THESIS DECLARATION

The undersigned

SURNAME	BATTAGLIA
NAME	MARIANNA
PhD Registration Number	1395136

Thesis title:

MIGRANTS AND MINORITIES: HEALTH AND EDUCATION CHOICES

PhD in	ECONOMICS
Cycle	XXIV
Candidate's tutor	ELIANA LA FERRARA
Year of discussion	2013

DECLARES

Under his responsibility:

1) that, according to the President's decree of 28.12.2000, No. 445, mendacious declarations, falsifying records and the use of false records are punishable under the penal code and special laws, should any of these hypotheses prove true, all benefits included in this declaration and those of the temporary embargo are automatically forfeited from the beginning;
2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree of 30th April 1999 protocol no. 224/1999, to keep copy of the thesis on deposit at the Biblioteche Nazionali Centrali di Roma e Firenze, where consultation is permitted, unless there is a temporary embargo in order to protect the rights of external bodies and industrial/commercial exploitation of the thesis;
3) that the Servizio Biblioteca Bocconi will file the thesis in its 'Archivio istituzionale ad accesso aperto' and will permit on-line consultation of the complete text (except in cases of a temporary embargo);
4) that in order keep the thesis on file at Biblioteca Bocconi, the University

requires that the thesis be delivered by the candidate to Società NORMADEC (acting on behalf of the University) by online procedure the contents of which must be unalterable and that NORMADEC will indicate in each footnote the following information:

- thesis MIGRANTS AND MINORITIES: HEALTH AND EDUCATION CHOICES
- by BATTAGLIA MARIANNA;
- discussed at Università Commerciale Luigi Bocconi – Milano in 2013 ;
- the thesis is protected by the regulations governing copyright (law of 22 April 1941, no. 633 and successive modifications). The exception is the right of Università Commerciale Luigi Bocconi to reproduce the same for research and teaching purposes, quoting the source;

5) that the copy of the thesis deposited with NORMADEC by online procedure is identical to those handed in/sent to the Examiners and to any other copy deposited in the University offices on paper or electronic copy and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (law of 22 April 1941, no. 633 and successive integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal and shall be exempt from any requests or claims from third parties;

7) that the PhD thesis is not the result of work included in the regulations governing industrial property, it was not produced as part of projects financed by public or private bodies with restrictions on the diffusion of the results; it is not subject to patent or protection registrations, and therefore not subject to an embargo;

Date 31/10/2012

SURNAME	BATTAGLIA	NAME	MARIANNA

MIGRANTS AND MINORITIES: HEALTH AND EDUCATION CHOICES

Contents

List of Tables

List of Figures

Acknowledgments

I am particularly grateful to my advisors Eliana La Ferrara and Michele Pellizzari for their continuous encouragement and support during these years. I also thank Jean-Marie Baland for his helpful comments and suggestions, both during my period in Namur and once back to Milan.

I am indebted to my co-author Lara Lebedinski, the Serbian Ministry of Education and Ipssos Strategic Marketing Agency for the help provided in the data collection process.

I would like to thank all colleagues and friends who shared this experience with me, both in Belgium and Italy. Special thanks go to Güneş, Daniele, Marinella, Lore, Annaig, Annamaria, Selim, Paolo, Luna and Francesco.

I will always be thankful to my parents, I never tell them enough.

Introduction

This dissertation consists of three essays in the field of development economics. Each paper makes use of microeconometric techniques to answer research questions on migrants and minorities and their health and education choices. The first paper investigates the impact of migration on fertility and health care choices of those remaining in the country of origin. The second and third papers focus on the Roma minority group and the effects of a remedial education programme on their schooling outcomes and educational aspirations.

The first chapter – *Migration, Health Knowledge and Teenage Fertility: Evidence from Mexico* – studies the role of migration as a channel to transfer norms from migrants' destinations to country of origin, a topic so far not largely investigated. In particular I focus on fertility. The results suggest specific policy interventions to improve sexual education and preventive infant health-care policies in the country of destination, targeting explicitly migrants. I focus on Mexico, a country with a long lasting history of migration, and I use a nationally representative demographic survey – ENADID – for the years 1997 and

2006. I primarily investigate impacts on teenage women because delaying onset of fertility is still a policy priority in the country. Although fertility rates have decreased considerably in the last decades, adolescent fertility rates have declined much more slowly and the proportion of teenage mothers is still high. This constitutes a problem because adolescent pregnancy and childbearing are associated with a range of adverse health outcomes, including high risk of pregnancy-related diseases, unsafe abortion practices, and maternal mortality. Adolescent mothers are also less likely to use either antenatal or delivery care than women aged 19-23 and they are less likely to have their infants immunized. Furthermore, especially among the poor, adolescent childbearing is linked with lower future monthly earnings for mothers, and contributes to the persistence of poverty from one generation to the next, affecting long term human capital accumulation and growth.

In this study I use historic migration rates interacted with the number of adult males in the household as an instrument to account for the endogeneity of migrant status. Data suggest that women – especially teenagers – in households with at least one member who migrated to the United States have a lower occurrence of pregnancy. This finding can be partially explained by the fact that teenagers in migrant households have a higher knowledge of contraceptive methods and likely practice active birth control. Migrants share similar economic and institutional environments and acquire information and behavioural norms of the context of residence. They both assimilate – change in pref-

erences – and adapt – adjustment in fertility behaviour due to economic opportunities and constraints – to alternative models of gender roles and family relationships. Migrants might therefore act as a channel of transmission of such norms and affect the behaviour of other family members in the home country: they can provide behaviour examples of behaviours to be considered and copied by those remained at home. A part of the effect I found is undoubtedly related to the characteristics of the context I study, and especially applies to those countries experiencing temporary and recurrent migration. Strong family ties of the Mexican society and the high percentage of people moving back and forth to the United States allow migration networks to play a role in providing information.

The second – *Equal access to education: An Evaluation of the Roma Teaching Assistant Programme in Serbia* – and the third chapter – *The Curse of Low Aspirations: Remedial Education and Perceived Returns to Education of Roma People* – are part of a project conducted together with my co-author Lara Lebedinski. We collected first-hand data from both schools and families involved in the Roma Teaching Assistant Programme (RTA), a remedial education programme introduced in Serbia in 2009. The programme assigns each school one teaching assistant. Her major tasks are helping children during regular classes, organising after-school extra classes, and act as direct link between the school and parents. To the best of our knowledge, there are not systematic studies in economic literature that try to investigate how to improve life circumstances of Roma, in

general, and Roma kids, in particular. These papers are a first solid attempt in this direction and they contribute to the existing literature by providing an accurate overview of the attainments of Roma pupils in Serbian schools, for which so far there were not data available, and by contrasting their achievement to average Non Roma pupils. More broadly, the papers add evidence on short-term effects of remedial education programmes on minority groups and investigate the role of these programmes in changing aspirations. They suggest replicable examples in contexts where minorities suffer low attainment rates and social exclusion. For Roma people this is the case in many other European countries. Roma constitute the largest ethnic minority in the continent. They suffer severe social exclusion, especially in terms of high poverty levels, low educational attainments and low social mobility. Understanding the underlying reasons for their social exclusion and finding ways to improve their living standard is the core aim of this project. We tackle this issue by examining the effect on schooling outcomes and educational aspirations of the RTA programme. Schooling has always been considered a needed measure to improve living conditions of Roma people and attempts in this direction have been devised by various countries for many years. Half of the Roma population is younger than 18 years old, so that focusing on children and young people is broadly recognised as a crucial step towards Roma inclusion. Higher enrolment rates and better achievement at school are expected to lead to persistent effects in the labour market and in the reduction of poverty in the long-run.

We first investigate the impact of this remedial education programme on schooling outcomes in its first year of introduction. More precisely, in the *second chapter*, we examine whether this remedial education programme is effective in reducing dropouts, raising attendance and improving the marks of Roma pupils. By using administrative data collected in schools across the country, we employ two different identification strategies and their combination. First, we compare *Early* and *Late Enrollees* by exploiting the gradual implementation and the intensity of the programme. Second, we compare children exposed to the programme to older cohorts not exposed to it in the same school. We find that, on average, marks have improved and dropouts have reduced for those children exposed to the programme in their first grade. There is also evidence that overall children exposed to the programme went on average more to school. Higher and more systematic impacts are attained in schools with a lower number of Roma.

A remedial education programme for primary school-age children may also affect parental aspirations about their children's future. In the *third chapter*, we investigate whether expectations on labour market perspectives and educational achievement change as a consequence of exposure to the Roma Teaching Assistant Programme. We argue that these changes are likely to occur mainly through a role model mechanism: in the programme all the assistants are Roma and from the same social background of the pupils they help. The presence of a person belonging to the same community, who showed to be successful, motivates

parents to believe their children can succeed. We use data collected among 300 Roma households in Belgrade in Fall 2010 and exploit the gradual implementation of the programme. Results suggest that parents of pupils in treated schools expect higher returns to education for their kids. They are also more likely to expect them to achieve a secondary level of education. One year of a remedial education programme may not be enough to break the curse of low aspirations, but we find encouraging results in this direction.

Chapter 1

Migration, Health Knowledge and Teenage Fertility: Evidence from Mexico

Migration may affect fertility and preventive health care of those remaining in the country of origin. Mexican data suggest that women - especially teenagers - in households with at least one member migrated to the United States have a lower occurrence of pregnancy. This finding can be partially explained by the fact that teenagers in migrant households have a higher knowledge of contraceptive methods and likely practice active birth control. I use the historic migration rates interacted with the number of adult males in the household as an instrument to account for the endogeneity of migrant status.

1.1 Introduction

Far from being a new phenomenon, Mexican migration to the United States has seen increasing flows of migrants over time. Their number has been both in absolute and relative terms larger than any other immigrant influx in the past century, corresponding nowadays to approximately seven million of people and therefore 31% of the foreign born population (U.S. Census Bureau, 2006).[1] According to the last Mexican Census (INEGI, 2000), from 1995 to 2000, 4.1% of all Mexican households saw at least one family member migrating to live in the United States, while an additional 1.8% of households had family members migrating back and forth between the two countries or returning to Mexico. Overall, almost 10% of the population born in Mexico resides now in the United States (UNDP, 2007).[2] The paper aims at investigating the impact of this international migration on fertility, knowledge of contraceptive methods and preventive child health care for those remaining in Mexico. Migration is

[1] In 2006, U.S. civilian labour force counted 149,668 million of people, 22,586 of which foreign born. The fraction of the US-workforce of 16 years and over composed of Mexican-born workers increased rapidly after 1970, passing from 0.4% in 1970 to 4.7% in 2006. Workforce migration levels analogous to those of the '70s were already reached in 1920 and continued rising until the late '20s, favoured by the outbreak of World War I and the Congressional action that restricted immigration from Europe (Massey et al., 2002; Lemay and Barkan, 1999). The following decades were characterized by a long steady decline in the Mexican immigration share. It lasted until the Bracero Program was introduced in 1942 to ease the labour force shortage caused by World War II in the agricultural sector.

[2] Mexico-U.S. migration is largely illegal and these are official data. Hence, they might underestimate the phenomenon.

expected to act mainly as a transfer of behavioural norms from host to migrants' home country: migrants, through their exposure to U.S. practices, acquire information, adopt behaviours and transmit them to the other family members in Mexico.

I focus on teenage women because delaying onset of fertility is still a policy priority in the country. Although fertility rates have decreased considerably in the last decades - the number of children per women for childbearing age went from 6.8 in 1970 to 2.2 in 2006[3] - adolescent fertility rates have declined much more slowly and the proportion of teenagers who are mothers is still around 17% (INEGI, 2005). The pregnancy rate for women between 12 and 19 years old was still 79 out of 1,000 in 2005 and in public institutions 21% of childbirths were given by women less than 20 (CONAPO, 2007).[4] Adolescent pregnancy and childbearing are associated with a range of adverse health outcomes, including high risk of pregnancy-related diseases, unsafe abortion practices, and maternal mortality.[5] Pregnancy and

[3]Since 1974 the Mexican government has adopted a family-planning program relied on persuasion: it has continuously broadcasted the jingle "Small Families Live Better" on television (*"La Familia Pequena Vive Mejor"* in the '70s and *"Pocos Hijos Para Darles Mucho"* in the '80s). Although some researchers estimate that government programs have only a trivial effect, others say that such programs explain as much as 40% in fertility patterns between the 1960s and 1990s (Weil, 2005).

[4]In 2005 in the United States the birth rate among teenagers aged 15-19 years was 40.5 out of 1,000 women. 10.2% of births were to women aged 19 years and younger (U.S. HHS, 2007). The corresponding percentages for women less than 20 giving birth are around 4% in most developed countries, while they can reach 50% in some African countries (CONAPO, 2007).

[5]Despite the restrictive nature of the law in Mexico, abortion is widely practiced: in the last decades each year about 24% of women of reproductive age were estimated to have undergone an abortion and about 65 out

childbirth are the leading causes of death among women aged 15-19 and the risk of maternal death during childbirth is 1.2 times higher among adolescents than among women aged 20 or older (Save the Children, 2004). Adolescent mothers are less likely than women aged 19-23 to use either antenatal or delivery care - which would improve maternal outcomes through prevention and treatment of obstetric complications. They less likely have their infants immunized, particularly for diphtheria, pertussis, tetanus and measles (Reynolds et al., 2006).[6] Moreover, especially among the poor, adolescent childbearing is linked with lower future monthly earnings for mothers, and contributes to the persistence of poverty from one generation to the next, affecting long term human capital accumulation and growth. Adolescent child bearers are themselves born to adolescent mothers: two-thirds of the adolescent mothers in Mexico have mothers who gave birth in their teens (Buvinic, 1998).

The principal channels through which migration may affect fertility, family planning decisions and preventive child care are both wealth and health knowledge (Hildebrandt and McKenzie, 2005). On one side, remittances sent back from migrants likely change household income and allow families to spend resources on different categories of expenditure, such as education and preventive health care. On the other side, through exposure

of 100,000 annual maternal deaths were due to illegally induced abortions performed in unsanitary conditions and by unqualified personnel (United Nations, 2002). A revised abortion law has been introduced in 2007.

[6]Adolescents' babies have a 34% higher risk of death in the neonatal period, mainly because of their increased risk of being low-birth weight, and have a 26% higher risk of death by age five (Bicego and Ahmad, 1996).

to U.S. practices, migrants acquire health knowledge, use more contraceptive methods and change their individual preferences. They provide examples of behaviours that may be considered and copied by other family members at home. More specifically, higher levels of educational attainment and health investment are plausibly expected in households experiencing migration and receiving remittances, with consequences on human capital accumulation and family planning decisions (Adams, 2005; Duryea et al., 2005; Cox et al., 2004). Access to education of adolescent girls may delay the onset of fertility (Breierova and Duflo, 2004; Duflo et al., 2011), increase the opportunity cost of women's time leading to changes in preferences regarding the quality and quantity of children (Becker et al., 1960), and endow girls with a better ability to process information, potentially increasing knowledge on contraception options (Rosenzweig and Schultz, 1987; Duncan et al., 1991).[7] Still, health knowledge - in the form of information and behaviours that flow from receiving to sending country communities - can represent an additional pathway through which migration affects health outcomes and fertility behaviours. Kovsted et al. (2002), for instance, suggest that although maternal education is important in determining child health and mortality, its effect diminishes or disappears when health knowledge is introduced as an explanatory variable.

[7]Recently some elements of sex-education are introduced in schools of many countries - due to the spread of the HIV epidemic in Africa, directly affecting knowledge of contraception methods, as well as incentives to avoid risky sexual activity. In Mexico, the first program for the spread of information for the prevention of AIDS/HIV in schools was introduced only in 2008 by the Ministerial Declaration "Prevenir con Educación".

Health knowledge crowds out the effects of mother's education of the mother and needs thus not to be closely associated with the income or education of parents. The information and behaviours obtained through the experience in the receiving country have therefore the potential to result in less risky sexual behaviours and in the long run in better health outcomes also in the sending country.

As much of the literature suggests (Fargues, 2006; Beine et al., 2008; Lindstrom and Munoz-Franco, 2005; Lindstrom and Saucedo, 2002), migrants' behavioural norms tend to converge to those of their host country. Through the process of living and working in U.S., migrants acquire information and become aware of alternative models of gender roles and family relationships which they may accept and adopt. This assimilation process – a change in individual preferences – is gradual. The impact of host country values and norms of migrants' behaviour increases with the length of the migration: the longer a migrant has been in the United States, the more his outcomes should look like those of a US native. The adjustment in fertility behaviour may also occur in response to the economic opportunities and constraints present in the destination country and therefore taking place also in a short period of time. Both processes take place simultaneously so it is challenging to disentangle the effects of assimilation – change in preferences – from those of adaptation – adjustment in fertility behaviour due to economic opportunities and constraints (Lindstrom and Saucedo, 2002). Overall, what it is observed is that the Mexican-born population's fertil-

ity rate is converging towards the Not-Hispanic U.S. fertility rate (IPUMS, 2010). Migrants are likely to share similar economic and institutional environments with other North American and acquire information and behavioural norms of the context of residence. This information and behaviours are then transmitted to the other family members remained in Mexico. Information about contraception transmitted by relatives and friends is often more influential in making decisions about the adoption and practice of contraception than it is information received through impersonal or formal channels (Lindstrom and Munoz-Franco, 2005). The strong family ties of the Mexican society and the recurrence of the Mexican migration to the United States - a high percentage of people tend to move back and forth - make the role of migration networks highly plausible.

The major concern in this study is the endogeneity of migrant status: family planning decisions can be related to the characteristics of migrants themselves. The direction of the potential selection of individual into migration is a priori unclear. It is difficult to disentangle the extent to which the impacts on fertility, knowledge of contraceptive methods and preventive child care reflect the unobserved characteristics of migrants, their households or their communities as opposed to the migration experience itself. I tackle the problem by using an instrumental variable method. Historic migration networks are used as instrument for current levels of migration. Historically migrants have mainly come from the central-western region of the country where the

rail lines going to the United States were built.[8] The states of Jalisco, Michoacán, and Guanajuato - that together accounted for about 30% of all Mexico-U.S. migrants throughout the twentieth century - are still among the principal sending states.[9] In order to obtain variations at the household level, 1924 migration levels are interacted with the number of adult males in the household: in the Mexican context, the likelihood that a member of a household migrates increases with the number of males in the household. My results suggest that women - especially teenagers - in households with at least one member who migrated to the United States have a lower occurrence of pregnancy compared to women in no-migrant households. This finding can be partially explained by the fact that teenagers in migrant households have a higher knowledge of contraceptive methods and likely practice active birth control. Less significant and more controversial results are obtained for child care outcomes.

[8]In Mexico, between 1884 and 1900, three major north-south rail lines were built. The first, the Central Mexican Railroad, went south from what is now Ciudad Juarez - connected to the Southern Pacific and Texas Pacific railroads in Texas - to the state of Guanajuato, where it branched east to Mexico City and west through Guadalajara to Colima. A second line, the Mexican International Railroad ran from Durango through Chihuahua to Piedras Negras until Texas. A third, the Mexican National Railroad, travelled north from Mexico City through San Luis Potosi and Monterrey, reaching the border at Nuevo Laredo (Woodruff, 2007).

