish between male and female partner expectations. The reason is that most of the interviews are carried out with both partners present. However, in order to test for equality of expectations, I selected all households where both male and female partners were part of the family at baseline and I conditioned on having the female partner present during the interview. This allows comparing the answers for households where the female partner is alone and where both partners are present.

Table H4 presents descriptive statistics for households in which the male partner is absent during the interview and households in which the male partner is present. The interviewers were originally instructed to arrange interviews where both partners were present and, if not possible, with at least the presence of the household head. Therefore whether the male partner is a household head is a

[28]We collected information on several durable goods that the household could have owned. The variables used in the construction of the index are indicator variables equal to one if the household own at least one item of the good and zero otherwise. We collected information on the following items: cooker and stove (by type of fuel), boiler, refrigerator, washing machine, iron, sewing, vacuum cleaner, air conditioning, radio and tv, video recorder, personal computer, phone and mobile phone, musical instrument, bicycle, car and motorbike.

strong predictor of his presence during the interview. At the same time, a younger male partner, households living in Skopje and household with a relatively higher wealth have a higher probability to be present during the interview. To control for differences in reported expectations Table H5 presents the results of regressions on return to schooling, variance of income and probability for the child to be employed at age 25 on a dummy variable indicating whether the male partner is present at the interview. To account for observable differences among these groups, the difference is conditional on household and individual characteristics. It emerges that the presence of the male partner slightly increase the return to schooling, but no significant difference is recorded. Additionally controlling for observable characteristics doesn't have a strong impact on the coefficients. This is however a sub-optimal test, since the presence of the husband cannot account for unobservable differences linked to the presence of the male partner during the interview, but provides a first evidence on the equality of expectations across partners.

Figure H1: Wealth index distribution



Note: The wealth index is computed using Principal Component Analysis following Filmer and Pritchett (2001). The index is built using information on durable good ownership, access to utilities, type and property of the dwelling, land property and livestock ownership.

Table H1: Treatment modality, attrition and missing expectations

	15-19 years old			*15-18 years old*		
	OLS (1)	OLS (2)	Probit (3)	OLS (4)	OLS (5)	Probit (6)
A. Attrition						
Payment to mother	0.005	0.005	0.005	0.018	0.014	0.011
	(0.024)	(0.023)	(0.021)	(0.024)	(0.025)	(0.021)
B. Missing expectation at baseline						
Payment to mother	0.001	-0.003	-0.004	-0.018	-0.021	-0.023
	(0.021)	(0.020)	(0.016)	(0.022)	(0.022)	(0.018)
Individual characteristics		✓	✓		✓	✓
Regional dummies	✓	✓	✓	✓	✓	✓
Observations	1233	1233	1233	795	795	795

Note. Standard errors in parenthesis. *** denotes significance at 1%, ** at 5%, and * at 10%. In Panel A, the dependent variable is equal to 1 if the household has been interviewed at baseline, but not at follow-up and 0 if it has been interviewed in both waves. In Panel B, the dependent variable is equal to one if the child has been selected for the expectations section (younger male and female child in age category older than 10 and younger than 18) and the data is missing for incomplete reporting or refusal and is equal to 0 is the child is selected and the information is complete. Columns 1 and 4 are OLS regressions on the treatment dummy and the regional dummies only, Columns 2 and 5 control for household and individual characteristics and Columns 3 and 6 use Probit estimation (marginal effects are reported). Individual and household characteristics include age, gender, gender of head, education of head, age of head, indicator dummies for level of assets, household size and number of female and male children, an indicator variable whether the household lives in a urban settlement and an indicator variable whether the household lives in Skopje.

Table H2: Effect on Individual expenditure shares on Education: attrition checks

	OLS	OLS (weighted)	OLS	OLS (weighted)	OLS	OLS (weighted)
	(1)	(2)	(3)	(4)	(5)	(6)
A. No interaction						
Payment to mother	0.012	0.013				
	(0.011)	(0.012)				
B. Interaction with return to schooling						
Payment to mother * Return (1st tercile)			-0.011	-0.010		
			(0.018)	(0.019)		
Payment to mother * Return (2nd tercile)			0.015	0.019		
			(0.018)	(0.019)		
Payment to mother * Return (3rd tercile)			0.038*	0.034*		
			(0.014)	(0.014)		
C. Interaction with return to schooling in terms of employment						
Payment to mother * Return (1st tercile)					-0.016	-0.023
					(0.016)	(0.017)
Payment to mother * Return (2nd tercile)					0.019	0.025
					(0.014)	(0.015)
Payment to mother * Return (3rd tercile)					0.039*	0.042*
					(0.015)	(0.017)
Observations	911	911	911	911	911	911