[9]In the twentieth century residents of Jalisco, Michoacán, and Guanajuato were roughly twice as likely to migrate as the average Mexican. Just 1.5% of migrants came from Chiapas, Campeche, Yucatán and Quintana Roo - which are home to 7% of the population (Durand et al., 2001). The data indicate that the high-migration states were, on average, poorer than low-migration states during the first half of the century with lower level of premigration measures of health, education and income.

So far the literature has not largely investigated the interaction between migration and health outcomes or, even more specifically, sexual behaviour and fertility. When it did, it mainly studied the potential effects in the country of destination - access to welfare benefits (Bertrand et al., 2000) or to health services (Menjivar, 2002; Deri, 2005; Devillanova, 2008) and fertility (Chou, 2010): little has been said about the impact of migration on these outcomes in the sending region. There are some recent contributes in this direction. Beine et al. (2008) examine the relationship between international migration and source country fertility and provide evidence of a strong transfer of fertility norms from migrants to their country of origin. Analogously, in their study on rural Guatemala, Lindstrom and Munoz-Franco (2005) found that contraceptive use increases and fertility falls with variables such as having family members in urban or international destinations and living in a community where urban migration is common. Kanaiaupuni and Donato (1999); Frank and Hummer (2002); Hildebrandt and McKenzie (2005) use Mexican data to investigate the effects of migration on infant mortality and birth weight, finding positive effects for children living in households with at least one migrant. In Pakistan, Mansuri (2006) found that migration has a positive impact on early child growth outcomes, especially for young girls. My paper is related to this strand of the literature and attempts to provide a first deep investigation of the impact of migration on fertility in the country of origin - especially for adolescent women for whom delaying onset of fertility is still a policy priority in many countries.

The paper is structured as follows. Section 2 provides information on the data and variables used in the study. The empirical strategy is presented in Section 3. Section 4 illustrates the results. Conclusion follows.

1.2 Data and Context

This study uses data from the National Survey of Demographic Dynamics (ENADID - Encuesta Nacional de Dinámica Demográfica) for the years 1997 and 2006. The ENADID is a nationally representative demographic survey conducted by the Mexican national statistical agency - Instituto Nacional de Estadística y Geografía (INEGI) - in partnership with the national public health and demographical agencies, Instituto Nacional de Salud Pública (INSP) and Consejo Nacional de Población (CONAPO). The households for which information is collected are not the same across years. Therefore I use a pooled cross-section dataset.[10] Detailed individual and household characteristics, including demographics, education and dwelling infrastructures are available in the pooled sample for 468,503 individuals in 107,654 dwellings across all 32 Mexican states. The survey also offers information on permanent and temporary migration to the United States: for migrants, data are collected on age, sex, relationship with the head of the household, year and duration of migration, state of destination, reasons to migrate and remit-

[10] I chose to use both years - 1997 and 2006 - in order to get a bigger sample. Data from 1992 are also available, but the section on *migration* is not as reliable as in the other two years.

tances.[11] Moreover, all women aged 15 to 54 in each household are asked detailed questions about their fertility history, knowledge and use of contraceptive methods and child health care.

For the purpose of the analysis the attention has been confined to women for whom information on migration is available. The pooled dataset counts therefore 39,203 women aged 15 to 54 in 28,433 households. Among those women, almost 21% are less than 20 years old.

Individual and household characteristics Table 1 presents summary statistics for individual and households characteristics of the pooled sample, both for all women and teenagers.

[insert TABLE 1.1 here]

Overall, teenagers in households with at least one member who migrated to the United States have less educated parents and live in larger families, in a bigger extent located in rural areas, when compared with teenagers in no-migrant households. No statistically significant differences are recorded for the level of education: on average nine years of schooling (junior high school).[12] These descriptive results are in line with much of

[11] Every household member of at least 12 years old living in the household have been asked on her migration experience. For those household member still abroad information is given by the head of the household. Some data are available only for the 2006 survey.

[12] In the Mexican education system, a child aged 3 to 5 attends preschool (not compulsory), from 6 to 11 years old she attends primary school/*primaria*, from 12 to 14 junior high school/*secundaria*, from 15 to 17 high school/*bachillerato-profesional tecnico* and from 18 years old onwards university/*licenciatura*. Level 3 in the dataset corresponds to *secundaria*, equivalent to nine years of schooling.

the literature that suggests that Mexican migrants usually come from the lower middle ranges of the socioeconomic scale (Massey et al., 1994; Borjas and Katz, 2005).[13]

International migration The ENADID survey asks each household member whether she has ever been to the United States in the five years prior to the day of the interview.[14] This implies that those who migrated and returned to Mexico before five years prior to that day are not considered as migrants in my sample. I am not concerned that this definition would bias my results because Mexican migration to the United States tends to be recurrent. If one migrates once it is likely that she will migrate again.[15] In my pooled sample almost 54% of the individuals who migrated moved back and forth.

The outcomes of interest need not to be affected by the same shocks leading to the migration decision. I am interested in the impact of previous migrant experience on subsequent occurrence of pregnancy and preventive child care outcomes. As suggested by Hildebrandt and McKenzie (2005), I divide the five years period in two: the first two years and the second three. A household is defined as a *migrant household* if it has at least one member

[13]Women in migrant households have a higher monthly income than women in no-migrant households likely because these values include remittances. However, there is a high percentage of households reporting a monthly household income equal to zero Pesos. The variable is therefore not reliable for estimation.

[14]The surveys have been conducted in the last quarter of 1997 and in the first quarter of 2006.

[15]They spend three/four years in the United States before returning home after a single migration spell (Munshi, 2003).

older than 12 years old who had migrated to the United States between January 1st 1992 and January 1st 1994 or between January 1st 2001 and January 1st 2003, depending on the dataset.[16] The occurrence of pregnancy and child health care outcomes are therefore recorded since January 1st 1994 or January 1st 2003.[17] Migrant households correspond to 10% of the pooled sample. Migrants mainly move in search of work – about 82% – and towards California and Texas – respectively 36% and 15%. They are mainly young males who cross the border illegally: almost 84% of those in the sample are undocumented.[18] The typical migrant is either the (elder) son of the household head or the head of the household.

[insert TABLE 1.2 here]

Occurrence of pregnancy, knowledge of contraceptive methods and desired fertility My outcomes of interest are occurrence of pregnancy and child health care. ENADID asks women aged 15 to 54 whether they ever got pregnant, whether they gave birth or have gone through abortion, the number of

[16]In ENADID individuals are considered able to migrate in search of a job if they are older than 12 years old. Thus, in my study I define as *adult* a person older than 12 years old.

[17]ENADID does not ask the year of the first trip to the United States. The classification between migrant and no-migrant households depends on whether at least one migrant made her last trip prior to respectively 1st January 1994 and 2003 or not.

[18]The percentage found by the Mexican Migration Project (MMP) is smaller: Mexican migrants who cross the border without documents correspond to 67% (Munshi, 2003). Since 1982 the MMP carries out surveys in Mexican communities, most of which are located in the western part of the country, where migration to the United States is more prevalent.

children and the personnel who attended the delivery or the abortion.[19] Almost 41% of teenagers and 74% of all women in the sample got pregnant in the last three years prior to the survey. There is statistically significant difference in the mean between migrant and no-migrant households for both samples of teenagers and all women: women in migrant households - especially if less than 20 years old - are less likely to get pregnant, when compared to women in no-migrant households.

[insert TABLE 1.3 here]

Pregnancy can be wanted or unwanted. Unwanted pregnancy is often linked to scarce knowledge – if any – of methods to avoid it. The availability of information on contraceptive methods is the first step towards a control in the occurrence of pregnancy. According to Keyfitz (1989), the increased knowledge of contraceptives, together with their availability, explains between 10%-40% of the decline in fertility only; the rest of the decline is explained by changes in desired fertility - number of children that families want to have - which can be driven by the environment that families face. The health knowledge channel comprises both components - knowledge of contraceptive methods and desired fertility. Information flowing from receiving to sending country and - even more plausibly - behaviours adopted by those living in U.S. and reproduced by those remained in Mexico can play

[19]The year of birth (or death, in case of death after birth) and the year of abortion are recorded. Hence, when we are referring to the occurrence of pregnancy, we are exactly capturing the occurrence of getting pregnant and not only whether the woman gave birth: cases of abortion and especially voluntary abortion are included.

a role in reducing women' - and especially teenagers' - fertility. Exposure to U.S practices likely leads to more knowledge and creates a greater sensitivity to the opportunity costs of raising children and a preference to get pregnant later in time, changing also individual preferences and values of other family members. Those remained in the country of origin know more and know better.

In ENADID women between 15 and 54 years old are also asked detailed information on the knowledge of eleven different methods of contraception, their use, the place where obtaining them and whether the partner agrees on their use or not.[20] This information is obtained only at the moment of the interview. Therefore, it would be hard to infer whether the knowledge is a consequence of the migration experience itself. I cannot exclude that it was obtained before. However, I can show that, controlling for observable characteristics, the two types of households have a different knowledge of contraceptive methods. The most known contraceptive methods are the birth control pill and the condom, reported to be known on average by respectively 72% and 56% of teenagers. 18% of teenagers reports not to

[20]The questions in the survey are the following: 1. Could you please tell me of which contraceptive method/s have you heard about? *Quisiera usted decirme de qué métodos o medios ha oído hablar?* 2. [for every not mentioned] Have you heard about [the name of the method]? *Ha oído hablar de[método]?* 3. Have you ever used [the method] to avoid pregnancy? *Alguna vez usted o su pareja han usado [método] para evitar el embarazo?* Only the first question is here considered to capture the knowledge of contraceptive methods because the methods directly mentioned are most likely those which are really known and used by the woman interviewed.

know any.[21] An initial comparison of migrant to no-migrant households reveals that migrant households know more and this difference is statistically significant. Moreover, almost 70% of sexually active teenagers reports to have used or to use contraceptive methods.[22] Self-reported use of contraceptive methods is often not reliable: people tend to declare a higher use than the real one. Still, knowledge of contraceptive methods and its use are positively correlated. Therefore, I prefer to investigate the impact of migration on knowledge of at least a contraceptive method and infer from that its possible impact on the use. ENADID also gives information about desired fertility.[23] Overall, teenagers desire 2.6 children. When compared to those in no-migrant households teenagers in migrant households desire more. The difference is statistically significant.

[21]The eleven contraceptive practices are: birth control pill, condom, diaphragm or sponge, intrauterine device (IUD), injection, Norplant implant, tubal occlusion, vasectomy, natural family planning, withdrawal and morning-after pill.

[22]Teenagers report to use mainly birth control pill, IUD - intra-uterine device and condoms. A high percentage - especially in the 1997 sample - resort also to withdrawal.

[23]In both surveys the interviewee has been asked: [for mothers] If you would have been back to the time when you have not children and you could choose the number of children you would have had in your life, how many children would you like to have? *Si usted pudiera regresar a la época en la cual no tenía hijos y pudiera escoger el número de hijos por tener en toda su vida, cuántos hijos tendría?* ; [without children] If you could choose the number of children you would have in your life, how many children would you like to have? *Si usted pudiera escoger el número de hijos por tener en toda su vida, cuántos hijos tendría?*

Child health care Data on child health care are considered for women who got pregnant since January 1st 1994 or 2003.[24] The main preventive child care measures used in this paper are number of visits to the doctor of the child during her first year of life, likelihood of the child to be delivered by a doctor and likelihood of the child to receive at least some vaccinations.[25] For each outcome the actual number of observations depends on the amount of missing values.

Unlike for the occurrence of pregnancy, knowledge of contraceptive methods and desired fertility, no remarkable differences across surveys are found for these outcomes of interest. On average, teenagers have their children visited by a doctor 3.4 times in the first year. Almost 96% of teenagers have her child receiving at least some doses of each vaccine, although very few children receive every dose of each vaccine. Above 80% of them has her child delivered by a doctor, suggesting that about 19% of delivers may have been conducted in less safe conditions.

The overall picture shows migrant households' characteristics generally associated with higher fertility rates and lower education levels. Teenagers in migrant households have less educated parents and live in larger family, in a bigger extent located in rural areas. Nonetheless, an initial comparison of migrant to no-migrant households reveals some differences in the outcomes of

[24] Women who went through an abortion are here excluded. The sample includes women who gave birth to a surviving child or to a child who died at or after death. This does not imply that the reduction of the sample by one third corresponds to the cases of abortion. Missing values play here a non negligible role.

[25] The vaccines are tuberculosis (BCG), polio, measles and DPT.

interest which suggest a lower probability of getting pregnant, a higher knowledge of contraceptive methods and a slightly higher child care for teenagers in migrant households. A causal analysis will help to understand which mechanisms play a role.

1.3 Empirical Strategy

1.3.1 Econometric specification

In order to assess the impact of international migration on the outcomes of interest, the following regression function will be estimated:

$$F_{ijst} = \beta_0 + \beta_1 migrants_{jst} + \beta_2 X_{ijst} + \beta_3 S_{st} + \\ + \sum_{s=1}^{32} \delta_s state_s + \sum_{t=1}^{2} \phi_t year_t + \epsilon_{ijts} \tag{1.1}$$

where F_{ijst} stands for the outcomes of interest − occurrence of pregnancy, knowledge of at least a contraceptive method, desired fertility, preventive child health care − for a woman i in household j, state s and year t. $migrants_{jst}$ is an indicator of whether the household j in Mexican state s and year t has a migrant member to the United States. The set of exogenous individual and household characteristics X_{ijst} includes age and age squared, woman's level of education, number of adult males in the household, household size, whether the woman is in a relationship and whether the household is located in a rural area. S_{st} is a vector of observable state-level characteristics that capture

the socio-economical development of the Mexican state of residence. It includes two measures of health care provision taken from CONAPO (2002, 1998), namely number of hospitals per 10,000 inhabitants (by state) and number of doctors per 1,000 inhabitants (by state) in the two years of the survey. $state_s$ is a dummy for the 32 states of residence and $year_t$ is a dummy for the year of the survey (1 stands for 1997 and 2 for 2006). Two additional controls for the level of education of the mother and the presence of a grandmother in the household - or another woman older than 50 years old - are included when the sample is restricted to teenagers. The rationale for these two variables is that teenagers who live with a higher educated mother, or other adult women beyond the mother, may get better supervision and be endowed with a better ability to process information. Furthermore, given that differences between rural and urban areas are expected, the same analysis is conducted by considering the interaction between being in a migrant household and living in rural areas.

$$F_{ijst} = \beta_0 + \beta_1 migrants_{jst} + \beta_2 X_{ijst} + \beta_3 S_{st} + \\ + \beta_4 migrants_{jst} * rural_{jst} + \sum_{s=1}^{32} \delta_s state_s + \sum_{t=1}^{2} \phi_t year_t + \epsilon_{ijts} \tag{1.2}$$

X_{ijst} includes the location of the household.

1.3.2 The endogeneity problem

The parameter of major interest in equation (1) is β_1. Nonetheless, its identification is complicated. Migrants are not randomly drawn from the population: knowledge of contraceptive methods, family planning decisions and health care can be related to the characteristics of migrants themselves. More educated and wealthier people likely tend to have better knowledge of contraceptive methods and be more sensitive to family planning. They can decide to migrate to improve their children's opportunities. Migrants would thus have lower fertility and better health outcomes. Vice versa, if wealthier people decide not to migrate, then migrants result to be the poorest and have higher fertility and worse health than no-migrants. There is a problem of potential selection of individual into migration whose direction is a priori unclear. More formally, in the equation (1) the error term ϵ_{ijts} is composed of two parts:

$$\epsilon_{ijst} = \eta_{ijst} + u_{ijst} \tag{1.3}$$

where η_{ijst} is an unobservable individual term and u_{ijst} is a random term. $migrants_{jst}$ is likely to depend on some factors captured by η_{ijst}. A simple comparison of migrants to no-migrants would therefore incorrectly estimate the gains or loss from migration. In order to separate its effect from the impacts of the selection mechanism, an instrumental variable strategy is employed. The instrument needs to be correlated with the migration decision but uncorrelated with the individual unobservable

attributes.

Foerster (1925) informs us of the historic state-level migration networks from Mexico to the United State of the year 1924. They can be used as instrument for current levels of migration (Hildebrandt and McKenzie, 2005). The assumption is that in a community with high historical levels of migration, a household will have a higher likelihood of having one migrant member than an otherwise identical household living in a community with low initial migration rates. There is substantial evidence that migration networks, formed as a result of U.S. demand conditions and the pattern of development of the north-south railroad system in the early 1900s, play an important role in determining migration from Mexico (among others, Massey et al. (1994); Winters et al. (2001); Munshi (2003)). Those states which were the main sending states at the beginning of the century are still among the principal sending states nowadays.[26]

[insert FIGURE 1.1 here]

Nonetheless, the use of historical migration networks only at the state level may create possible concerns: historical migration networks may be correlated with state level unobservables which could likely affect the state average outcomes. It would be preferable to obtain variations at the household level. By following Mansuri (2006), historical migration networks are thus

[26]Historically migrants have mainly come from the central-western region of the country - Aguascalientes, Jalisco, Michoacán, Coahuila, Zacatecas and Durango. The lowest rates are registered in the southern states of Chiapas, Oaxaca, Yucatán and Guerrero.

interacted with the number of adult males (more than 12 years old) in the household, including those migrated.[27] In the Mexican context, the likelihood that the members of a household undertake migration increases with the number of males in the household. ENADID data suggest that households with a single adult male are less likely to undertake migration (9%) compared to households with three or more adult males (14%), conditional on the same household size. The typical migrants are the sons of the household or his male head.

$migrants_{jst}$ in a reduced form framework is therefore modelled as follows:

$$migrants_{jst} = \alpha_0 + \alpha_1 historicalnetworks_s * noadultmales_{jst} + \\ + \alpha_2 X_{ijst} + \alpha_3 S_{st} + \sum_{s=1}^{32} \sigma_s state_s + \sum_{t=1}^{2} \gamma_t year_t + \eta_{ijts} + v_{jts} \tag{1.4}$$

where $historicalnetworks_s * noadultmales_{jst}$ is the 1924 state migration rate interacted with the number of adult males in the household.[28] It affects the household's propensity to send members abroad but it is unlikely to be correlated with any unobservable household or women attributes that affect the outcomes of

[27] In her study on Pakistan, Mansuri (2006) uses prevalence rates of migration together with census level information on landownership in the village to get within village variation. By interacting the village-land group migration network with the number of adult males in a household, she obtains an instrument which varies at the household level.

[28] For the interaction term $migrants_{jst} * rural_{jst}$ in (2) the instrument $-$ $historicalnetworks_s * noadultmales_{jst}$ $-$ is interacted with the dummy for the location of the household $-$ $rural_{jst}$.

interest. The coefficients of the reduced form (4) with and without controls almost do not change, especially for teenagers. This mitigate concerns that household composition, and precisely the number of adult males in a household, is correlated with unobservables – i.e. labour force participation decisions of the household or female investment in education. Moreover, it is worthy to remind that I am mostly considering teenagers for which it is hard to believe that the number of brothers only, excluding sisters, would affect directly their fertility and health knowledge. It is difficult to imagine that the two scenarios of having one brother and a sister or two brothers and a sister affect differently the fertility and health knowledge of a teenager. Results are reported in columns 1 to 4 of Table 4.

[insert TABLE 1.4 here]

The first-stage results for the instrumental variable estimation show that F-statistics on the incidence of migration are above 10, suggesting that after controlling for the remaining exogenous regressors the historical migration rates interacted with the number of adult males in the household can be a strong determinant of whether a household has currently a migrant member.[26] The coefficients of the instrument are positive, as expected, and highly statistically significant.

[26]The strength of the rejection of the null hypothesis of independence between errors is important in determining the finite sample properties, particularly the bias, of the IV estimator. The Stock and Yogo (2005) test has also been estimated. Here theory does not apply exactly because of heteroskedasticity: the test uses default standard errors in the first-stage regression, while I use robust standard errors. The null hypothesis of weak instruments can be rejected at 5% of distortion.

Estimation methods

The estimates reported in this study are obtained through OLS and standard two-stage least squares estimation (2SLS). The IV estimation method for linear regressions has been preferred to a IV probit estimator also in the case of binary outcomes. It requires fewer distributional assumptions - IV probit requires joint normality - and will however consistently estimate average treatment effects in the case of binary endogenous variable (Newey, 1987; Angrist, 1991). Moreover, IV probit would not precisely estimate the first-stage. It still provides least-squares (LPM) estimates of the first-stage while the binary nature of the endogenous regressor - whether there is at least one migrant member in the household - would require probit estimates.[29]

The IV estimation method for linear regressions has also been preferred to the treatment effect model for the continuous dependent variables used. The latter adds more structure to explicitly account for the binary nature of the endogenous regressor by changing the first-stage model to a latent-variable model similar to the probit and may increase precision of estimation. Nonetheless, its cost is greater chance of misspecification error.[30]

[29] Two separated stages for probit estimates is an alternative. The results are not reported but they are available upon request.

[30] The results are not reported but they are available upon request.

1.4 Results

Table 5 presents the estimation results for the occurrence of pregnancy.

[insert TABLE 1.5 here]

The difference between OLS (columns 1-2) and 2SLS (columns 3-4) estimates indicates a positive selection into migration. By simply comparing migrant households and no-migrant households I would have obtained results driven by the characteristics of migrants' households themselves: migrant households' characteristics are generally associated with higher fertility rates than no-migrant households'. The coefficient estimates for the individual characteristics confirm what expected: the higher the education level of a woman, the lower her chance of getting pregnant - especially for teenagers. Moreover, the higher the education level of her mother, the lower the teenager's chance of getting pregnant. Being in a larger family reduces the occurrence of pregnancy for all women and the presence of a grandmother - or a woman older than 50 years old besides the mother - does not seem to affect the outcome.

Columns 3 and 4 report separately the 2SLS estimates for the sample of all women and teenagers.[31] The results suggest that women in households with at least one migrant member to the United States have a lower occurrence of pregnancy than women in no-migrant households. This is especially true for

[31] The sample includes only women whose partner is living with them. Women whose partner migrated are excluded from the sample.

teenagers, whose reduction corresponds on average to 0.330 probability points.[32] Column 5 controls for the location of the household: whether the household is located in a rural or urban area. The results suggest that there is not statistically significant difference between being in a migrant household in rural areas and being in a migrant household in urban areas.

The reduction in the occurrence of pregnancy can partially be explained by an increase in the knowledge of methods to avoid it. Knowledge is positively correlated with use of contraceptive methods and use of contraception leads to a reduction in the occurrence of pregnancy.[33] Table 6 reports the estimates for the knowledge of at least one contraceptive method.

[insert TABLE 1.6 here]

The dependent variable is a dummy equal to one when the woman knows at least one method to avoid pregnancy. Women in migrant households - especially teenagers - have a higher knowledge of contraceptive methods when compared to their peers in no-migrant households (columns 3 and 4). More precisely, being a teenager in a migrant household increases the knowledge of at least one contraceptive method by 0.389 probability points, on average.[34] There is evidence of the existence of a health informa-

[32]Being in a migrant household reduces on average the occurrence of pregnancy by 0.796 standard deviations for teenagers.

[33]Among the contraceptive methods, costless methods such as withdrawal and natural family planning are also included in the survey. The economic status of the individual should not affect per se the choice to use or not methods of contraception.

[34]Being a teenager in a migrant household increases the knowledge of at least one contraceptive method on average by 1.02 standard deviations,

tion channel which can partially drive the reduction in teenagers' occurrence of pregnancy.

Women's migration is expected to exert a stronger influence on fertility than men's migration. "This expectation is consistent with the emphasis on women's social networks that is found in research on the role of diffusion in contraceptive adoption and fertility decline"(Lindstrom and Saucedo, 2002, p.1346). Therefore, I investigated whether the gender of the migrant, precisely whether the existence of at least one sister who migrates, has an impact on the outcomes of interest. Nonetheless, households with female migrants are only 20% of my migrant sample and few observations can be considered in the analysis. Results are not anymore statistically significant.[35]

Changes in behaviours and preferences driven by the migration experience could also reflect in the number of children desired by a woman. Results of Table 7 suggest that desired fertility for women - especially teenagers - in migrant households increases. Teenagers in households with at least one migrant to U.S. desire at least two children more than peers in no-migrant households.[36]

[insert TABLE 1.7 here]

while for the sample of all women being in a migrant household increases the knowledge of at least one contraceptive method on average by 0.187 probability points (or 0.62 standard deviations).

[35]They are not reported but they are available upon request.

[36]The median number of desired children is three, both for women in migrant and in no-migrant households. For teenagers in no-migrant households the median number reduces to two children, while it remains three in migrant households.

The increase in desired fertility together with the reduction in occurrence of pregnancy suggest that women in migrant households control whether to have children or not. These findings increase my confidence that women - especially teenagers - in these households have a lower occurrence of pregnancy also because they have more knowledge of methods to avoid it. There is an active birth control behaviour. They would like to have more children but likely because of financial constraints – migrants come from the lower middle ranges of the socioeconomic scale – they decide not to have them.

For preventive child health care outcomes, results are more controversial and not statistically significant for teenagers, as the sample is considerably reduced.[37] The only remarkable difference between migrant and no-migrant households can be found for the likelihood of the child to be delivered by a doctor. Being in a household with at least one migrant is estimated to raise the probability that the child is delivered by a doctor by 0.447 probability points, on average.[38]

[insert TABLE 1.8 here]

[37] Outliers are excluded from the analysis (1% of the sample) for the number of visits to the doctor during the first year of life of the child. The number of observations is 18,643.

[38] Being in a household with at least one migrant is estimated to raise the probability that the child is delivered by a doctor by 1.23 standard deviations.

1.4.1 Wealth channel: the role of remittances

Migration is expected to affect fertility and family planning decisions also through the channel of wealth: remittances sent back from migrants change household income and allow households to spend resources on different categories of expenditure such as education and preventive health care. Access to education of adolescent girls may delay the onset of fertility and endow girls with a better ability to process information, potentially increasing knowledge on contraception options.

In ENADID direct information on remittances is collected but there are too many missing values and inconsistent amounts. I do not use this information. A trustable proxy of the money sent back to the country of origin is wealth, but these data are not directly available. I therefore create a wealth index. Wealth is proxied by household durables and utilities and it is defined through the first principal component analysis.[39]

As expected, remittances sent back from migrants increase households' wealth. Results of Table 9 suggest that a woman in a migrant household has a higher wealth index.[40] This finding confirms that migrant households are richer and can plausibly spend

[39]Filmer and Pritchett (2001) show that an index obtained through the first principal component can provide reasonable estimates of the wealth level effects in situations where wealth data are not directly available. The first principal component is the linear combination of the set of variables whose sample variance is greatest among all such linear combinations, subject to a normalization restriction. The single wealth index explains the maximum variation in durables and utilities over the different durables' and utilities' variables.

[40]The wealth index varies from -8.3791 to 6.0967, its mean is equal to 0.0022 whilst its median is 0.67.

more resources on education and preventive health care. Migration affects fertility and family planning decisions also through economical remittances. Both channels, health knowledge and wealth knowledge, play a role.

[insert TABLE 1.9 here]

1.5 Conclusion

Mexican migration to the United States is a longstanding phenomenon which has seen increasing flows of migrants over time and has importantly affected Mexican society. In this paper I investigated the impact of this migration on fertility, knowledge of contraceptive methods and preventive child health care for those remained in the country of origin. The principal channels through which migration affects fertility, family planning decisions and preventive child care are both wealth and health knowledge. In this paper I focus on health knowledge, in the form of information and behaviours that flow from receiving to sending country communities. Migration acts as a channel to transfer behavioural norms from host to migrants' home country.

The main econometric challenge for the paper is the endogeneity of the migrant status, solved by the employment of an instrumental variable approach. Historic migration rates - interacted with the number of adult males in the household - are used as instrument for current household migration to the United States.

Overall, the results suggest that women - especially teenagers - in households with at least one member migrated to the United States have a lower occurrence of pregnancy compared to women in no-migrant households. Evidence of the existence of a health knowledge channel which can partially drive this phenomenon is found: teenagers in migrant households have a significantly higher knowledge of contraceptive methods and likely practice active birth control behaviours. Less significant and more controversial results are obtained for child care outcomes. For teenagers no statistically significant results are found.

The impact of transmission of norms is not easy to measure. A part of the impact I found is undoubtedly related to the characteristics of the context I study and applies especially to those countries experiencing temporary and recurrent migration. However, my results are interesting in a more broad perspective. They suggest specific policy interventions to improve sexual education and preventive infant health-care policies in the country of destination, targeting explicitly migrants.

1.6 Bibliography

Adams, R. H. J. (2005). Remittances, Household Expenditure and Investment in Guatemala. *The World Bank Policy Research Working Paper Series No. 3532*.

Angrist, J. (1991). Instrumental Variables Estimation of Average Treatment Effects in Econometrics and Epidemology. *National Bureau of Economic Research Working Paper No.115*.

Becker, G., J. Duesenberry, and B. Okun (1960). *An Economic Analysis of Fertility in Demographic and Economic Change in Developed Countries*. National Bureau of Economic Research, Inc.

Beine, M., F. Docquier, and M. Schiff (2008). International Migration, Transfers of Norms and Home Country Fertility. *CESifo Working Paper No.2512*.

Bertrand, M., E. Luttmer, and S. Mullainathan (2000). Network Effects and Welfare Cultures. *Quarterly Journal of Economics 115(3)*, 1019–1055.

Bicego, G. and O. B. Ahmad (1996). Infant and Child mortality. *Demographic and Health Surveys, Comparative Studies. Macro International Inc., Calverton, Maryland 20.*

Borjas, G. and L. Katz (2005). The Evolution of the Mexican-Born Workforce in the United States. *NBER Working Paper No. 11281*.

Breierova, L. and E. Duflo (2004). The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers? *NBER Working Paper No. 10513*.

Buvinic, M. (1998). The Costs of Adolescent Childbearing: Evidence from Chile, Barbados, Guatemala, and Mexico. *Studies in Family Planning 29(2)*, 201–209.

Chou, T. (2010). Is Fertility Decline Contagious? Trans-Atlantic Norm Shifts and Fertility in the Post-Vatican II Period. Job Market Paper.

CONAPO (1997). Informacion Basica sobre Recursos y Servicios del Sistema Nacional de Salud. *Salud Publica de Mexico 39(2)*, 766–817.

CONAPO (1998). Boletin de Informacion Estadistica. Recursos y Servicios, 1997. Technical report, Direccion General de Estadistica e Informatica.

CONAPO (2002). Proyecciones de la Poblacion de Mexico, 2000-2050. Technical report, Direccion General de Informacion en Salud. Sistema Nacional de Informacion en Salud. Secretaria de Salud, CONAPO. Mexico.

CONAPO (2007). Programa Nacional de Salud 2007-2012. Technical report, Direccion General de Informacion en Salud. Sistema Nacional de Informacion en Salud. Secretaria de Salud, CONAPO. Mexico.

Cordova, E. (2005). Globalization, Migration and Development: The Role of Mexican Migrant Remittances. *Economia 6(1)*, 217–256.

Cox, D., B. E. Hansen, and E. Jimenez (2004). How Responsive are Private Transfers to Income? Evidence from a Laissez-Faire Economy. *Journal of Public Economics 88(9-10)*, 2193–2219.

Deri, C. (2005). Social Networks and Health Service Utilization. *Journal of Health Economics 24(6)*, 1076–1107.

Devillanova, C. (2008). Social Networks, Information and Health Care Utilization: Evidence from Undocumented Immigrants in Milan. *Journal of Health Economics 27(2)*, 265–286.

Duflo, E., P. Dupas, and M. Kremer (2011). Education, HIV and Early Fertility: Experimental Evidence from Kenya. Working paper.

Duncan, T., J. Strauss, and M. Henriques (1991). How Does Mother's Education Affect Child Height? *Journal of Human Resources 26(2)*, 183–211.

Durand, J., D. Massey, and E. Parrado (1999). The New Era of Mexican Migration to the United States. *The Journal of American History 86(2)*, 518–536.

Durand, J., D. Massey, and R. Zenteno (2001). Mexican Immigration to the United States: Continuities and Changes. *Latin American Research Review 36(1)*, 107–127.

Durand, J., K. William, E. Parrado, and D. Massey (1996). International Migration and Development in Mexican Communities. *Demography 33(2)*, 249–264.

Duryea, S., E. L. Cordova, and A. Olmedo (2005). Migrant Remittances and Infant Mortality: Evidence from Mexico. *Mimeo. Inter-American Development Bank*.

Fargues, P. (2006). The Demogaphic Benefit of International Migration: Hypothesis and Application to Middle Eastern and North African Contexts. *World Bank Policy Research Working Paper 4050*.

Filmer, D. and L. Pritchett (2001). Estimating Wealth Effects Without Expenditure Data or Tears: An Application to Educational Enrollments in States of India. *Demography 38 (1)*, 115–132.

Foerster, R. (1925). *The Racial Problems Involved in Immigration from Latin America and the West Indies to the United States*. Washington, DC: The United States Department of Labor.

Frank, R. and R. A. Hummer (2002). The Other Side of the Paradox: The Risk of Low Birth Weight among Infants of Migrant and Non-migrant Households within Mexico. *International Migration Review 36(3)*, 746–765.

Glewwe, P. (1999). Why Does Mother's Schooling Raise Child Health in Developing Countries? Evidence from Marocco. *Journal of Human Resources 34(1)*, 124–159.

Hildebrandt, N. and D. J. McKenzie (2005). The Effects of Migration on Child Health in Mexico. *Economia 6(1)*, 257–289.

INEGI (2000). Mexican census. Instituto Nacional de Estadistica y Geografia.

INEGI (2005). Conteo de poblacion y vivienda. Instituto Nacional de Estadistica y Geografia.

IPUMS (2010). Tables sda 3.3. - number of own children in the household by origin. CSM, UC Berkeley.

Kaestner, R. and O. Malamud (2010). Self-selection and International Migration: New Evidence from Mexico. *NBER Working Paper No. 15765*.

Kanaiaupuni, S. M. and K. Donato (1999). Migradollars and Mortality: The Effect of U.S. Migration on Infant Survival in Mexico. *Demography 36(3)*, 339–353.

Kovsted, J., C. C. Portner, and F. Trap (2002). Child Health and Mortality: Does Health Knowledge Matter? *Journal of Africa Economies 11(4)*, 542–560.

La Ferrara, E., A. Chong, and S. Duryea (2012). Soap Operas and Fertility. Evidence from Brazil. *American Economic Journal: Applied Economics (forthcoming)*.

Lemay, M. and E. Barkan (1999). *U.S. Immigration and Naturalization Laws and Issues: A Documentary History*. Greenwood Press.

Lindstrom, D. P. and E. Munoz-Franco (2005). Migration and Diffusion of Modern Contraceptive Knowledge and Use in Rural Guatemala. *Studies in Family Planning 36(4)*, 277–288.

Lindstrom, D. P. and S. G. Saucedo (2002). The Short- and Long-Term Effects of U.S. Migration Experience on Mexican Women's Fertility. *Social Forces 80(4)*, 1341–1368.

Macours, K. and R. Vakis (2007). Seasonal Migration and Early Childhood Development. *Social Protection Discussion Paper No. 0702, The World Bank*.

Mansuri, G. (2006). Migration, Sex Bias, and Child Growth in Rural Pakistan. *World Bank Policy Research Working Paper 3946*.

Martin, P. (1998). Migration and Development in Mexico. *Social Science Quarterly 79(1)*, 26–32.

Massey, D., J. Durand, and N. J. Malone (2002). *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*. New York: Russell Sage Foundation.

Massey, D., L. Goldring, and J. Durand (1994). Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities. *The American Journal of Sociology 99(6)*, 1492–1533.

McKenzie, D. and M. Sasin (2007). Migration, Remittances, Poverty, and Human Capital: Conceptual and Empirical Challenges. *World Bank Policy Research Working Paper No. 4272*.

Menjivar, C. (2002). The Ties That Heal: Guatemalan Immigrant Women's Networks and Medical Treatment. *International Migration Review 36(2)*, 437–466.

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market. *The Quarterly Journal of Economics 118(2)*, 549–599.

Newey, W. (1987). Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables. *The Journal of Econometrics 36(3)*, 231–250.

Reynolds, H. W., E. L. Wong, and H. Tucker (2006). Adolescents' Use of Maternal and Child Health Services in Developing Countries. *International Family Planning Perspectives 32(1)*, 6–16.

Rosenzweig, M. R. and T. P. Schultz (1987). Fertility and Investments in Human Capital: Estimates of the Consequence of Imperfect Fertility Control in Malaysia. *Journal of Econometrics 36(1-2)*, 163–184.

Save the Children (2004). *Annual Report*. Save the Children.

UNDP (2007). *Annual Report*. UNDP.

United Nations (2002). *Annual Report*. The Population Policy Data Bank.

United Nations Environment Programme (2010). *Globalis Indicator*. UNEP.

U.S. Census Bureau (2006). U.s. census. U.S. Census Bureau.

U.S. HHS (2007). Child Health USA 2007. Rockville, Maryland: U.S. Department of Health and Human Services, 2008.

Weil, D. N. (2005). *Economic Growth (ch.4)*. Boston: Pearson/Addison-Wesley, c2005.

Winters, P., A. de Janvry, and E. Sadoulet (2001). Family and Community Networks in Mexico-U.S. Migration. *The Journal of Human Resources 36(1)*, 159–184.

Woodruff, C. (2007). How does Migration Affect Local Develoment? What Mexico's Experience Tells Us. In *Immigrants and Their International Money Flows, S. Pozo*. Western Michigan University.