Note. Standard errors clustered at municipality level in parenthesis. The dependent variable is computed as the ratio between the education expenditure on the children and the total expenditure of the household. Returns to schooling and expected incomes are computed assuming a triangular distribution and using log-income. I include controls for gender, age and education of household head, ethnicity, religion, household size, number of female and male children in age 13-18, presence of mother and father in the household, rural and Skopje dummies. Year and semester of birth dummies and Regional dummies are included. Return in terms of employment is defined as the difference between the probability of being employed after secondary school and the probability for being employed after primary school at age 25.

159

Table H3: Share of households owning an asset, by type of good and wealth quantile

	Wealth quantile			
	Lower (1)	Middle (2)	Higher (3)	All (4)
Characteristics of the dwelling				
Owner of the dwelling	0.475	0.507	0.521	0.501
Public water available	0.552	0.781	0.948	0.760
Electricity available	0.938	0.988	0.998	0.974
Telephone line available	0.306	0.465	0.621	0.464
Toilet connected to sewerage	0.172	0.413	0.773	0.452
Toilet connected to septic tank	0.274	0.346	0.214	0.278
Toilet not connected to sewerage or latrine	0.118	0.053	0.003	0.059
Toilet not shared with other households	0.918	0.948	0.965	0.944
Asset ownership				
Solid fuel cooker	0.915	0.973	0.766	0.885
Electric or gas cooker	0.286	0.701	0.930	0.639
Boiler	0.214	0.794	0.978	0.661
Refrigerator	0.731	0.925	0.988	0.881
Washing machine	0.244	0.672	0.940	0.618
Vacuum cleaner	0.122	0.607	0.900	0.543
Personal computer	0.017	0.0970	0.382	0.165
Mobile phone	0.774	0.828	0.853	0.818
Bycicle	0.052	0.102	0.254	0.136
Land and livestock property				
Household owns land	0.021	0.019	0.010	0.017
Household owns cattle	0.107	0.060	0.010	0.059

Note. Wealth quantiles are determined by the tercile in the wealth index, which is computed using Principal Component Analysis following Filmer and Pritchett (2001). The index is built using information on durable good ownership, access to utilities, type and property of the dwelling, land property and livestock ownership. For limited space, I don't include in the table the following group of indicators that were used in the computation of the wealth index: other types of dwelling property or rental, ownership of other types of animals, type of stove.

Table H4: Comparison of household characteristics by presence of male partner at the interview

	Presence of male partner		
	Absent	Present	Difference
Household members	4.824	4.686	-0.137
	[1.198]	[1.098]	(0.151)
Boys (6-13 years old)	0.336	0.254	-0.0817
	[0.509]	[0.460]	(0.0491)
Girls (6-13 years old)	0.395	0.283	-0.112*
	[0.641]	[0.507]	(0.0565)
Boys (14-19 years old)	0.782	0.866	0.0841
	[0.653]	[0.700]	(0.0714)
Girls (14-19 years old)	0.891	0.792	-0.0986
	[0.757]	[0.751]	(0.0864)
Male (head)	0.773	0.921	0.148***
	[0.421]	[0.270]	(0.0455)
Age (wife)	41.36	42.17	0.815
	[5.065]	[5.756]	(0.512)
Age (husband)	44.50	45.57	1.067**
	[4.700]	[5.439]	(0.519)
- *Lower primary or less (mother)*	0.252	0.197	-0.0550
	[0.436]	[0.398]	(0.0457)
- *Upper primary (mother)*	0.496	0.538	0.0418
	[0.502]	[0.499]	(0.0503)
- *Secondary school or more (mother)*	0.143	0.158	0.0148
	[0.351]	[0.365]	(0.0384)
- *Lower primary or less (father)*	0.227	0.201	-0.0262
	[0.421]	[0.401]	(0.0435)
- *Upper primary (father)*	0.496	0.529	0.0329
	[0.502]	[0.500]	(0.0551)
- *Secondary school or more (father)*	0.261	0.262	0.00114
	[0.441]	[0.440]	(0.0367)
- *Macedonian and others*	0.345	0.455	0.111*
	[0.477]	[0.498]	(0.0662)
- *Albanian*	0.429	0.292	-0.136*
	[0.497]	[0.455]	(0.0708)
- *Roma*	0.109	0.136	0.0270
	[0.313]	[0.343]	(0.0362)
- *Turkish*	0.134	0.118	-0.0162
	[0.343]	[0.323]	(0.0384)
Rural	0.504	0.459	-0.0454
	[0.502]	[0.499]	(0.0819)
Part of City of Skopje	0.235	0.102	-0.133**
	[0.426]	[0.303]	(0.0561)
Wealth (low)	0.261	0.348	0.0872*
	[0.441]	[0.477]	(0.0484)
Wealth (middle)	0.311	0.333	0.0224
	[0.465]	[0.472]	(0.0495)
Wealth (high)	0.429	0.319	-0.110*
	[0.497]	[0.467]	(0.0641)
Observations	119	558	677