1.7 Tables

Table 1.1: Summary Statistics - *Individual characteristics*

TEENS	(1) Full sample	(2) No-migrant household	(3) Migrant household	(4) Difference (2-3)	(5) p-value
age	17.74	17.76	17.52	0.24	[0.000]
	(1.71)	(1.71)	(1.72)		
education level[a]	2.86	2.85	2.87	-0.02	[0.491]
	(1.00)	(1.01)	(0.98)		
father's education level	2.6	2.65	2.33	0.32	[0.000]
	(1.45)	(1.48)	(1.18)		
mother's education level	2.4	2.46	2.25	0.21	[0.000]
	(1.28)	(1.29)	(1.10)		
couple (=1)	0.38	0.39	0.27	0.12	[0.000]
household size	6.7	6.7	7.1	-0.4	[0.000]
	(3.09)	(3.10)	(3.00)		
monthly household income	3692	3621	4217	-588	[0.002]
(mexican pesos[b])	(7670)	(7932)	(4968)		
rural (=1)	0.65	0.64	0.74	-0.1	[0.000]
No. observations (max)	8158	7281	877		
ALL WOMEN					
age	28.50	28.30	29.40	-1.1	[0.000]
	(9.13)	(9.04)	(9.78)		
education level[a]	2.96	2.98	2.86	0.12	[0.000]
	(1.46)	(1.47)	(1.36)		
father's education level	2.98	3.03	2.64	0.39	[0.000]
	(1.70)	(1.72)	(1.44)		
mother's education level	2.73	2.76	2.5	0.26	[0.000]
	(1.48)	(1.49)	(1.33)		
couple (=1)	0.73	0.73	0.69	0.4	[0.001]
household size	5.99	5.94	6.37	-0.43	[0.000]
	(2.73)	(2.71)	(2.83)		
monthly household income	3720	3661	4209	-548	[0.000]
(mexican pesos[b])	(7188)	(7337)	(5792)		
rural (=1)	0.60	0.59	0.69	-0.10	[0.000]
No. observations (max)	39133	34899	4234		

Standard deviations in parentheses. Columns 5 reports p-value for Student's t-test and test of proportions for dummies.

[a] Level 3 corresponds to junior high school, equivalent to nine years of schooling.

[b] 1 Mexican Peso equal to 0.12697 US dollars in 1997 and to 0.0919 in 2006.

Table 1.2: International migration

Migrant-households		10.15%
Individuals in migrant-households		10.82%
Male migrant		78%
Age of migrant	mean	27.6
	median	25
Relationship in the household	head	29.65%
	son/daughter	47.6%
Illegal migrant		84%
Returnee		53.5%
Destination in US		
	California	36.05%
	Texas	15.55%
	Arizona	7.14%
Reason to migrate		
	looking for a job	74.35%
	found a job	7.66%
Household receiving remittances	international	12.4%
(among migrant-households)	internal[a]	5.5%
Monthly mean remittances	international	2303
(mexican pesos)[b]	internal[a]	982

[a] International remittances are coming from international migration (Mexico-U.S.A). Internal remittances are coming from internal migration, within Mexico.

[b] 1 Mexican Peso equal to 0.12697 US dollars in 1997 and to 0.0919 in 2006.

Table 1.3: Summary Statistics - *Outcomes of interest*

TEENS	(1) Full sample	(2) No-migrant household	(3) Migrant household	(4) Difference (2-3)	(5) p-value
occurrence of pregnancy	0.41	0.43	0.29	0.14	[0.000]
knowledge of at least one contraceptive method	0.82	0.82	0.85	-0.03	[0.055]
desired fertility	2.6	2.58	2.76	-0.18	[0.000]
No. observations (max)	8157	7280	877		
no. visits in 1st year	3.4	3.3	3.5	-0.1	[0.402]
	(2.84)	(2.84)	(2.82)		
delivered by a doctor	0.81	0.81	0.86	-0.05	[0.026]
vaccinations	0.96	0.96	0.96	0.00	[0.53]
No. observations (max)	3362	3107	250		
ALL WOMEN					
occurrence of pregnancy	0.74	0.75	0.68	0.07	[0.000]
knowledge of at least one contraceptive method	0.90	0.90	0.92	-0.02	[0.000]
desired fertility	2.99	2.97	3.19	-0.22	[0.000]
No. observations (max)	39122	34890	4232		
no. visits in 1st year	3.6	3.6	3.6	0.00	[0.739]
	(3.03)	(3.03)	(2.99)		
delivered by a doctor	0.84	0.84	0.89	-0.05	[0.000]
vaccinations	0.97	0.97	0.97	0.00	[0.24]
No. observations (max)	25779	23374	2391		

Standard deviations in parentheses. Columns 5 reports Student's t-test and test of proportions for dummies.

Table 1.4: First stage

	(1) OLS all	(2) OLS teenagers	(3) OLS all	(4) OLS teenagers
Dep.var.: whether there is at least one migrant in the household				
historical migration rates*	0.169***	0.143***	0.087***	0.147***
no. adult males (household)	(0.015)	(0.029)	(0.018)	(0.052)
no. adult males (household)			0.013***	0.006
			(0.002)	(0.007)
age			0.002*	0.024
			(0.001)	(0.080)
age squared			-0.000	-0.001
			(0.000)	(0.002)
education level			-0.002	-0.003
			(0.001)	(0.005)
household size			0.001	0.002
			(0.001)	(0.003)
couple			-0.021***	0.004
			(0.004)	(0.014)
rural			0.042***	0.052***
			(0.003)	(0.013)
mother's education level				-0.009**
				(0.004)
presence of a grandmother				-0.027
				(0.017)
state FEs	yes	yes	yes	yes
year FEs	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes
No. observations	38448	8022	36421	3095
R-squared	0.051	0.063	0.059	0.096
F-statistic on historical migration rates* no. adult males (household)	134.44	24.79	78.66	20.13

Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.5: Occurrence of pregnancy

	(1) OLS all	(2) OLS teenagers	(3) 2SLS all	(4) 2SLS teenagers	(5) 2SLS teenagers[a]
migrants	0.003	-0.018	-0.009	-0.330*	-0.517
	(0.005)	(0.018)	(0.092)	(0.199)	(0.327)
age	0.055***	0.114	0.055***	0.120	0.100
	(0.001)	(0.080)	(0.001)	(0.084)	(0.086)
age squared	-0.001***	-0.002	-0.001***	-0.002	-0.001
	(0.000)	(0.002)	(0.000)	(0.002)	(0.002)
education level	-0.014***	-0.039***	-0.014***	-0.040***	-0.040***
	(0.001)	(0.006)	(0.001)	(0.007)	(0.006)
no. adult males (household)	-0.012***	-0.011	-0.012***	-0.014*	-0.013*
	(0.002)	(0.007)	(0.002)	(0.007)	(0.007)
household size	-0.014***	-0.011***	-0.014***	-0.008**	-0.008*
	(0.001)	(0.003)	(0.001)	(0.004)	(0.004)
couple	0.377***	0.584***	0.377***	0.589***	0.591***
	(0.005)	(0.017)	(0.005)	(0.018)	(0.018)
mother's education level		-0.012**		-0.015***	-0.015***
		(0.005)		(0.006)	(0.006)
presence of a grandmother		0.021		0.013	0.012
		(0.016)		(0.018)	(0.018)
rural	-0.001	0.010	-0.000	0.027	0.001
	(0.003)	(0.014)	(0.005)	(0.018)	(0.030)
rural*migrants					0.244
					(0.265)
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	36421	3095	36421	3095	3095

Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.
[a] Column 5 reports estimates for specification (2). Both *migrants* and *migrants*rural* are instrumented. The instruments are *historicalnetworks*noadultmales* and *historicalnetworks*noadultmales*rural*.

Table 1.6: Knowledge of at least one contraceptive method

	(1) OLS all	(2) OLS teenagers	(3) 2SLS all	(4) 2SLS teenagers	(5) 2SLS teenagers[a]
migrants	0.021***	0.059***	0.187**	0.389*	0.611*
	(0.005)	(0.018)	(0.092)	(0.228)	(0.361)
age	0.018***	0.146	0.018***	0.140	0.164*
	(0.001)	(0.090)	(0.001)	(0.093)	(0.096)
age squared	-0.000***	-0.004	-0.000***	-0.004	-0.004
	(0.000)	(0.003)	(0.000)	(0.003)	(0.003)
education level	0.022***	0.076***	0.023***	0.076***	0.076***
	(0.001)	(0.006)	(0.001)	(0.006)	(0.006)
no. adult males (household)	0.003	0.008	0.003	0.011	0.010
	(0.002)	(0.007)	(0.002)	(0.007)	(0.007)
household size	-0.009***	-0.013***	-0.010***	-0.016***	-0.017***
	(0.001)	(0.003)	(0.001)	(0.004)	(0.004)
couple	0.064***	0.089***	0.067***	0.083***	0.082***
	(0.004)	(0.015)	(0.004)	(0.016)	(0.017)
mother's education level		-0.006		-0.003	-0.002
		(0.005)		(0.005)	(0.006)
presence of a grandmother		0.034*		0.043**	0.045**
		(0.020)		(0.021)	(0.022)
rural	-0.054***	-0.094***	-0.061***	-0.112***	-0.081***
	(0.003)	(0.013)	(0.005)	(0.018)	(0.031)
rural*migrants					-0.291
					(0.280)
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	36413	3094	36413	3094	3094

Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.
[a] Column 5 reports estimates for specification (2). Both *migrants* and *migrants*rural* are instrumented. The instruments are *historicalnetworks*noadultmales* and *historicalnetworks*noadultmales*rural*.

Table 1.7: Desired fertility: number of desired children

	(1) OLS all	(2) OLS teenagers	(3) 2SLS all	(4) 2SLS teenagers	(5) 2SLS teenagers[a]
migrants	0.070***	0.086	1.035*	2.058**	3.356*
	(0.026)	(0.064)	(0.542)	(0.896)	(1.952)
age	-0.013**	0.540*	-0.014**	0.531	0.666*
	(0.006)	(0.300)	(0.006)	(0.339)	(0.381)
age squared	0.001***	-0.016*	0.001***	-0.016	-0.019*
	(0.000)	(0.008)	(0.000)	(0.010)	(0.011)
education level	-0.118***	-0.073***	-0.117***	-0.068***	-0.069***
	(0.005)	(0.020)	(0.005)	(0.023)	(0.023)
no. adult males (household)	-0.095***	0.043*	-0.095***	0.063**	0.058**
	(0.011)	(0.023)	(0.011)	(0.029)	(0.029)
household size	0.091***	0.010	0.083***	-0.009	-0.010
	(0.005)	(0.011)	(0.006)	(0.016)	(0.017)
couple	0.593***	0.243***	0.606***	0.208***	0.201***
	(0.019)	(0.052)	(0.021)	(0.061)	(0.064)
mother's education level		-0.006		0.015	0.018
		(0.016)		(0.021)	(0.024)
presence of a grandmother		0.026		0.093	0.098
		(0.063)		(0.077)	(0.083)
rural	0.341***	0.275***	0.301***	0.169**	0.354**
	(0.015)	(0.047)	(0.028)	(0.074)	(0.148)
rural*migrants					-1.692
					(1.620)
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	35373	3040	35373	3040	3040

Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.
[a] Column 5 reports estimates for specification (2). Both *migrants* and *migrants*rural* are instrumented. The instruments are *historicalnetworks*noadultmales* and *historicalnetworks*noadultmales*rural*.

Table 1.8: Preventive child care

	(1) OLS all	(2) OLS teenagers	(3) 2SLS all	(4) 2SLS teenagers	(5) 2SLS teenagers[a]
Dep.var.: number of visits in the first year of life of the child					
migrants	0.054	-0.268	0.414	3.065	6.302
	(0.077)	(0.289)	(1.630)	(6.589)	(9.039)
rural	-0.652***	-0.336	-0.666***	-0.608	-0.143
	(0.051)	(0.219)	(0.081)	(0.580)	(0.513)
rural*migrants					-4.950
					(5.617)
individual characteristics	yes	yes	yes	yes	yes
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	17377	892	17377	892	892
Dep.var.: whether the child has received some vaccinations					
migrants	-0.007*	-0.018	-0.029	-0.071	0.458
	(0.004)	(0.023)	(0.077)	(0.158)	(0.357)
rural	-0.004*	0.001	-0.003	0.005	0.052*
	(0.002)	(0.014)	(0.003)	(0.015)	(0.029)
rural*migrants					-0.582*
					(0.316)
individual characteristics	yes	yes	yes	yes	yes
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	22239	1023	22239	1023	1023
Dep.var.: whether the child has been delivered by a doctor					
migrants	0.033***	0.027	0.447**	-0.137	-1.044
	(0.007)	(0.033)	(0.182)	(0.393)	(0.727)
rural	-0.109***	-0.128***	-0.124***	-0.116***	-0.195***
	(0.004)	(0.021)	(0.008)	(0.033)	(0.048)
rural*migrants					1.008*
					(0.545)
individual characteristics	yes	yes	yes	yes	yes
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	24304	1099	24304	1099	1099

Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Column 5 reports estimates for specification (2). Both *migrants* and *migrants*rural* are instrumented. The instruments are *historicalnetworks*noadultmales* and *historicalnetworks*noadultmales*rural*.

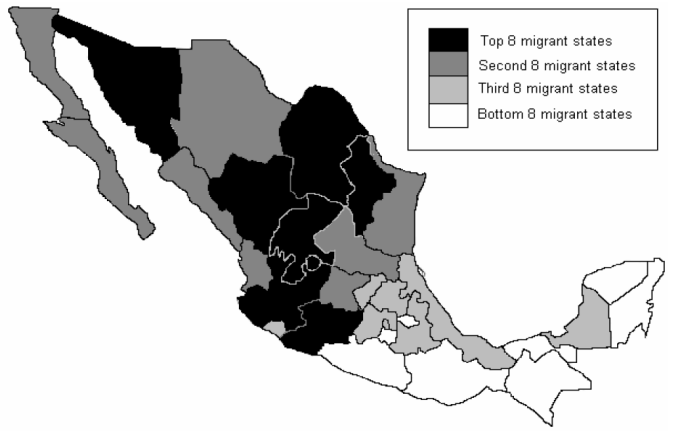
Table 1.9: Wealth index - remittances

	(1) OLS all	(2) OLS teenagers	(3) 2SLS all	(4) 2SLS teenagers	(5) 2SLS teenagers[a]
migrants	0.086**	0.001	2.861***	2.104*	3.039*
	(0.041)	(0.136)	(0.801)	(1.122)	(1.816)
age	-0.007	1.350**	-0.007	1.294**	1.420**
	(0.009)	(0.627)	(0.010)	(0.655)	(0.672)
age squared	0.000*	-0.038**	0.000	-0.036*	-0.040**
	(0.000)	(0.018)	(0.000)	(0.019)	(0.019)
education level	0.185***	0.228***	0.190***	0.234***	0.235***
	(0.010)	(0.046)	(0.010)	(0.050)	(0.049)
no. adult males (household)	0.066***	0.056	0.074***	0.077	0.070
	(0.017)	(0.046)	(0.018)	(0.050)	(0.050)
household size	-0.022***	0.002	-0.043***	-0.016	-0.013
	(0.007)	(0.023)	(0.010)	(0.026)	(0.026)
couple	-0.038	0.002	0.022	-0.046	-0.057
	(0.033)	(0.107)	(0.040)	(0.115)	(0.116)
mother's education level		0.226***		0.245***	0.246***
		(0.044)		(0.046)	(0.046)
presence of a grandmother		-0.079		-0.040	-0.045
		(0.123)		(0.129)	(0.128)
rural	-1.075***	-1.046***	-1.202***	-1.190***	-1.053***
	(0.030)	(0.108)	(0.048)	(0.136)	(0.217)
rural*migrants					-1.249
					(1.680)
state FEs	yes	yes	yes	yes	yes
year FEs	yes	yes	yes	yes	yes
state*year FEs	yes	yes	yes	yes	yes
No. observations	27886	2411	27886	2411	2411

Robust standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.
[a] Column 5 reports estimates for specification (2). Both *migrants* and *migrants*rural* are instrumented. The instruments are *historicalnetworks*noadultmales* and *historicalnetworks*noadultmales*rural*.

1.8 Figures

Figure 1.1: Historical migration rates and Railroad system in early 1900s *Source: Hildebrant N. and McKenzie D.J. (2005) and Woodruff C. (2007)*



Chapter 2

Equal access to education: An Evaluation of the Roma Teaching Assistant Programme in Serbia

(joint with Lara Lebedinski)

Roma constitute a large ethnic minority suffering severe social exclusion, especially in terms of high poverty levels and low educational attainments. This paper investigates the impact of the Roma Teaching Assistant Programme in Serbia in its first year of introduction on the following schooling outcomes: marks, absences and probability to drop out. By using first-hand collected data, we employ two different identification strategies and their combination. First, we exploit the gradual implementation and the intensity of the programme in order to base the evaluation of its impact on a comparison of *Early* and *Late Enrollees*. Second, we compare children exposed to the programme to older cohorts not exposed to it in the same school. We find that, on aver-

age, marks have improved and dropouts have reduced for those children exposed to the programme in their first grade. There is also evidence that overall children exposed to the programme went on average more to school. Higher and more systematic impacts are obtained in schools with a lower number of Roma. We confirm the robustness of our results with placebo tests for the years prior to the introduction of the programme.

2.1 Introduction

Roma are mainly located in South Eastern Europe and with a population of approximately 6 million of people they constitute the largest ethnic minority in the continent (Open Society Institute, 2008).[1] In all countries they suffer severe social exclusion which can be observed in high poverty and unemployment levels, low educational attainments and no participation in the political and cultural life. Roma people are poorer than other population groups and more likely to fall into poverty and remain poor. They have persistent disadvantages in education, including low school attendance and overrepresentation in special schools and schools for adult education.[2] Family permanent incomes are low, also due to big household size and low wages, mainly coming from low skilled jobs. They often lack access to credit and property ownership and are overdependent on social benefits.

The problems of the Roma minority have moved at the forefront of the media's attention especially with the expansion of the European Union. The visa liberalisation and the adhesion to the Union of countries like Romania and Bulgaria, in which the

[1]The number of Roma and the subsequent numbers refer to the following countries: Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Kosovo, Latvia, Lithuania, Republic of Macedonia, Moldova, Montenegro, Poland, Romania, Serbia, Slovakia, Slovenia and Ukraine.

[2]Special schools are schools for children with special educational needs. Schools for adult education were initially introduced with the idea to provide basic literacy knowledge to adult pupils. Nowadays, however, they are mainly attended by pupils who are late at enrolling and by pupils who decided to return to school after dropping out.

percentage of Roma population is high, have indirectly led to significant migration flows towards Western Europe. Appearance of informal settlements, increased number of unemployed and inadequacy of the education system in receiving new foreign pupils are some of the problems which arose in the receiving areas. The extraordinariness of the phenomenon has led to hot discussions within the European countries and civil society. It increased the interest of the European Union in those countries which will likely enter the Union in the future and where a high percentage of Roma population resides, e.g. Serbia. Understanding the impact of programmes targeting Roma is useful not only for these countries but also for the receiving ones, where a minority group becomes also a migrant group.

The aim of this paper is to evaluate the impact of the Roma Teaching Assistant Programme in the first year of its complete introduction. More precisely, we want to examine whether this remedial education programme is effective in reducing dropouts, raising attendance and improving the marks of Roma pupils, by means of a target increase in instruction time, help in homework and assignments and direct link between assistant and parents.

Schooling has always been considered a needed measure to improve living conditions of Roma people and attempts in this direction have been devised by various countries for many years. Half of the Roma population is younger than 18 years old and focusing on children and young people is broadly recognised as a crucial step towards Roma inclusion: higher enrolment rates and better achievement at school are expected to lead to persistent

effects in the labour market and in the reduction of poverty in the long-run. Nonetheless, nowadays the net enrolment rate of Roma in primary education varies among the countries and it is still low, in the range of 40% to 60%. Moreover, students may enrol at the beginning of the year, but may not actually attend school: the percentage of completion rates of primary school is in the range of only 30% to 40% for most countries (Open Society Institute, 2008). However, to the best of our knowledge, there are not systematic studies in economic literature that investigate how to improve life circumstances of Roma, in general, and Roma kids, in particular. This paper is a first solid attempt in this direction and it contributes to the existing literature by providing an accurate overview of the attainments of Roma pupils in Serbian schools, for which so far there were not data available, and by contrasting their achievement to the average Non-Roma pupils. More broadly, it adds evidence on short-term effects of remedial education programmes on minority groups and suggests replicable examples in contexts where minorities suffer low attainment rates and social exclusion. For Roma people this is the case in many other European countries.

The Roma Teaching Assistant Programme (RTA) is the main programme targeting Roma inclusion in education in South Eastern Europe.[3] It began in Serbia in 2002 as a pilot programme car-

[3]The Serbian name of the programme is *Romski Asistenti - Pomoć u Nastavi*. The Serbian Government - together with Montenegro, Croatia, Macedonia, Hungary, Romania, Bulgaria, the Czech Republic and Slovakia - is participating in the *Decade of Roma Inclusion*, an international initiative running from 2005 to 2015 in Central and South-Eastern Europe. The initiative brings together governments, international and non-governmental

ried out by different NGOs and from 2007 to 2009 was led by the Organisation for Security and Cooperation in Europe (OSCE). Since 2009 the Ministry of Education has been responsible for the coordination of the programme, which for the first time had a broad country coverage. Roma assistants - one per each school - participate in regular lessons where they provide additional help to Roma pupils who have difficulties in following classes, especially from lower grades. Moreover, they organise additional lessons, help them with their homework and assignments and once per week they visit their parents. In September 2010 the name of Roma assistants has been changed to pedagogical assistants and their target group is no longer only Roma but all children from marginalised groups.[4] Nonetheless, the Ministry of Education expects that mainly Roma children will benefit from this programme.

We use primary data we collected during 5 months in the Summer-Autumn 2010. We employ two different identification strategies and their combination. First, we exploit the gradual implementation and the intensity of the programme in order to base the evaluation of its impact on a comparison of *Early* and *Late Enrollees*. Second, we compare children exposed to the programme to older cohorts not exposed to it. Our results suggest

organizations to improve the welfare of the Roma population, focusing on healthcare, education, employment and housing. Examples of other programmes which introduce Roma school assistants can be found in Czech Republic, Slovakia, Bulgaria and Croatia.

[4]The programme is now financed by the European Union and it is named *Education for All - Increasing the Availability and Quality of Education for Children from Marginalised Groups*.

that, on average, marks have improved by almost 0.28 standard deviations in Mathematics, 0.35 standard deviation in Serbian and dropouts have reduced by 6.6 percentage points for those children exposed to the programme in their first grade. There is also evidence that all children exposed to the programme went on average at least 17 hours more to school, corresponding to 3/4 days. Higher and more systematic impacts are obtained in schools with a lower number of Roma: the higher is their number, the lower the impact of the programme on the outcomes of interest. This is especially the case for female, for whom being in a school with a lower number of Roma turns out to be more favourable.

The paper is related to three strands of the literature: on remedial education programmes targeting underachieving students, on programmes aiming at improving schooling outcomes of minority communities and at narrowing differences between racial groups, and on programmes aiming at achieving better schooling outcomes of the poor.

Policies targeting low-performing students are generally difficult to evaluate because children with learning difficulties are not randomly assigned to programmes: their characteristics affect both the selection into the programme and its success, making difficult to distinguish between the two effects, especially because the selection mechanism is not typically fully observable. Few studies are able to overcome this problem and find proper counterfactuals: this literature is still scanty and country specific. Among others, Lavy and Schlosser (2005) succeed

in evaluating the effects of a remedial education programme for underperforming high school students in Israel. The intervention prepared students for the matriculation exams and aimed at increasing the school mean matriculation rates. It took place gradually allowing the authors to use as a control group those schools enrolled in the programme later. They found that the programme raised the school mean matriculation rate by 3.3 percentage points. In the United States Hanushek et al. (2002) investigate the effects of targeted programmes on learning-disabled or emotionally disturbed students. They follow those who move in and out of these special programmes and identify programme effectiveness from changes over time in individual performance by comparing academic performance before and after placement into special education. Results suggest that the average special education programme significantly boosts Mathematics achievement of special-education students by roughly 0.1 standard deviations. Jacob and Lefgren (2004) study the effect of summer school and grade retention programmes in the United States by using a regression discontinuity design. They find that their net effect is to improve academic performance in reading and Mathematics among low-achieving students and that these positive effects remain substantial at least two years following the completion of the experience. Another way to overcome the potential selection bias and understand the impact of these programmes is offered by randomised evaluations. An interesting and successful randomized experiment, for instance, has been conduced in schools in urban India by Banerjee et al. (2007). As part of the

programme, underachieving third grade students would have met for two hours each day with an instructor during school hours.[5] Mark of children in schools with remedial education improved in both the first and second year of the programme (Banerjee et al., 2007).

The second strand of the literature this paper is related to is the literature on programmes aiming at improving schooling outcomes of minority communities and at narrowing differences between racial groups. The black-white mark gap has been intensively investigated in the United states. In the past five decades there have been many attempts to close the racial gap even before kids enter school. The first and most known programme is the Perry Preschool programme introduced in 1962: it targeted children from disadvantaged socioeconomic backgrounds and consisted of a 2-5-hour daily preschool programme for children aged three years old and weekly home visits by teachers.[6] Other interventions for disadvantaged families followed such as the Abecedarian Project in the '70s, which provided childcare services for four cohorts of children from infancy through age five, and the Early Training Project, consisting in summertime experiences and weekly home visits during the three summers before entering first grade. Attempts have been also made during the primary school through the introduction of after-school

[5]The remedial classes consisted of 15-20 students and they focused on the core competencies such as literacy and numeracy skills.

[6]Schweinhart et al. (1993) find that students in the programme had higher mark between the ages of 5 to 27, 21% less grade retention or special services required and 21% higher graduation rates.

programmes (Lauer et al., 2006), of merit pay for principals, teachers, and students(Podgursky and Springer, 2007; Roland G. Fryer, 2010), of professional development for teachers (Boyd et al., 2008), and by getting parents to be more involved (Domina, 2005), by placing disadvantaged students in better schools through desegregation busing (Angrist and Lang, 2004) or alter the neighbourhoods in which they live (Jacob, 2004; Sanbonmatsu et al., 2006). The evidence on the efficacy of these interventions is mixed: certain programmes have left the racial achievement gap essentially unchanged. However, according to Roland G. Fryer (2010), racial differences in social and economic outcomes are greatly reduced when one accounts for educational achievement and poverty levels. This points to the fact that there is little empirical evidence for discrimination in the recent data. The same has been concluded by Kertesi and Kezdi (2011) in their study on Roma in Hungary. They find that the gap between Roma and Non-Roma is substantially larger than the gap between African Americans and whites in United States, but that accounting for health, parenting, school and class fixed effects, and family background, the mark gap disappears in reading and decreases by 85% in Mathematics.

The third relevant strand of the literature is on programmes aiming at achieving better schooling outcomes of the poor. This literature suggests that conditional cash transfers, modelled after the Mexican programme PROGRESA, are successful in improving enrolment and attendance in many developing countries. However, policies that promote school enrolment may not pro-

mote learning: early contributions indicate that programmes which are effective at reducing absence from school often do not have an impact on mark of the average student(Schultz, 2004; Miguel and Kremer, 2004). Moreover,Das et al. (2011) show that, although unanticipated school grants lead to significant improvements in student mark, anticipated grants have no impact on them. Analogously, Roland G. Fryer (2010), through school-based randomized trials in schools designed to test the impact of incentives on student achievement, shows that incentives can raise achievement among even the poorest minority students in the lowest performing schools only if the incentives are given for certain inputs, such as reading books, increasing in attendance and students pass. Providing incentives for achievement in mark are much less effective. Finally, only providing school books and other school material or subsidised school meals does not seem to improve students achievements in the case of students with weaker academic backgrounds (Glewwe et al., 2009; Vermeersch and Kremer, 2004).

The rest of the paper is organised as follow. Section 2 gives a general overview of the Serbian education system and summarises the main characteristics of Roma in Serbia. Section 3 gives a description of the Roma Teaching Assistant Programme. Sections 4 and 5 describe our data, the empirical strategy and present our results. Section 6 discusses our findings and concludes.

2.2 The Education System and Roma in Serbia

2.2.1 Primary Education System in Serbia

In Serbia, school is compulsory until the age of 15. Children enrol at primary school if they are aged at least 6.5 years at the start of the scholastic year in September. Since 2007 the attendance of at least 6 months of a cost free preschool programme is compulsory; in 2010 the length of the compulsory preschool has been extended to 9 months.[7]

Primary school consists of 8 years. In the first four grades pupils get one teacher who teaches all compulsory subjects except English, while in the upper four years pupils get one teacher per subject. In the first grade children receive descriptive marks; from the second grade on, the range of marks is 1 to 5 with 1 being the insufficient and worst mark. If a pupil gets at least one insufficient mark at the end of the year, her teacher can decide whether to let her pass to the upper grade or to ask her to take the retake exam in August. In the last few years the Ministry of Education has suggested to schools to reduce repetition rates, especially in the lower four grades.

[7]The obligatory preschool programme has been introduced in order to facilitate the transition to school for children from lower socioeconomic backgrounds. In the initial years the capacities of preschool institutions were not sufficient to enrol all preschool children. Hence, some children, mainly from poorer families or in rural areas, could not be enrolled in preschool. However, due to the lack in the enforcement of the law, they were let to enrol in school also without having attended the compulsory preschool programme.

There are no school fees for primary school, but indirect costs such as books and other school material can pose a considerable cost for some parents.[8] The Ministry of Education aims at reducing the cost of education and the first graders in 2009/2010 are the first generation which received free text books. The plan is that this generation and all younger generations obtain free school books in the future.

2.2.2 Context

Data on Roma in Serbia are inaccurate and scarce. Official census data from 2002 suggest that in the country there are 108,000 Roma, although estimates put forward a number of somewhere between 350,000 to 500,000, approximately 6% of the overall population(Open Society Institute, 2007). Most Roma live in segregated settlements and have considerably different demographic characteristics from the rest of the population. According to the World Bank Living Standard Measurement Survey (LSMS) 2003 - which provides a boosted sample of Roma in Serbia - the average household size of Roma population is of 4.5 household members and thus larger than the national average of 3.2. The average number of children younger than 18 years is 2.4 per Roma households, while the population average is only 0.9. 25% of Roma are younger than 10 and approximately 50% of the Roma population is younger than 23. Consequently,

[8] On average, in Serbia costs associated with school (books and other school material) correspond to almost 2% of yearly household income (LSMS 2003). Based on a survey we conducted in Belgrade, for Roma people these costs account for 6% of their yearly household income.

the average age of Roma is 25, whereas the average age in the country is 42. The percentage of male Roma who declare to have worked over the last week is similar to the national average (69%). Nonetheless, the participation of females is only around 34% and therefore considerably lower than the national average (53%). Overall, approximately 60% of Roma have a consumption below the poverty line and weekly consumption of food per household member in Roma households is half the national average.

Turning to education, 60% of Roma younger than 18 years old have not completed primary education. In contrast, only 20% of overall population do not have a primary school diploma. Out of all children of primary school-age, 30% of Roma do not attend school whereas this is the case for only 1% of the overall population of primary school-age. By using data from the National Assessment Study conducted with third grade students, Baucal (2009) finds that after the first 3 years of school Roma pupils lag 2.2 - 2.5 years behind the average student. Moreover, children from Roma ethnic minority performed worse on standardised tests than Non-Roma children with the same socioeconomic background.

The main barriers of access to education for Roma are: absence of documents, financial constraints, parents' low educational background, child labour, discrimination from teachers and pupils and language barriers.

In the recent years Serbian schools started enrolling children with incomplete documents, but there is still a minor number of

children not able to enrol due to lack of documents. According to the law, the local government is responsible of informing schools and parents that children who reach the school-age in the municipality have to enrol at school. But Roma are often not regularly registered as residents in the municipality and the local government is not able to reach out to some of them. School books and additional school material are a significant burden for the budget of poor families and the most poor among Roma children do not even own adequate clothing for winter months and live in overcrowded homes where they do not have comfortable conditions to pursue their studies. A majority of Roma parents has low educational attainment and this implies that they often cannot help their children with their school work. In addition, some parents attach little value to schooling and education. These reasons together imply that the perceived benefits of going to school are extremely low compared to the respective costs. Moreover, in some cases Roma children help their parents in their work, e.g. they would go with their parents to collect rubbish or they would help them selling goods on the market, or have to take care of their younger brothers and sisters while the parents are working. Also, Roma pupils can face discrimination from teachers and other pupils. There is anecdotal evidence that they are often seated in the last row in classrooms, that teachers do not read their homework and that teachers do not encourage them in their studies. Another problematic issue is they a considerable share of them is sent to special schools. Finally, in a survey conducted by UNICEF - Multiple Indicator Cluster Survey, 2006

- only 10% of Roma declare Serbian to be their mother tongue. As a consequence, children may face difficulties at school due to limited knowledge of Serbian.

2.3 The Roma Teaching Assistant Programme

The Roma Teaching Assistant Programme started as a pilot programme implemented by various NGOs in 2002. In 2007 the OSCE took over its coordination and financing. Since 2009 the programme started to have a country coverage and it is under the coordination of the Ministry of Education. In the scholastic year 2009/2010 there were 48 primary schools which had a Roma assistant: 22 schools started with the programme at different points of time between 2002 and 2007; 26 schools started in 2009. The Ministry expanded the programme to other 77 schools starting from November 2010.

Based on when the programme started in a school, the schools can be divided in two groups: schools which start the programme in September 2009 (*Early Enrollees*) and schools which receive a Roma assistant in the scholastic year 2010/2011 (*Late Enrollees*). The 22 schools which joined the programme between 2002 and 2007 are excluded from our analysis. The selection on these schools was not centralised since they were chosen by NGOs due to their considerable percentage of Roma pupils. For the purpose of our analysis the schools involved in the programme are therefore 26 *Early Enrollees* and 77 *Late Enrollees*.

Both schools and potential Roma assistants had to apply in order to participate in the programme. Among 78 schools which applied in 2009 a commission, representing the government institutions together with OSCE representatives, chose 26 *Early Enrollees* schools based on the following two criteria: first, the percentage of Roma students between 5% and 40% and second, preferably, the availability of preschool programme in the school.[9] The requirements for Roma assistants were knowledge of Romani, secondary school diploma and experience in working with children. 158 applications were received for 26 assistant positions.[10]

In 2010 the programme has been renamed to *Education for all* and starting with the scholastic year 2010/2011 Roma teaching assistants have been renamed to pedagogical assistants. The same selection criteria for the percentage of Roma students as in 2009/2010 applied for further 77 schools out of 252 which entered the programme in 2010/2011 (*Late Enrollees*). The only difference was that in 2010 also schools not offering the compulsory preschool programme could apply for an assistant. The rea-

[9] 64 out of 78 schools which applied had a percentage of Roma between 5% and 40%. Among these 64, OSCE selected 19 schools (out of 26) with a preschool programme, 5 schools (out of 37) with no preschool programme and a school for which no information is available.

[10] The following criteria were taken into account for the evaluation of the assistants' applications: highest level of education completed or enrolled (from 10 to 30 points), experience in working with Roma children (0 to 10 points), experience in working on projects related to education (0 to 10 points), motivation (0 to 10 points), attendance of relevant seminars and/or courses (0 to 10 points), experience as Roma teaching assistant (0 to 10 points), knowledge of Romani (0 to 10 points) and additional points (0 to 10 points).

son is that in 2010/2011 pedagogical assistants were also introduced in 50 kindergartens which offer the compulsory preschool programme. Schools which were not offering the preschool programme could have then been close to kindergartens offering it. The Roma pupil would have been followed by an assistant from her entry in the school anyhow.[11] Selection criteria for now pedagogical assistants remained unchanged and out of the 329 applications for the position, 77 were accepted to work at schools and another 50 were accepted for kindergartens.

We believe that parents were not aware of the existence of the RTA before enrolling their children at school. Data also confirm that *Early Enrollees* were not attracting more Roma students than *Late Enrollees* in the first year of the programme.[12] Therefore, we are confident that our analysis is not affected by possible selection of children into schools.

Every school receives only one assistant. Schools receive from the Ministry a description of her duties, but they are free to decide how to allocate the time of the assistant depending on the need of the school. There is a time allocation suggested. The 30 weekly hours of the assistant could be distributed in the following

[11]Unfortunately we do not have information on the availability of a preschool programme for schools applying in 2010/2011. Nonetheless, it is worthy to recall that some schools without the compulsory preschool programme have also been selected in the previous year.

[12]Roma pupils joining *Early Enrollees* schools in the pretreatment year - 2008/2009 - corresponded to 2.4% of all Roma enrolled in these schools. In *Late Enrollees* they were 2.1%. In the first year of the programme - 2009/2010 - these percentages were respectively 1.6% and 1.3%. Thus, the number of Roma pupils enrolling at school for the first time reduced between the two years and it did it proportionally in both types of schools.

way: work at school (19 hours), work with the local community (8 hours) and writing reports and documentation (3 hours). Activities at school involve both working during regular classes and after-school work. Work with local communities comprises duties such as collecting information about children who did not enrol or who left school, gathering documents for school enrolment, visiting families,[13] cooperation with Roma NGOs, etc. The assistants were advised to work mainly with lower grades, especially the first. Their objectives are making sure that children go to school, preventing them from dropping out and helping them to succeed at school.

In 2009 the Ministry of Education has organised a series of seminars to provide the necessary knowledge and skills to Roma teaching assistants. In total, the assistants have attended 19 working days of seminars and courses in the scholastic years 2009/2010. Regular seminars provided the opportunity to the Ministry to understand the problems of the assistants and guide them through the initial difficulties. In 2010 a set of 9 modules, which all assistants have to attend in their first year of service, has been devised.

[13] In most cases Roma live in segregated settlements so that assistants can go to the settlement and visit several families.

2.4 Preliminary Analysis

2.4.1 Data and Trends of the Variables

We use primary data we collected during 5 months in the Summer-Autumn 2010. They come from administrative records of 23 schools (out of 26) among *Early Enrollees* and 15 schools (out of 77) among *Late Enrollees*.[14] We select the subsample of 15 *Late Enrollees* schools according to the following criteria: firstly, they have to be in the same district of a *Early Enrollees* school;[15] secondly, they have to be in a rural/urban municipality as the nearby *Early Enrollees* school; thirdly, they have to share a similar school size to the nearby *Early Enrollees* school and finally a similar percentage of Roma pupils.[16]

[insert TABLE 2.1 here]

Schools are mainly in Belgrade/Central Serbia and in the South/South-Eastern part of the country, and they are fairly located in both rural and urban areas.[17] Figure 2.1 reports the

[14] In total, there were 26 schools which got an assistant in 2009/2010. In 3 schools we were not allowed to collect data. These schools do not differ from the other schools neither in the number of pupils nor in the percentage of Roma children and they are located in different areas: one in Belgrade, one in Central Serbia and one in the South.

[15] A district is made by more municipalities. In Serbia there are 24 districts and 160 municipalities.

[16] In few cases the school chosen was not available and we needed to select the second option.

[17] 10 schools are located in Belgrade; 8 schools in the central area of the country (5 schools in the municipality of Valjevo and 3 in the municipality of Novi Sad); 12 schools in South-Eastern Serbia (3 schools in the munic-

distribution of schools from which we have collected data. In *pink* municipalities there are only *Early Enrollees* schools; in *green* municipalities there are only *Late Enrollees* schools; and in *dark blue* municipalities there are both *Early* and *Late Enrollees*.