Note. Standard errors in parenthesis clustered at municipality level. The table presents the comparison of reported subjective expectations conditional on having the female partner present at the interview at Baseline. Husband present is a dummy equal to 1 if the male partner is present and equal to 0 otherwise. Individual and municipality controls are included.

Table H5: Comparison of perceived returns by presence of male partner at the interview

	Dependent variable:									
			Conditional on completing Primary school				*Conditional on completing Secondary school*			
	Return (1)	Return (2)	Variance (3)	Variance (4)	Prob. job. (5)	Prob. job. (6)	Variance (7)	Variance (8)	Prob. job. (9)	Prob. job. (10)
Husband present	0.016	0.022	-0.007	-0.007	-0.014	-0.018	-0.003	-0.003	0.007	-0.001
	(0.048)	(0.048)	(0.005)	(0.005)	(0.020)	(0.022)	(0.003)	(0.003)	(0.022)	(0.023)
Individual controls		✓		✓		✓		✓		✓
Regional dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	844	844	844	844	844	844	844	844	844	844

Note. Standard errors in parenthesis clustered at municipality level. The table presents the comparison of reported subjective expectations conditional on having the female partner present at the interview at Baseline. Husband present is a dummy equal to 1 if the male partner is present and equal to 0 otherwise. Individual and municipality controls are included.

162

Chapter 6

Conclusions

This thesis is based on understanding whether a variation in intra-household allocation of resources has an effect on household decisions in developing countries and whether this has a relationship with information and subjective expectations. A first approach was to consider an exogenous variation in variables affecting the formation of the household, such as the marriage market. Another strategy was to study a policy that exogenously switch resources between household members. In the thesis I followed both approaches by looking at the Macedonian Conditional Cash Transfer for Secondary School Education, which is a program aiming at increasing secondary school attendance across poor households in the Republic of Macedonia and whose payments are randomised across the country in relation to the gender of the recipient (the mother versus the household head in the household). Although there are several "first generation" impact evaluation studies assessing the impact of conditional cash transfer programs, there are far fewer "second generation" studies focusing on the question of how to design CCTs more efficiently and specifically whether the gender of the recipient matters.

As a first approach to understand whether variables affecting the formation of the household, such as the marriage market, has an effect on intra-household dynamics, chapter 2 tested the assertion that the status of the marriage market impacts on intra-household allocation of expenditure by looking at the pre-policy structure of expenditures of poor households in Macedonia. I used measures of both the marriage and the re-marriage market to capture differences about these two variables. I find that sex ratios with a higher percentage of women affect intra-household allocation of expenditure through an increase in the expenditure share of education and health and through a decrease of the share allocated to

food. This provides new evidence on the relationship of sex ratio with labour market characteristics since we would expect rarity of women to have a positive effect on the decision power of women within the household. Additionally, by focusing on the municipality-level share of weddings in which the spouse had previously divorced from another partner, I find evidence against the unitary rationality hypothesis and I provide evidence that is consistent with the effect of the sex ratio and in favour of the efficiency assumption.