[insert FIGURE 2.1 here]

The data set contains information on 4 scholastic years – from 2006/2007 to 2009/2010 – for the lower four grades of primary school for 18,268 children, both Roma and Non-Roma. It contains for each year and for each pupil the final mark in Mathematics, final mark in Serbian, end of year average and number of hours of absences in a year. For the scholastic years 2008/2009 and 2009/2010 we have also semester outcomes for Mathematics, Serbian, average and hours of absences. The data set contains personal characteristics, such as gender, year of birth, month of birth and place of birth.[18] School specific data include school size, number of Roma - in both school and class - and whether the school is in a urban setting.

Tables 2.2 and 2.3 summarise respectively the averages of the control variables and main outcomes of interest for Roma and

ipality of Jagodina, 2 in Kragujevac, 3 in Kruševac, 3 in Zaječar and 1 in Požarevac); 8 schools in the South of the country (6 schools in the municipality of Leskovac and 2 in the municipality of Niš). We define *urban area* a municipality with more than 35,000 inhabitants.

[18] It is worthy to mention that Roma in Serbia are mainly sedentary: they do not move much within the country. Nonetheless, there is a substantial out-migration, especially towards the European Union, and in the last years in-migration has increased due to the wars in Ex-Jugoslavia. Many Roma refugees in Serbia, for instance, come from Kosovo.

Non-Roma children in the pre- and treatment year.[19]

[insert TABLES 2.2 and 2.3 here]

As reported in Table 2.2, in the pre-treatment year the mean characteristics of the schools that were enrolled in the programme later (column 2) resemble those of the schools that enrolled first (column 1). Boys and girls are equally distributed in schools and pupils were mainly born in the same town where they attend school. The average percentage of Roma per school is 22% in *Early Enrollees* and 19% in *Late Enrollees*. The number of Roma per class is roughly 4/5 children and the class size is 22.16 in *Early Enrollees* and 23.97 in *Late Enrollees*. The two tables show no statistically significant differences between *Early Enrollees* and *Late Enrollees* nor in the student's and school characteristics nor in the outcomes of interest.

This similarity between *Early Enrollees* and *Late Enrollees* schools is found also in the treatment year, providing some support for our claim that *Early Enrollees* and *Late Enrollees* are comparable.[20] The same holds for the assistants: *Early Enrollees* and *Late Enrollees* assistants are comparable in terms of observable characteristics (see Table 2.4).[21]

[insert TABLE 2.4 here]

[19]The same tables are obtained for the years 2006/2007 and 2007/2008 and they are available upon request.

[20]The only significant difference is found for the place of birth of Roma children: there are less migrant children in treated schools.

[21]Unfortunately we have not got demographic information about the assistants besides the gender.

Three important aspects need to be stressed when comparing Roma and Non-Roma children. On a grading scale of 1 to 5, the difference of almost two grades between Roma and Non-Roma pupils in Serbian and Mathematics is very large: for instance, the average in Mathematics for Roma in *Late Enrollees* is 2.37 in 2008/2009 whereas it is 4.17 for Non-Roma; for Serbian it is 2.55 for Roma and 4.33 for Non-Roma. Moreover, dropouts seem to be almost exclusively of Roma children: in 2008/2009 in *Late Enrollees* schools 1.9% among Roma children dropout while among Non-Roma children only 0.06% did it. Lastly, Roma children show to be absent from school approximately three to four times as much as Non-Roma children. In terms of absences, a Non-Roma child is absent from school approximately 40 hours in a year (corresponding to 8 days), while a Roma child misses school somewhere between 118 and 155 hours in a year (corresponding to 24 and 25 days).

By simply comparing the averages of outcomes of pre- and treatment year in the two types of schools, we can see that for Roma children there is a slight improvement in all marks. These effects are larger in *Early Enrollees* than in *Late Enrollees*. Dropouts almost double in the last year in both types of school. The reason for this sharp increase is likely related to the liberalisation of the visa regime with the European Union which induced a certain number of Roma families to migrate to the EU. Finally, absences increase in 2009/2010 in both *Early Enrollees* and *Late Enrollees* for both Roma and Non-Roma, but for Roma they increase by less in *Early Enrollees* schools.

Our data also allow to see whether inequality in marks – the difference between higher and lower marks – decreases as a response to the programme. We use both Roma and Non-Roma children's marks to calculate our Gini inequality index.[22] As usual, under perfect equality the Gini coefficient equals 0; on the other hand, Gini index is equal to 1 when there is total inequality. To calculate the Gini index we use the common formula for income inequality, but we replace the income with pupils' grades.[23] A general improvement in marks of pupils would translate into a reduction of the Gini coefficient and hence a reduction in inequality. Figure 2.2 shows the trends of the Gini coefficient over the four year period – from 2006 to 2010 – in Serbian and Mathematics.

[insert FIGURE 2.2 here]

The graph suggests that inequality is lower in *Late Enrollees* schools for both subjects. An encouraging fact emerging from the graph is that inequality is decreasing in both *Late Enrollees* and *Early Enrollees* schools over the period. Nonetheless, we do not see a sharp reduction in the inequality index among *Early Enrollees* schools compared to *Late Enrollees* schools in the year of the treatment. It is thus unlikely that the programme was effective in reducing inequality in marks in the first year of its implementation.

[22]Similar results are obtained using the Atkinson, Theil index and decomposing the aforementioned indices in within and between-group inequality.

[23]The Gini formula is given by $G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n^2\overline{x}}$, where n is the number of individuals, x_i the mark in Serbian or Mathematics of individual i and $\overline{x}$ the mean mark in Serbian or Mathematics of the whole population.

2.5 Identification Strategy

The aim of this paper is to evaluate the effects of the Roma Teaching Assistant Programme on educational outcomes of pupils in the first four grades of school. More precisely, we want to examine the impact of the programme on dropouts, attendance and grades of Roma pupils in the first year of its implementation. In the following analysis we therefore select only Roma children. We intend to address the following questions:

- Does the programme have an impact on Roma pupils' grades?
- Does the programme reduce dropouts rates of Roma pupils?
- Does the programme increase attendance rates of Roma pupils?

The ideal experiment would require having a random assignment of schools to the programme. Unfortunately, we are not in this setting: schools were not chosen randomly to participate. Nonetheless, the gradual implementation of the programme allows us to base the evaluation on a comparison of *Early* and *Late Enrollees*. Our treatment group is made of schools which started to implement the programme in September 2009 (*Early Enrollees*) whereas the control group is a subsample of schools which got the assistants starting from November 2010 (*Late Enrollees*). There is certainly the concern that schools and assistants starting the programme in two different years may differ because they had to apply in order to get selected. Although

the selection criteria remain almost the same, we do not know what motivates schools to apply before others and whether these motivations are related to differences in the principle or in school quality.[24] Schools which applied in the first year could also apply in the second year. However, some schools which applied in 2009 did not apply anymore in the year after.[25] Thus, if they really were more motivated and of better quality that those applying later, it is hard to understand why they did not want to be part of the programme anymore in 2010. Other schools that applied and meet the requirements in both years got selected only in 2010. They should not differ from those selected in 2009. Moreover, we do know that observable characteristics do not differ between those schools which applied in the first year and those in the second year: schools which applied in 2010 are in the same areas of schools of 2009 and they have almost the same percentage of Roma, on average 10.5% compared to 12.2%.[26] We also do know that the committee which decided the schools selected - composed by the Minister of Education and other representatives of the Ministry, representatives of National Council, OSCE

[24]In both rounds the programme was advertised in newspapers *Politika* and *Prosvetni Pregled*, the last being a newspaper for people working in the education sector; in addition to the advertisement, in 2010/2011 schools' directorates - one directorate may be responsible for more than a municipality - were encharged to inform schools directly.

[25]45% of schools which applied in 2009, and did not get selected, did it again in 2010 and two third of them got selected in the second year (16 out of 24 schools applying in both years). Among these schools, only 2 schools, corresponding to 15%, is present in our subsample.

[26]These are the only information we have got on schools which applied and did not get selected.

and of the Ministry for Human and Minority Rights - gave priority to schools in the poorest municipalities or with huge Roma settlements[27] and rated them based on their shown interest and motivation (application) in the same way, in both years. The same holds for the assistants.

Placebo tests are one possibility to ensure the robustness of our results. Another possibility is to compare older cohorts less exposed to the programme (control group) to younger cohorts (treated group) exposed to the programme in *Early Enrollees*-treated schools.

The advantages and disadvantages of both strategies need to be mentioned. In the first strategy, the control group consists of schools which enrolled later in the programme. The main advantage of this group is that the impact of the programme would not be confounded with other government policies that took place in the year of its introduction. For instance, in 2009/2010 all first grade pupils got free text books and in the last few years the Ministry strongly suggests to schools to reduce repetition rates especially in the lower grades. The disadvantage of this control group is that we are not able to control for unobservable differences which have led some schools to enter the programme before other schools. By using older cohorts in the treatment schools as a control group eliminates our possible problem of selection bias. Nonetheless, this identification strategy relies on the strong assumptions that there were no government interventions over the period - which is not exactly our case - and that the outcomes

[27] Subotica, Novi Sad, Niš, Kragujevac, Belgrade.

have a regular trend over the years. A possible way to better take into account strengths and weaknesses of both approaches is to combine the *Early - Late Enrollees* analysis with the cohort one. We therefore look at the triple difference between cohorts of treated and control schools.

2.5.1 First Approach: Comparison of *Early Enrollees* vs. *Late Enrollees*

Our first identification strategy exploits the fact that some schools received assistants prior to other schools. We compare *Early Enrollee* schools with *Late Enrollees* schools in the years 2008/2009 - when there was no programme - and 2009/2010 - when the programme got introduced.

Average treatment approach

Our specification 2.1 is the difference-in-difference model with school fixed effects:

$$Y_{ijt} = \beta_0 + \delta_t + \rho_j + \beta_1 treatment_j * post_t + \beta_2 X'_{ijt} + \varepsilon_{ijt} \quad (2.1)$$

The outcome variables Y_{ijt} are final marks in Serbian and Mathematics, probability to dropout and hours of absences of individual i, in school j at time t. δ_t is a time fixed effect, ρ_j corresponds to school fixed effects, and $treatment_j * post_t$ is the interaction term between the dummies for treatment status of the school and treatment year. With school fixed effects we are

able to control for time-invariant unobservable school characteristics as well as unobservable geographical characteristics. The control variables X'_{ijt} are school size, school size squared, number of Roma in school, number of Roma in school squared, percentage of Roma per class, class size, class size squared, the gender of the child (=1 if the child is female), age, age squared, and whether the kid is a migrant (=1 if the child was born in the same town where she attends school). The coefficient of interest is the difference-in-difference estimator of the interaction term between treatment and time that captures the difference in outcomes between the treatment and control schools.

Results for the different outcomes of interest are reported in Table 2.5.[28] For all outcomes we estimate the regressions without and with controls (columns 1 and 2). We then split our sample by gender (columns 3 and 4). It is reasonable to expect differences in the impact of government interventions due to different scholastic achievements by gender and different social roles attached to the different sexes in Roma culture. For this reason we also control for the gender of the assistant, but it does not turn out to be significant in any specification.[29] Then we split the sample by grade to examine the presence of possible differential effects (columns 5 to 8). We expect that pupils from

[28]The coefficients for control variables are not reported. Overall, the higher the percentage of Roma in a class, the worse their average marks and the higher their hours of absences. Class size is statistically insignificant in all regressions, but school size turns out to be significant in some specifications and, as we would expect, it has a negative impact on marks. Complete results are available upon request.

[29]Results are available upon request.

the first grade are the most responsive ones for two reasons. First, the assistants work mostly with the first graders: they are lagging less than higher graders. Second, first graders are the ones who do not have any habits about school (e.g. attendance and doing homework), that is, they are the ones who can be influenced most by regular work.[30]

[insert TABLE 2.5 here]

Overall, results suggest that the programme have a statistically significant impact only on hours of absences: pupils exposed to the programme are on average 17 hours (3 to 4 days), corresponding to 0.12 standard deviations, less absent from school than pupils not exposed to it.[31] This is especially the case for male, whose reduction in absences is of 26 hours (5 days) or equivalently 0.18 standard deviation. Marks in Mathematics and Serbian would suggest that the programme have a positive impact on Roma pupils, but the coefficients are not significant. The fact that the inclusion of control variables does not change the magnitude of our coefficients suggests that our coefficients are very robustly estimated.

We exclude that children in *Early Enrollees* and children in *Late Enrollees* schools interact. Only in three cases children

[30] In an attempt to better understand who exactly has been targeted by the programme, we have split our sample based on whether a child has enrolled at school at the right age or at a higher age. We expected to find that children who started school later were targeted by the programme, but we did not find support for this hypothesis.

[31] On average, Roma pupils are absent from school 143 hours (28 days) in a year.

from a control and children from a treatment school live in the same settlement. However, if one were to believe that there are spillover effects from treated children on children from control schools, this would imply that coefficients in our regressions are underestimated.

Intensity of treatment approach

It is worth investigating the intensity of the programme. Each school has only one assistant implying that the higher is the number of Roma children the less intense is the programme. If the assistant has to follow a high number of students, it is likely that she could follow less each of them: she would be less present both in regular classes and in helping them with their homework and assignments.

The following specification is a variation of the previous approach and it exploits the within school variation of Roma and the fact that the programme intensity depends on the number of Roma in a school. Schools are divided in two groups and each group has an equal number of schools.[32] One group contains 19 schools with fewer Roma and the other group contains 19 schools with a higher number of Roma.[33] The main difference to

[32]Similarly, we have tried to divide the schools in four quantiles. However, with four quartiles the data did not pass balance tests for pretreatment year and we have therefore decided to group schools in less - two groups.

[33]The average number of Roma between the two years - pre- and treatment year - is used in order to define the two groups. The threshold is here 50 pupils in order to get a balanced sample. We believe that differentiating the schools in groups helps to better understand the role of the school size on the impacts of the programme. Alternatively, we use the share of Roma

the prior model is that we interact the dummy ($more\_Roma$) with treatment and time. The coefficient of interest is now β_6.

The intensity of treatment is modelled as follows:

$$\begin{aligned} Y_{ijt} &= \beta_0 + \delta_t + \beta_1 treatment_j + \beta_2 treatment_j * post_t + \\ &+ \beta_3 more\_Roma_{jt} + \beta_4 more\_Roma_{jt} * post_t + \\ &+ \beta_5 more\_Roma_{jt} * treatment_j + \\ &+ \beta_6 more\_Roma_{jt} * treatment_j * post_t + \varepsilon_{ijt} \end{aligned} \tag{2.2}$$

Results are in Table 2.6. Again, we look at the impacts by gender.[34]

[insert TABLE 2.6 here]

The intensity of the programme clearly plays a role in explaining its effects. The lower is the number of Roma in a school, and thus the more the assistant can follow them, the higher is the impact on the outcomes of interest. Absences, for instance, reduce on average by 40 hours (8 days) in a year in schools with less Roma, compared to *Late Enrollees* schools. These effects disappear in schools with a higher number of Roma. Marks in both Mathematics and Serbian increase for pupils in *Early Enrollee* schools with a lower number of Roma, but again these effects do not remain in schools with a higher number of Roma. The impacts are especially large for females, for whom being in a school

in each school and we obtain comparable results.

[34] We are not able to split our sample by grades because our data does not pass some balance tests for the pretreatment year if we divide schools in two groups and by grades.

with a lower number of Roma seems to be more favourable: on average, if exposed to the programme in a school with less Roma, their marks in Mathematics and Serbian increase respectively by 0.511 (0.44 standard deviations) and 0.345 (0.29 standard deviations).[35]

Placebo Regressions

The difference-in-difference approach relies on the parallel trends assumption. That is, we assume that, in the absence of the programme, treatment and comparison schools would have had a parallel trend in the average outcomes of interest. The most obvious way to examine the robustness of our results is to estimate the same regressions (specifications (2.1) and (2.2)) for the years 2006/2007 versus 2007/2008 and for the years 2007/2008 versus 2008/2009 (Table 2.9). These two placebo tests allow us to test if treatment and comparison schools are comparable; in other words, in this way we can test if the outcomes in the two groups of schools had a parallel trend before the introduction of the programme. Significant difference-in-difference coefficients in the case of average treatment approach in the years prior to the introduction of the programme would question the adequacy of our comparison group. The two placebo tests suggest that our results are robust.

[35] Their average mark in Mathematics is 2.56 and in Serbian 2.82 in control schools with less Roma.

2.5.2 Second approach: Cohort regressions and triple difference

In our second approach we try to circumvent the problem of possible selection bias by using as control schools older cohorts from treated schools. We compare kids in the first grade (young cohorts) with kids in older grades - second, third and fourth - (old cohorts) in the pre- and treatment year. We know that assistants worked mostly with the first grade and our results from the average treatment approach are suggestive that the first grade has benefitted most from the programme. Hence in this section we use a specification which informs us whether the programme was successful for the children enrolled in the first grade.

We first estimate the following regression for *Early Enrollees*:

$$Y_{ijt} = \beta_0 + \beta_1 young_i + \beta_2 post_t + \beta_3 young_i * post_t + \varepsilon_{ijt} \quad (2.3)$$

where Y_{ijt} are again final marks in Serbian and Mathematics, probability to dropout and hours of absences of individual i, in school j and at time t; $young_t$ is equal to 1 when the child is at the first grade; $post_t$ is equal to 1 in the year of the treatment (2009/2010). The coefficient of interest is now β_3 which tells us how the first graders have performed compared to the older grades.

The same regression (2.3) is then estimated for *Late Enrollees* and the triple difference between treated and control schools and

cohorts is captured by γ_3 in the following specification:

$$\begin{aligned} Y_{ijt} &= \beta_0 + \beta_1 young_i + \beta_2 post_t + \beta_3 young_i * post_t + \\ &+ \gamma_1 treatment_j * post_t + \gamma_2 young_i * treatment_j + \\ &+ \gamma_3 young_i * post_t * treatment_j + \varepsilon_{ijt} \end{aligned} \tag{2.4}$$

The regressions are estimated without and with controls, as in previous specifications. We also look at the impacts by gender. Results are in Tables 2.7 and 2.8. T below the column number in the table stands for the subsample with only *Early Enrollee* schools, while C stands for the subsample with only *Late Enrollee* schools.

[insert TABLES 2.7 and 2.8 here]

When comparing first graders with older pupils in only *Early Enrollee* schools, our coefficients of interest have, with the exception of absences, the correct sign, but are not statistically significant. Results of the triple interaction for the full sample are reported in column 6. Pupils in the first grade exposed to the programme get higher marks than first graders in control schools with respect to their older mates. On average, being in a *Early Enrolles* school increases marks in Mathematics and Serbian by respectively 0.28 and 0.35 standard deviations for first graders. Moreover, it reduces their probability to dropout on average by 6.6 percentage points. The coefficient of interest of the cohort analysis in the *Late Enrolles* schools (β_3 in column 5) suggest that in the absence of the programme dropouts take place more

among first graders than among older graders, confirming the importance for the assistants to focus primarily on this group of students. One might argue that this phenomenon leads to a selection of better pupils in higher grades so that the two groups of young cohorts and old cohorts would not be comparable. Test statistics for the difference in average dropouts between the two groups suggest that it is not the case here.[36] Participation of assistants in regular lessons, organisation of additional classes, their assistance with homework and visits to parents help children to perform better at school.

Surprisingly, hours of absences increase in both treated and control schools: we would have expected them to reduce because of the programme. A possible reason for this finding is that some Roma families have migrated to European countries and have withdrawn their children from school as a consequence of the new visa liberalisation regime. However, we find that the increase in absences is smaller in *Early Enrollees* than in *Late Enrollees* schools, confirming the results obtained by the average treatment approach also in the magnitude of the difference (32 hours or 6 days). Although the coefficient γ_3 of column 6 is not statistically significant, it suggests that, without the programme, the absences would have possibly had an even larger increase. Overall, the effects on the triple difference between cohorts in treated and control schools seem to be mainly driven by differences between schools than differences between cohorts in

[36] The p-value of the test for the difference between the means in dropouts is 0.3860 for the control schools and 0.1593 for the treated schools.

the same school.

Placebo Regressions

We need to control for the robustness of our results by estimating regressions (regression (2.3), (2.4)) for the years 2006/2007 versus 2007/2008 and for the years 2007/2008 versus 2008/2009 (see Table 2.10). The two placebo tests suggest that our results are robust.

2.5.3 Spillover effects - Non-Roma pupils

It remains to investigate whether this programme also affects Non-Roma pupils. We employ both identification strategies and their combination and find that neither marks improved nor absences reduced for Non-Roma students. The presence of a Roma assistant do not improve Non-Roma schooling outcomes. Results are reported in Tables 2.11 and 2.12.

[insert TABLES 2.11 and 2.12 here]

These results, combined together with the previous ones, provide some evidence that the programme is succeeding in reducing the gap between Roma and Non-Roma children, both in school achievements and attendance.

2.6 Conclusion

In this paper we estimate the impact of the Roma Teaching Assistant Programme in its first year of implementation on different

outcomes of interest. We use a difference-in-difference approach. Our first estimation strategy exploits the fact that the introduction of assistants in schools was gradual: some schools entered the programme before others. Given that schools and assistants needed to apply to the programme, a problem of potential selection bias may arise: there are unobservable characteristics, likely related to differences in the principle or in school quality, we cannot control for. In order to circumvent this problem, we use a second identification strategy. We compare pupils of the first grade from treated schools with older cohorts from the same schools. This identification strategy controls well for schools specific characteristics, but we are not able to control for government interventions which might have taken place over the period. Therefore, we combine the two approaches (*Early - Late Enrollees* with cohort analysis) and estimate the triple difference between young and old cohorts in treated and control schools.

Results of our analysis suggest that the programme have a positive effect and started to reduce the gap between Roma and Non-Roma students both in school achievements and attendance. There is evidence that children in treated schools went on average at least 17 hours (3 to 4 days) more to school. Results suggest that, on average, marks have improved by almost 0.28 standard deviations in Mathematics, 0.35 standard deviations in Serbian and dropouts have reduced by 6.6 percentage points for those children exposed to the programme in their first grade. Higher and more systematic impacts are obtained in schools with a lower number of Roma: the higher is their number, the less the

assistant can follow them, and the lower is the impact of the programme on the outcomes of interest. This seems to be especially the case for females, for whom being in a school with a lower number of Roma turns out to be more favourable. Marks of girls in schools with less Roma improve by 0.44 standard deviations in Mathematics and 0.29 standard deviations in Serbian.

While first graders in treated school perform better than their older colleagues, overall the programme does not seem to have a significant impact on pupils' achievement. This is likely due to the fact that assistants work mainly with lower grades and that young cohorts are those really exposed to them. Therefore, the general modest effects should not be interpreted as a failure of the programme. Moreover, this study looks only at its impact in the first year. It is possible that assistants and schools need some time to adjust to the new role of the assistant and that the full benefit from them will come at a later stage. Nevertheless, our results suggest that the programme is more effective in schools with less Roma. We are aware that it is possible that there are systematic differences between schools with a lower and a higher share of Roma. One could argue, for instance, that Roma in schools with a lower share are more willing to adapt and assimilate to the majority population. We cannot be certain that the same effects could be attained in schools with higher percentage of Roma if more assistants were assigned to these schools. However, we believe that it would be worth rethinking to assign more than one assistant to schools attended by a large number of Roma.

2.7 Bibliography

Angrist, J. D. and K. Lang (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *The American Economic Review 94(5)*, 1613–1634.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics 122*(3), 1235–1264.

Baucal, A. (2009). Development of Mathematical and Language Literacy among Roma Students. *Psihologija 39*(2), 207–227.

Boyd, D., P. Grossman, H. Lankford, S. Loeb, and J. Wyckoff (2008). Teacher Preparation and Student Achievement:. *Working Paper no.14314*.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman (2011). School Inputs, Household Substitution, and Test Scores.

Domina, T. (2005). Leveling the Home Advantage: Assessing the Effectiveness of Parental Involvement in Elementary School. *Sociology of Education 78(3)*, 233–249.

Glewwe, P., M. Kremer, and S. Moulin (2009). Many Children Left Behind? Textbooks and Test Scores in Kenya. *American Economic Journal: Applied Economics 1*(1), 112–135.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2002). Inferring Program Effects for Specialized Populations: Does Special Education Raise Achievement for Students with Disabilities. *Review of Economics and Statistics 84*(4), 584–599.

Jacob, B. A. (2004). Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago. *American Economic Review 94(1)*, 233–258.

Jacob, B. A. and L. Lefgren (2004). Remedial Education and Student Achievement: A Regression Discontinuity Analysis. *Review of Economics and Statistics 86*(1), 226–244.

Kertesi, G. and G. Kezdi (2011). The Roma/Non-Roma Test Score Gap in Hungary. *American Economic Review: Papers & Proceedings 101:3*, 519–525.

Lauer, P. A., M. Akiba, S. B. Wilkerson, H. S. Apthorp, D. Snow, and M. L. Martin-Glenn (2006). Out-of-School-Time Programs: A Meta-Analysis of Effects for At-Risk Students. *Review of Educational Research 76(2)*, 275–313.

Lavy, V. and A. Schlosser (2005). Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits. *Journal of Labour Economics 23*(4), 839–874.

Miguel, E. and M. Kremer (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica 72*(1), 159Ű–217.

Open Society Institute (2007). Equal Access to Quality Education For Roma.

Open Society Institute (2008). International Comparative Data Set.

Podgursky, M. J. and M. G. Springer (2007). Teacher Performance Pay: A Review. *Journal of Policy Analysis and Management 26(4)*, 909–949.

Roland G. Fryer, J. (2010). Racial Inequality in the 21st Century: The Declining Significance of Discrimination. *NBER, Working Paper 16256*.

Sanbonmatsu, L., J. R. Kling, G. J. Duncan, and J. Brooks-Gunn (2006). Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment. *The Journal of Human Resources 41(4)*, 649–691.

Schultz, P. T. (2004). School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. *Journal of Development Economics 74*(1), 199–250.

Schweinhart, L. J., H. V. Barnes, and D. P. Weikart (1993). Significant Benefits: The High/Scope Perry Preschool Study Through Age 27. Technical report, High Scope Press, Ypsilanti, MI.

Vermeersch, C. and M. Kremer (2004). School Meals, Educational Achievement and School Competition: Evidence from a

Randomized Evaluation. *World Bank Policy Research Working Paper No. 3523*.

2.8 Tables

Table 2.1: Programme timeline

	2009 Early Enrollees	2010 Late Enrollees
Number of schools applying to the programme	128	252
Number of assistants applying to the programme	158	329
Number of schools (and assistants) joining the programme	26	77
Number of schools in our sample	23	15

Table 2.2: Averages of control variables in pre- and treatment year: Roma and Non-Roma

	Pre-treatment year			Treatment year		
	Early Enrollees (1)	Late Enrollees (2)	Difference (1-2) (3)	Early Enrollees (1)	Late Enrollees (2)	Difference (1-2) (3)
Female:						
Roma	0.5	0.47	0.03	0.49	0.47	0.02
			(0.02)			(0.26)
Non-Roma	0.49	0.48	0.01	0.47	0.49	-0.02
			(0.014)			(0.013)
Born in the same town:						
Roma	0.86	0.81	0.05	0.88	0.81	0.07*
			(0.04)			(0.35)
Non-Roma	0.92	0.91	0.01	0.93	0.92	0.01
			(0.011)			(0.011)
Roma per School	0.22	0.19	0.03	0.19	0.23	-0.04
			(0.06)			(0.06)
School size	305	361	-56	301	363	-62
			(52.96)			(56.04)
No. of Roma per Class	4.91	4.39	0.52	5.25	4.49	0.76
			(1.33)			(1.48)
No. of Roma per Class	5.56	4.64	0.92	5.9	4.6	1.3
(if at least a Roma)			(1.35)			(1.48)
Class size	22.16	23.97	-1.8	22.44	24.21	-1.77
			(1.42)			(1.38)
Number of schools	23	15		23	15	
Number of Roma pupils	1241	811		1268	847	
Number of Non-Roma pupils	4303	3374		4122	3514	

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.3: Averages of outcomes in pre- and treatment year: Roma and Non-Roma

	Pre-treatment year			Treatment year		
	Early Enrollees (1)	Late Enrollees (2)	Difference (1-2) (3)	Early Enrollees (1)	Late Enrollees (2)	Difference (1-2) (3)
Mathematics:						
Roma	2.28	2.37	-0.9	2.36	2.40	-0.04
			(0.05)			(0.05)
Non-Roma	4.25	4.17	0.07	4.3	4.2	0.01
			(0.02)			(0.02)
Serbian:						
Roma	2.43	2.55	-0.12	2.49	2.56	-0.7
			(0.05)			(0.05)
Non-Roma	4.4	4.33	0.07	4.43	4.34	0.09
			(0.02)			(0.02)
Dropout:						
Roma	0.015	0.019	-0.004	0.026	0.035	-0.009
			(0.006)			(0.007)
Non-Roma	0.001	0.0006	0.0004	0.001	0	0.001**
			(0.006)			(0.0005)
Absences (hours):						
Roma	118	125	-7	134	155	-21
			(6.51)			(6.74)
Non-Roma	39	36	3	42	40	2
			(1.02)			(0.97)

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.4: Characteristics of the assistants

	Early Enrollees (1)	Late Enrollees (2)	Difference (1-2) (3)
Female	0.52	0.5	0.02
			(0.18)
Maximum level of education			
Secondary school	0.65	0.58	0.07
			(0.18)
University	0.35	0.33	0.02
			(0.17)
Experience with Roma	0.91	0.67	0.24
			(0.15)
Experience in NGO	0.91	1	-0.9
			(0.06)
Number of assistants	23	12	

Standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.
We could not get information about three assistants of the *Late Enrollees* schools.

Table 2.5: Average treatment approach

	(1) without controls	(2) with controls	(3) female	(4) male	(5) first grade	(6) second grade	(7) third grade	(8) fourth grade
MATHEMATICS								
post	0.051	0.065	0.096	0.041	-0.076	0.040	0.208*	0.168*
	(0.069)	(0.062)	(0.080)	(0.056)	(0.175)	(0.126)	(0.117)	(0.085)
treatment*post	0.046	0.030	0.015	0.053	0.213	-0.085	0.040	-0.156
	(0.081)	(0.077)	(0.091)	(0.085)	(0.221)	(0.192)	(0.146)	(0.158)
SERBIAN								
post	0.039	0.046	0.079	0.027	-0.048	0.099	0.222***	0.109
	(0.060)	(0.048)	(0.055)	(0.050)	(0.132)	(0.133)	(0.079)	(0.076)
treatment*post	0.044	0.012	-0.035	0.058	0.121	-0.130	-0.110	-0.075
	(0.069)	(0.066)	(0.075)	(0.080)	(0.177)	(0.207)	(0.097)	(0.177)
No. observations	4085	3961	1916	2045	989	1111	988	873
DROPOUT								
post	0.017**	0.015**	0.001	0.027**	0.067***	-0.014*	0.011***	-0.007
	(0.007)	(0.006)	(0.010)	(0.012)	(0.016)	(0.008)	(0.004)	(0.011)
treatment*post	-0.006	0.003	0.028*	-0.018	-0.037	0.027*	-0.004	0.045**
	(0.009)	(0.009)	(0.014)	(0.014)	(0.025)	(0.013)	(0.011)	(0.020)
No. observations	4167	4039	1951	2088	1005	1140	1009	885
ABSENCES								
post	31.236***	32.853***	22.456***	42.034***	64.693**	19.900	21.515*	-0.007
	(7.856)	(9.078)	(10.797)	(10.764)	(28.341)	(23.608)	(17.049)	(0.020)
treatment*post	-17.299**	-16.679*	-4.713	-26.119**	-28.336	-19.406	-10.048	0.045**
	(7.856)	(9.078)	(10.797)	(10.764)	(28.341)	(23.608)	(17.049)	(0.020)
No. observations	3980	3868	1871	1997	945	1084	980	885

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.6: Intensity of treatment

	(1) without controls	(2) with controls	(3) female	(4) male
MATHEMATICS				
treatment*post	0.402***	0.339***	0.511***	0.173*
	(0.129)	(0.113)	(0.149)	(0.100)
more_Roma*treatment*post	-0.454***	-0.417***	-0.680***	-0.161
	(0.152)	(0.143)	(0.179)	(0.138)
SERBIAN				
treatment*post	0.303*	0.250**	0.345**	0.149
	(0.150)	(0.121)	(0.159)	(0.113)
more_Roma*treatment*post	-0.330*	-0.307**	-0.500**	-0.109
	(0.170)	(0.150)	(0.186)	(0.151)
No. observations	4085	3961	1916	2045
DROPOUT				
treatment*post	0.001	-0.000	0.053	-0.046
	(0.024)	(0.022)	(0.042)	(0.032)
more_Roma*treatment*post	-0.008	-0.001	-0.044	0.038
	(0.026)	(0.023)	(0.044)	(0.035)
No. observations	4167	4039	1951	2088
ABSENCES				
treatment*post	-28.012	-40.478*	-7.635	-68.816*
	(25.997)	(22.323)	(34.882)	(34.882)
more_Roma*treatment*post	19.312	36.390	9.521	59.759**
	(27.112)	(24.179)	(36.097)	(24.684)
No. observations	3980	3868	1871	1997

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.7: Cohort regression - A

	without controls			with controls			female			male		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
MATHEMATICS	T	C	ALL	T	C	ALL	T	C	ALL	T	C	ALL
young	0.108	0.172	0.213	-0.141	-0.104	-0.077	-0.141	-0.063	-0.105	-0.119	-0.161	-0.043
	(0.146)	(0.127)	(0.152)	(0.150)	(0.145)	(0.132)	(0.134)	(0.224)	(0.215)	(0.251)	(0.230)	(0.171)
post	0.060	0.121*	0.162	0.083	0.106	0.132**	0.116**	0.165*	0.190**	0.067	0.072	0.098*
	(0.048)	(0.061)	(0.103)	(0.050)	(0.062)	(0.054)	(0.049)	(0.087)	(0.071)	(0.070)	(0.062)	(0.058)
young*post	0.070	-0.324**	-0.365**	0.082	-0.241	-0.291*	-0.019	-0.344	-0.426*	0.180	-0.205	-0.239*
	(0.131)	(0.147)	(0.177)	(0.116)	(0.143)	(0.152)	(0.132)	(0.234)	(0.241)	(0.155)	(0.142)	(0.135)
treatment*post			-0.127			-0.059			-0.080			-0.045
			(0.145)			(0.074)			(0.088)			(0.086)
young*treatment			-0.131			-0.102			-0.041			-0.157
			(0.241)			(0.198)			(0.240)			(0.250)
young*post*treatment			0.460*			0.381*			0.428			0.412*
			(0.235)			(0.194)			(0.279)			(0.207)
SERBIAN												
young	0.173	0.290***	0.340***	-0.073	-0.011	0.035	-0.175	-0.035	-0.060	0.040	0.004	0.130
	(0.110)	(0.078)	(0.087)	(0.118)	(0.154)	(0.119)	(0.117)	(0.160)	(0.145)	(0.192)	(0.278)	(0.202)
post	0.047	0.123**	0.173	0.051	0.077	0.118**	0.037	0.104	0.159**	0.082	0.079	0.106**
	(0.053)	(0.056)	(0.110)	(0.056)	(0.050)	(0.045)	(0.051)	(0.072)	(0.063)	(0.081)	(0.049)	(0.041)
young*post	0.058	-0.391***	-0.441***	0.079	-0.255**	-0.300***	0.052	-0.284	-0.352**	0.101	-0.295*	-0.328**
	(0.115)	(0.109)	(0.152)	(0.102)	(0.104)	(0.101)	(0.113)	(0.173)	(0.170)	(0.161)	(0.154)	(0.139)
treatment*post			-0.158			-0.079			-0.131			-0.043
			(0.154)			(0.074)			(0.086)			(0.086)
young*treatment			-0.199			-0.147			-0.123			-0.168
			(0.167)			(0.138)			(0.154)			(0.222)
young*post*treatment			0.531**			0.382**			0.416*			0.423*
			(0.212)			(0.149)			(0.215)			(0.212)
No. observations	2462	1624	4086	2395	1567	3962	1180	736	1916	1215	831	2046

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.8: Cohort regression - B

	without controls			with controls			female			male		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
DROPOUT												
young	-0.020***	-0.014*	-0.012**	-0.002	0.011	0.002	-0.002	0.013	0.009	-0.005	0.008	-0.008
	(0.004)	(0.007)	(0.006)	(0.006)	(0.014)	(0.007)	(0.013)	(0.017)	(0.017)	(0.010)	(0.015)	(0.009)
post	0.003	-0.004	-0.002	0.015**	-0.002	-0.003	0.021*	-0.019	-0.016	0.009	0.010	0.007
	(0.006)	(0.006)	(0.007)	(0.006)	(0.010)	(0.006)	(0.010)	(0.012)	(0.010)	(0.009)	(0.015)	(0.014)
young*post	0.030**	0.080***	0.078***	0.014	0.079***	0.080***	0.031	0.071***	0.077***	-0.003	0.087***	0.083***
	(0.014)	(0.024)	(0.023)	(0.013)	(0.019)	(0.018)	(0.020)	(0.019)	(0.018)	(0.012)	(0.028)	(0.028)
treatment*post			0.005			0.018**			0.038**			0.003
			(0.009)			(0.009)			(0.015)			(0.016)
young*treatment			-0.009			0.005			-0.003			0.015
			(0.006)			(0.006)			(0.016)			(0.012)
young*post*treatment			-0.047*			-0.066***			-0.047*			-0.087***
			(0.027)			(0.022)			(0.027)			(0.029)
No. observations	2509	1658	4167	2438	1601	4039	1200	751	1951	1238	850	2088
ABSENCES												
young	-21.835***	-1.674	-0.141	3.123	37.687*	18.656	-8.373	39.502	26.379	12.221	32.323	8.644
	(6.679)	(14.939)	(13.725)	(8.355)	(20.554)	(17.503)	(11.240)	(23.527)	(21.174)	(14.033)	(23.473)	(17.397)
post	6.732	19.031**	20.564	10.022	30.724***	20.728**	7.655	21.190*	8.570	13.007	37.766***	30.488***
	(5.386)	(8.406)	(13.534)	(5.975)	(9.912)	(7.856)	(7.172)	(10.707)	(8.771)	(8.865)	(11.015)	(9.396)
young*post	37.800***	45.788*	44.255*	23.579*	57.493**	54.639**	40.707**	67.618**	67.603**	9.740	52.423**	49.480**
	(11.139)	(24.209)	(24.838)	(11.537)	(21.263)	(22.235)	(16.000)	(28.938)	(26.469)	(14.055)	(22.036)	(24.287)
treatment*post			-14.801			-10.409			-0.785			-17.036
			(18.481)			(10.354)			(11.898)			(12.891)
young*treatment			-22.663			-0.868			-19.597			16.848
			(17.445)			(16.384)			(19.622)			(16.766)
young*post*treatment			-5.487			-31.867			-28.524			-40.337
			(28.803)			(24.945)			(30.772)			(27.852)
No. observations	2394	1586	3980	2336	1532	3868	1152	719	1871	1184	813	1997