Chapter 2 showed that households don't behave like a unitary individual and that we need to take into account intra-household decision processes in order to understand household decisions. This is particularly important since there is still no consensus about how household members interact to take important investment decisions, such as determining human capital investments for the children. Understanding how this decision is taken is particularly important in developing countries, where households tend to underinvest due to market failures related to information, such as incomplete or asymmetric information. For this reason, during the 2010 and 2012 data collection waves of the Macedonian "Secondary School Conditional Cash Transfer" evaluation household survey, information about the parental perceived returns to schooling was collected to understand how information interact with intra-household dynamics to determine child human capital investment. Data on subjective expectations have been collected using the method proposed by Guiso et al. (2002). Under distributional assumptions, this method allows eliciting the subjective expected earning distribution for each respondent. By using the randomisation in the order of questions and testing whether respondents understand the question and use sufficient mental effort to report their answers, chapter 3 showed that when asked about the probability to earn less than a certain threshold and more than the same threshold, respondents tend to reply with sufficient mental effort only to the first question they are asked. This suggests it would be sensible to use only the first information reported by the respondent, specifically because complementary answers related to probabilities tend to fail in their second component. At the same time, the chapter provided evidence on the robustness of the method in relation to different distributional assumptions. These results are particularly important since there is still no consensus in literature about the best practice to collect subjective expectations in developing countries.

The Macedonian CCT Evaluation database and its module on subjective expectations related to schooling allowed studying the relationship between ex-ante

expectations and schooling decisions. Chapter 4 investigated the role of parental expected returns to schooling as determinants for schooling decisions two years after the subjective expectation was reported.

The chapter analyses the relationship between schooling decisions and ex-ante parental perceived returns to secondary school. 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. In addition, by using the longitudinal dimension of the data, I provide evidence against cognitive biases in expectation reporting and against endogeneity issues, which provide support for the use of subjective data in decision models. The chapter provided evidence that ex-ante parental expectations are important in explaining secondary schooling decisions for children. Additionally, important differences exists across gender: 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.

When studying the relation between schooling decisions and self-reported income expectations, criticisms claim that results might be biased by cognitive dissonance, e.g. the tendency of the respondent to over-report their expectations to support their decisions. In the Macedonian setting, cognitive dissonance would affect the updating of expectations such that expectations linked to choices made during the two data collection points 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, chapter 4 provided important 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.

Chapters 2 and 4 showed that both intra-household power allocation and the subjective expectations are credible determinants of schooling decisions. Specifically, chapter 5 studied how the interaction between intra-household allocation of resources and expected returns to schooling influences human capital investment among poor households. To do this I used a nationally implemented randomised programme in the Republic of Macedonia. The programme provided cash transfers to poor households conditional on secondary school enrolment of their children and payments were transferred either to mothers or to household heads. The availability of the unique dataset with information on parental subjective expectations of returns to schooling and employment risk allowed estimating the heterogeneity of the effect of the intervention. This chapter provided evidence that targeting mothers allowed increasing secondary school enrolment only for children whose returns are sufficiently high at the beginning of the program. This outcome was driven by increases in individual expenditure shares on education for children in the highest tercile of the return distribution. Heterogeneous effects along the expected return distribution are supported by large ex-ante heterogeneity in parental expectations as discussed in chapter 4. At the same time, no effect was recorded for other inputs, such as monitoring of children school attendance and parental time use.

These findings suggest that in order to understand the role of each member in familial interactions is key not only to clarify how decision power is distributed across members, but as well how this relates to subjective expectations. This is particularly important in developing countries, since perceived returns to education are often are different than the market returns. Individuals have largely heterogeneous expectations related to education and the perceived returns do correlate with future choices and outcomes. If shifts in parental expectations have indirect impacts on the decision to invest on human capital, then much work need to be done in order to understand how subjective expectations form, how they evolve over time and how they interact with individual and collective choices.

Bibliography

Akresh, R., De Walque, D., and Kazianga, H. (2012). Alternative cash transfer delivery mechanisms: impacts on routine preventative health clinic visits in burkina faso. Technical report, National Bureau of Economic Research.

Angrist, J. (2002). How do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation. *Quarterly Journal of Economics*, 117(3):997–1038.

Armand, A. and Carneiro, P. (2013). Impact evaluation of conditional cash transfers for secondary school attendance in macedonia: Results from the first follow-up. Technical report, Ministry of Labour and Social Policy, Republic of Macedonia.

Ashworth, K., Hardman, J., Hartfree, Y., Maguire, S., Middleton, S., Smith, D., Dearden, L., Emmerson, C., Frayne, C., and Meghir, C. (2002). Education maintenance allowance: The first two years. a quantitative evaluation. Research report rr352, Department for Education and Skills. Nottingham, UK: Queen's Printer.

Attanasio, O. (2009). Expectations and Perceptions in Developing Countries: Their Measurement and Their Use. *American Economic Review*, 99(2):87–92.

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

Attanasio, O. and Lechene, V. (2002). Tests of income pooling in household decisions. *Review of Economic Dynamics*, 5(4):720–748.

Attanasio, O. and Lechene, V. (2011). Efficient responses to targeted cash transfers. *Journal of Political Economy, forthcoming*.

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

Attanasio, O. P., Meghir, C., and Santiago, A. (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.

Baird, S. J., Garfein, R. S., McIntosh, C. T., and Özler, B. (2012). Effect of a cash transfer programme for schooling on prevalence of hiv and herpes simplex type 2 in malawi: a cluster randomised trial. *The Lancet*, 379(9823):1320–1329.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., and Perez-Calle, F. (2011). Improving the design of conditional transfer programs: Evidence from a randomized education experiment in colombia. *American Economic Journal: Applied Economics*, 3(2):167–195.

Basu, K. (2006). Gender and say: a model of household behaviour with endogenously determined balance of power. *Economic Journal - London*, 116(511):558.

Becker, G. (1991). *A Treatise on the Family*. Harvard Univ Pr.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P., and Pouliquen, V. (2013). Turning a shove into a nudge? a "labeled cash transfer" for education. Technical report, National Bureau of Economic Research.

Betts, J. R. (1996). What do students know about wages? evidence from a survey of undergraduates. *Journal of Human Resources*, pages 27–56.

Blau, F. D. (1990). Career plans and expectations of young women and men: the earnings gap and labor force participation. Technical report, National Bureau of Economic Research.

Bloch, F. and Rao, V. (2002). Terror as a bargaining instrument: A case study of dowry violence in rural India. *The American economic review*, 92(4):1029–1043.

Blundell, R., Chiappori, P.-A., and Meghir, C. (2005). Collective labor supply with children. *Journal of Political Economy*, 113(6):1277–1306.

Blundell, R. and Meghir, C. (1987). Bivariate alternatives to the Tobit model. *Journal of Econometrics*, 34(1-2):179–200.

Bobonis, G. (2009). Is the Allocation of Resources within the Household Efficient? New Evidence from a Randomized Experiment. *Journal of Political Economy*, 117(3):453–503.

Bourguignon, F., Browning, M., and Chiappori, P. (2009). Efficient intra-household allocations and distribution factors: Implications and identification. *Review of Economic Studies*, 76(2):503–528.

Bourguignon, F., Browning, M., Chiappori, P., and Lechene, V. (1993). Intra household allocation of consumption: a model and some evidence from French data. *Annales d'Economie et de Statistique*, pages 137–156.

Bourguignon, F., Browning, M., Chiappori, P., and Lechene, V. (1994). Incomes and Outcomes: a Structural Model of Intra-Household Allocation. *Journal of Political Economy*, 102(6):1067–1096.

Browning, M. and Chiappori, P. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, pages 1241–1278.

Chiappori, P., Fortin, B., and Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy*, 110(1):37–72.

Chiappori, P.-A. (1992). Collective labor supply and welfare. *Journal of Political Economy*, pages 437–467.

De Brauw, A. and Hoddinott, J. (2011). Must conditional cash transfer programs be conditioned to be effective? the impact of conditioning transfers on school enrollment in mexico. *Journal of Development Economics*, 96(2):359–370.

Delavande, A., Giné, X., and McKenzie, D. (2010). Measuring subjective expectations in developing countries: a critical review and new evidence. *Journal of Development Economics*.

Delavande, A. and Kohler, H. (2009). Subjective expectations in the context of hiv/aids in malawi. *Demographic research*, 20:817.

Dominitz, J. and Manski, C. (1997). Using expectation data to study subjective income expectations. *Journal of the American Statistical Association*, 27 (1):30–41.

Dominitz, J. and Manski, C. F. (1996). Eliciting student expectations of the returns to schooling. *Journal of Human Resources*, 31:1–26.

Doss, C. (2006). The effects of intrahousehold property ownership on expenditure patterns in ghana. *Journal of African Economies*, 15(1):149–180.

Duflo, E. (2003). Grandmothers and Granddaughters: Old Age Pension and Intra-household Allocation in South Africa. *The World Bank Economic Review*, 17(1).

Duflo, E. (2005). Gender Equality in Development. *BREAD Policy Paper*, 11.

Duflo, E. and Udry, C. (2004). *Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices*. National Bureau of Economic Research Cambridge, Mass., USA.

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

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

Fiszbein, A. and Schady, N. (2009). *Conditional cash transfers: reducing present and future poverty*. World Bank Publications.

Freeman, R. (1971). The market for college trained manpower.

Giustinelli, P. (2011). Group decision making with uncertain outcomes: Unpacking child-parent choices of high school tracks. Technical report.

Grossbard-Shechtman, S. and Neideffer, M. (1997). Women's hours of work and marriage market imbalances. *Economics of the Family and Family Policies*, pages 100–118.

Grossbard-Shechtman, S. and Treble, J. (1993). *On the economics of marriage: A theory of marriage, labor, and divorce*. Westview.

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

Hill, R. (2006). Coffee price risk in the market: Exporter, producer and trader data from uganda.

Hoddinott, J. and Haddad, L. (1995). Does female income share influence household expenditures? Evidence from Côte d'Ivoire. *Oxford Bulletin of Economics and Statistics*, 57(1):77–96.

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

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

Ligon, E. (2002). Dynamic bargaining in households (with application to Bangladesh). *UC Berkeley: Department of Agricultural and Resource Economics, UCB*.

Lundberg, S. and Pollak, R. (1993). Separate spheres bargaining and the marriage market. *Journal of Political Economy*, 101(6):988.

Lundberg, S. and Pollak, R. (2003). Efficiency in marriage. *Review of Economics of the Household*, 1(3):153–167.

Lundberg, S., Pollak, R., and Wales, T. (1997). Do husbands and wives pool their resources? Evidence from the United Kingdom Child Benefit. *The Journal of Human Resources*, 32(3):463–480.

Luseno John, G. and Winnie, K. (2003). Assessing the value of climate forecast information for pastoralists: Evidence from southern ethiopia and northern kenya. *World Development*, 31(9):1477–1494.

Lybbert, T., Barrett, C., McPeak, J., and Luseno, W. (2007). Bayesian herders: Updating of rainfall beliefs in response to external forecasts. *World Development*, 35(3):480–497.

Mahajan, A., Tarozzi, A., Yoong, J., and Blackburn, B. (2008). Bednets, information and malaria in orissa. In *World Bank Poverty and Applied Micro Seminar Series, World Bank, Washington, DC (December 10)*.

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

McKenzie, D., Gibson, J., and Stillman, S. (2007). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *World Bank Policy Research*, (Working Paper 4141).

McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Journal of Population Economics*, 18(2):229–260.

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

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. *Unpublished manuscript*.

Oreffice, S. (2007). Did the legalization of abortion increase women's household bargaining power? Evidence from labor supply. *Review of Economics of the Household*, 5(2):181–207.

Phipps, S. and Burton, P. (1998). What's Mine is yours? The influence of male and female incomes on patterns of household expenditure. *Economica*, 65(260):599–613.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Rubalcava, L. and Thomas, D. (2000). Family bargaining and welfare. *UC Los Angeles: California Center for Population Research*.

Schultz, T. (1990). Testing the neoclassical model of family labor supply and fertility. *The Journal of Human Resources*, 25(4):599–634.

Smith, H. L. and Powell, B. (1990). Great expectations: Variations in income expectations among college seniors. *Sociology of Education*, pages 194–207.

State Statistical Office, S. (2002). Census of population, households and dwellings in the republic of macedonia. *Book XIII*.

State Statistical Office, S. (2007). Census of agriculture.

Thomas, D. (1990). Intra-household resource allocation: An inferential approach. *The Journal of Human Resources*, 25(4):635–664.

Thomas, D., Contreras, D., and Frankenberg, E. (1997). Distribution of power within the household and child health. *RAND: Santa Monica, CA*.

Udry, C. (1996). Gender, agricultural production, and the theory of the household. *The Journal of Political Economy*, 104(5):1010–1046.

Verme, P. (2008). A review of the social protection system in the republic of macedonia. *Ministry of Labour and Social Policy, Republic of Macedonia*.

Ward-Batts, J. (2008). Out of the wallet and into the purse using micro data to test income pooling. *Journal of Human Resources*, 43(2):325–351.

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