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.9: Average treatment approach - Placebo

	2006/2007 and 2007/2008			2007/2008 and 2008/2009		
	(1)	(2)	(3)	(4)	(5)	(6)
	all	female	male	all	female	male
MATHEMATICS						
treatment*post	0.080	0.093	0.067	-0.057	-0.102	0.007
	(0.066)	(0.075)	(0.080)	(0.077)	(0.098)	(0.080)
SERBIAN						
treatment*post	0.059	0.141	-0.025	-0.094	-0.103	-0.053
	(0.080)	(0.105)	(0.103)	(0.077)	(0.093)	(0.072)
No. observations	3585	1750	1835	3816	1876	1970
DROPOUT						
treatment*post	-0.003	0.006	-0.010	0.015	0.014	0.014
	(0.009)	(0.017)	(0.016)	(0.010)	(0.025)	(0.014)
No. observations	3640	1776	1864	3897	1897	2000
ABSENCES						
treatment*post	0.955	-2.002	3.566	9.558	2.921	14.935
	(12.592)	(17.435)	(14.753)	(13.864)	(19.448)	(10.602)
No. observations	3542	1732	1810	3788	1850	1938

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.10: Cohort regression - Placebo

	2006/2007 and 2007/2008			2007/2008 and 2008/2009		
	(1)	(2)	(3)	(1)	(2)	(3)
	T	C	ALL	T	C	ALL
MATHEMATICS						
young*post	0.164	0.117	0.086	-0.101	-0.115	-0.113
	(0.237)	(0.180)	(0.185)	(0.145)	(0.197)	(0.196)
young*post*treatment			0.122			-0.022
			(0.286)			(0.239)
SERBIAN						
young*post	0.030	0.143	0.105	-0.167	0.034	0.043
	(0.207)	(0.194)	(0.203)	(0.178)	(0.154)	(0.156)
young*post*treatment			-0.024			-0.244
			(0.278)			(0.231)
No. observations	2232	1354	3586	2364	1482	3846
DROPOUT						
young*post	0.003	0.008	0.011	-0.020	-0.036**	-0.037***
	(0.017)	(0.017)	(0.015)	(0.015)	(0.012)	(0.013)
young*post*treatment			-0.010			0.019
			(0.024)			(0.020)
No. observations	2259	1381	3640	2389	1508	3897
ABSENCES						
young*post	-14.021	-56.385*	-54.237*	-9.425	-13.207	-13.596
	(17.068)	(27.822)	(28.643)	(13.600)	(16.765)	(18.055)
young*post*treatment			34.591			10.100
			(32.454)			(23.240)
No. observations	2203	1339	3542	2331	1457	3788

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.11: Average treatment approach - Non-Roma

				placebo 2007/2008		
	(1)	(2)	(3)	(4)	(5)	(6)
	Mathematics	Serbian	Absences	Mathematics	Serbian	Absences
post	0.011	-0.011	5.025***	0.058**	0.052**	-0.518
	(0.018)	(0.031)	(1.185)	(0.027)	(0.025)	(2.874)
treatment*post	0.029	0.054	-1.586	-0.043	-0.029	3.749
	(0.025)	(0.034)	(1.725)	(0.033)	(0.029)	(3.106)
No. observations	14981	14982	14686	15345	15345	15052

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.12: Cohort regression - Non-Roma

	Mathematics			Serbian			Absences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	T	C	ALL	T	C	ALL	T	C	ALL
young*post	0.044	-0.062	-0.070	0.033	-0.056	-0.066	0.658	3.178	3.259
	(0.058)	(0.085)	(0.079)	(0.072)	(0.077)	(0.069)	(2.682)	(3.274)	(3.057)
young*post*treatment			-0.010						
			(0.010)						
No. observations	8232	6749	14981	8232	6750	14982	8099	6587	14686
				placebo 2007/2008					
	Mathematics			Serbian			Absences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	T	C	ALL	T	C	ALL	T	C	ALL
young*post	-0.099*	-0.056	-0.053	-0.118**	-0.089	-0.087	0.610	-1.718	-2.440
	(0.050)	(0.104)	(0.099)	(0.047)	(0.101)	(0.097)	(2.847)	(3.922)	(3.935)
young*post*treatment			-0.049			-0.037			2.787
			(0.111)			(0.107)			(4.986)
No. observations	8573	6772	15345	8573	6772	15345	8435	6617	15052

Robust standard errors corrected for clustering at the school level are reported in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

2.9 Figures

Figure 2.1: Location of the schools with assistants

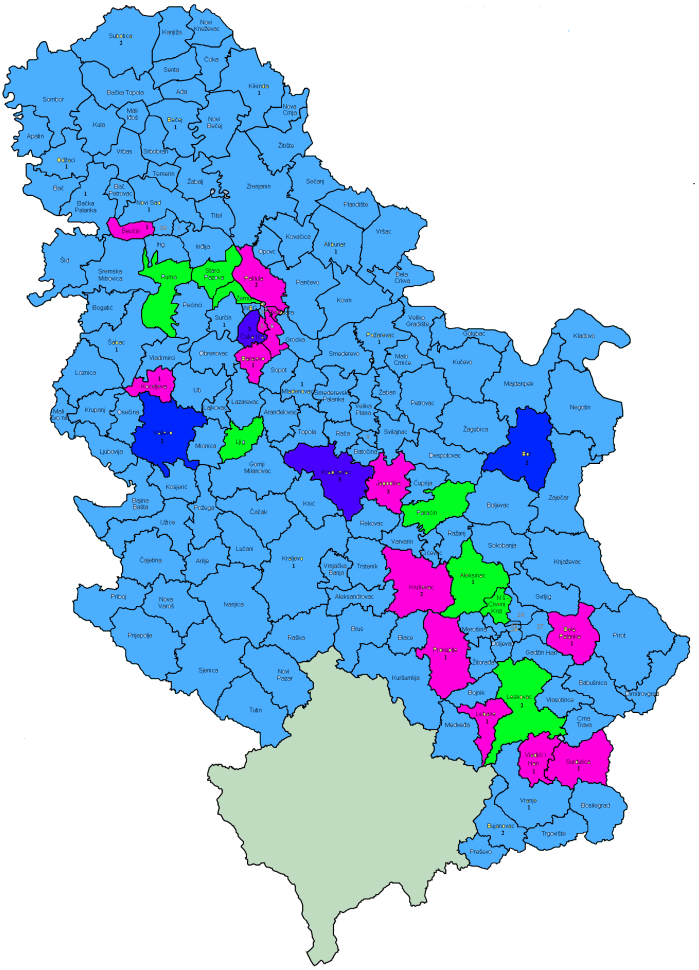
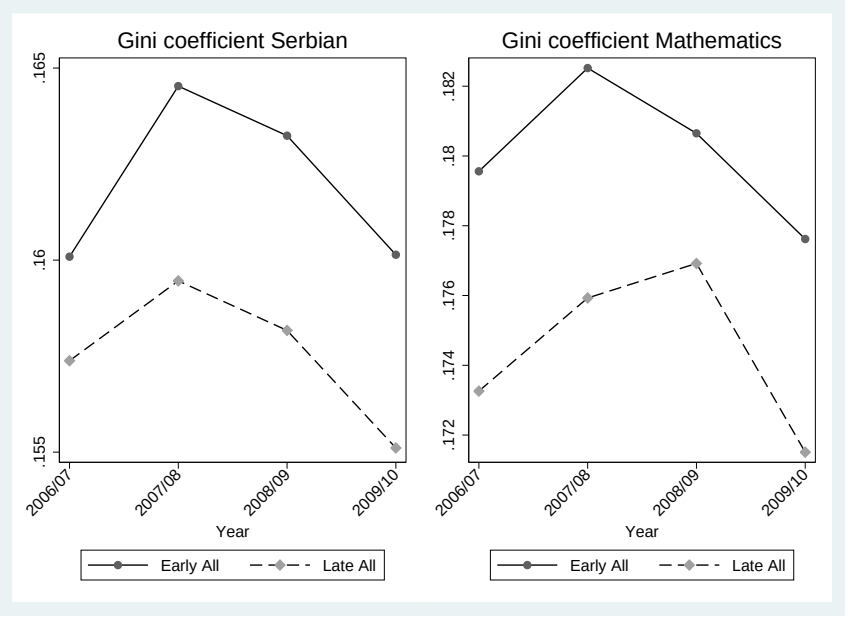


Figure 2.2: Gini coefficient of inequality in Serbian and Mathematics



Chapter 3

The Curse of Low Aspirations: Remedial Education and Perceived Returns to Education of Roma People

(joint with Lara Lebedinski)

We examine how a remedial education programme for primary school-age children affects parental aspirations about their children's future. Using original survey data we collected in Serbia, we investigate whether expectations on labour market perspectives and educational achievement change as a consequence of exposure to the Roma Teaching Assistant Programme. We argue that these changes are likely to occur mainly through a role model mechanism: in the programme all the assistants are Roma and from the same social background of the pupils they help. The presence of a person belonging to the same community, who

showed to be successful, motivates parents to believe their children can succeed. Our results show that parents of pupils in treated schools expect higher returns to education for their kids. They are also more likely to expect them to achieve a secondary level of education.

3.1 Introduction

Aspirations for the future consistently affect choices made in the present. The social environment where one lives plays a role in shaping aspirations. The aim of this paper is to examine the impact on parental aspirations of a remedial education programme for primary school-age children targeting the marginalised Roma minority group. We investigate whether expectations on labour market perspectives and educational achievement change as a consequence of exposure to the Roma Teaching Assistant Programme (RTA), a remedial education programme introduced in Serbia in 2009. We argue that these changes are likely to occur through a role model mechanism. In the RTA Programme, all the assistants are Roma and from the same social background of the pupils they help. They can share their successful experience with them and their parents who will be motivated to believe that their children can achieve analogous results. Households may thus respond to changes in perceived returns when making schooling decisions and increase current investment in education.

Data on education referring to Roma people are striking. In most countries enrolment rates in primary school among Roma are in the range of 40% to 60%. Their completion rates are even lower: only 30% to 40% of Roma adults have completed compulsory primary education.[1] Underinvestment in education can

[1]There are reasons to believe that these numbers are upper bounds. First, some schools keep children who do not come to schools in their school books. Second, a large number of Roma finish evening schools or special schools which count as finished primary schools although the requirements in these schools are much lower. Special schools are schools for children

be due to financial constraints. However, Roma people may not invest in education because they do not expect school to give them enough future opportunities. In the formal job market there is often discrimination for people coming from minority groups. The informal job market does not often require education. Roma people are mainly involved in casual and seasonal jobs, performed without any written contract, i.e. they collect rubbish, sell goods on the market or do low skilled jobs. The cost of investing in education is too high as compared to the discounted stream of expected future benefits. Therefore, there is no incentive to invest. A vicious cycle of low aspirations arises.

For the purpose of our analysis, we have conducted an extensive survey with 300 Roma households in Belgrade. In Autumn 2010 we interviewed both parents and their children in 12 different settlements of the city. The RTA Programme started in 2009 and we look at its impact after a year of implementation. The programme was introduced gradually: some schools received their teaching assistant before others. Parents and children who attend schools receiving the teaching assistant in Autumn 2009 are our treated group. Parents and children who attend schools that received the teaching assistant at a later point in time (Autumn 2010) are our control group.[2] The allocation of Roma teaching assistants was not randomised: schools and assistants

with special educational needs. Schools for adult education were initially introduced with the idea to provide basic literacy knowledge to adult pupils. Nowadays they are mainly attended by pupils who are late in enrolling and by pupils who decided to return to school after dropping out.

[2] All schools involved in the programme are public schools.

needed to apply to be part of the programme. We therefore need to tackle the possible problem of selection bias.

We find that parents whose children were exposed to the programme expect higher returns to education for their kids. They are also more likely to expect them to achieve a secondary level of education. Moreover, an examination of heterogeneous effects suggests first that our results on highest expected level of education are driven by responses from Non-Muslim parents and parents living in *mixed (Roma and Non-Roma)* settlements. Second, parents revise their expectations in response to the programme mainly for younger kids (6 to 10 years).

The importance of role models for minorities is not new in the education literature. A series of researchers and policy makers in the 90s was pushing for an increased hiring of minority teachers in the US (Graham, 1987; Ladson-Billings, 1994). In fact, the importance of having a teacher with the same background has been found significant in improving the achievement gap for minorities (Dee, 2004). However, to the best of our knowledge this is the first study addressing changes in aspirations of minority groups arising as a consequence of such an experience. Only Krishnan and Krutikova (2010) evaluate the long-term effects of an after-school programme for children living in slums in Bombay and find rather weak evidence on expected life evaluation and aspirations. Nonetheless, they do not look at a minority group and elicit directly from children their role models, whereas we argue that assistants of the RTA Programme are perceived as such. Our paper, together with its companion paper (Battaglia

and Lebedinski, 2011), adds evidence on short-term effects of remedial education programmes on minority groups. It also suggests replicable examples in contexts where minorities suffer low attainment rates and social exclusion. For Roma people, for instance, this is the case in many other European countries and so far there are few attempts to investigate how to improve their life circumstances in general and of children in particular. Furthermore, we contribute to the existing literature by providing primary data in a context where data are scarce.

Our paper is in line with the contributions of Ray (2004), Genicot and Ray (2010), Nguyen (2008) and Beaman et al. (2012). We know that individual desires and their standards of behaviour depend, in part, on past experience and the observations of their peers. In societies where the poor do not observe someone with their similar background becoming rich, downward mobility and underinvestment in education is expected. A reasonable distance between one's current living standards and where one wants to be motivates to believe one can succeed. Our paper is also linked to the strand of literature on subjective expectations and information gap between perceived and actual returns to schooling.[3] Standard economic theory suggests that, in the presence of perfect information, individuals choose their level of education by equating the marginal benefits of education to its marginal costs. Underinvestment in education can be due

[3]Literature suggests that this gap can be filled also by providing additional information through statistics (Jensen, 2010). These tools turn out to be most cost−effective solutions than incentives, like cash transfers or private school vouchers.

to credit constraints, high discount rates and low school quality.[4] However, several works emphasised the importance of subjective expectations (Manski, 1993; Jensen, 2010; Nguyen, 2008; Kaufmann and Attanasio, 2009). Schooling decisions are affected by returns perceived by individuals, and these perceptions may be inaccurate, due to limited or imperfect information.

In the background of our work there is the literature on residential segregation and neighbourhood effects that studies the relevance of neighbourhoods and one's peers in influencing socioeconomic outcomes.[5] For instance, segregation of the African Americans has been identified as one of the reasons for the persistence of inner city poverty in the US (Cutler and Glaeser, 1997). Moreover, the neighbourhood where one lives can clearly affect one's labour market (Clark and Drinkwater, 2002; Edin et al., 2003; Bayer et al., 2008; Boeri et al., 2011) and educational outcomes (Card and Rothstein, 2007). Lastly, the ethnic composition of a municipality can be important for the quality of local public goods such as schools (Alesina et al., 1999; La Ferrara and Mele, 2006).

The rest of the paper is organised as follows. Section 2 gives information on the background and the way the survey has been designed. Section 3 describes our data and provides some descriptive statistics. Section 4 presents the estimation strategy and our results. Section 5 discusses findings and concludes.

[4]See Glewwe (2006) for an extensive summary on education in developing countries.

[5]For an excellent review of the literature on neighbourhood effects see Durlauf (2004) and Blume and Durlauf (2006).

3.2 Background: Context, Roma Teaching Assistant Programme and Survey Design

3.2.1 Roma in Serbia

The Roma people are the largest ethnic minority in Europe and have been experiencing discrimination for centuries in all the countries where they live.[6] They are poorer than other population groups and more likely to fall into poverty and remain poor. They suffer of severe social exclusion in terms of overrepresentation among low skilled jobs and no participation in the political and cultural life and this is persistent over time.

Official data on Roma in Serbia are scarce and inaccurate.[7] Roma people often do not declare themselves as Roma in surveys. Most of them consider themselves both Roma and Serbian and the question of nationality allows only one answer.[8] Thus, the 2002 Census counts 108,000 Roma, corresponding to 1.44% of the total Serbian population, while estimates suggest a number between 350,000 and 500,000, approximately 4-6% of the overall

[6] The Roma people are mainly located in South Eastern Europe: Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Czech republic, Hungary, Kosovo, Republic of Macedonia, Moldova, Montenegro, Romania, Slovakia, Slovenia and Ukraine.

[7] This is the case for most Central and Eastern European countries where the majority of the Roma population lives.

[8] The most appropriate approach when asking for one's identity would be to allow for multiple identities, but this approach has been rather uncommon in this type of surveys.

population (Open Society Institute, 2007).

The Living Standard Measurement Survey (LSMS) from 2003 provides rich information on the living conditions of the Roma population in the country. It is important to note that this survey includes only Roma living in segregated settlements, which according to the 2002 Census is the case for 83% of the Roma population. The numbers from the LSMS are alarming. Two out of three Roma households are poor implying that their average consumption is below the absolute poverty line.[9] Almost half of the Roma population (40%) is younger than 18 years old[10] and only 71% of children from Roma settlements aged 6 to 15 attend school. Among the adults, 25% have no schooling at all and another 36% have not finished primary school.[11] Conversely, 99% of Non-Roma of the same age are enrolled in school and only 13% of adults have not completed primary school. The employment rate among Roma males is very similar to that of the Non-Roma

[9]The percentage of the extremely poor among the Roma interviewed is 11.9%. Those who are considered extremely poor are those who cannot satisfy even their basic needs for food.

[10]The average age of Roma people is 25, whereas the average age in Serbia is 42. The average number of children younger than 18 years old is 2.4 per Roma households, while the population average is only 0.9. The average household size of Roma population is of 4.5 household members; the national average of 3.2.

[11]In Serbia, school is compulsory until the age of 15. Children enrol at primary school if they are aged at least 6.5 years at the start of the scholastic year in September. Since 2007 the attendance of at least 6 months of a cost free preschool programme is compulsory; in 2010 the length of the compulsory preschool has been extended to 9 months. Primary school consists of 8 years. In the first four grades pupils get one teacher who teaches all compulsory subjects except English, while in the upper four years pupils get one teacher per subject.

population (69%), but the female employment rate is very low with 34% versus 53%. The LSMS confirms that Roma live in difficult conditions and that they constitute a marginalised minority.

Data on education and job market are in line with those of other countries. They underinvest in education due to credit constraints and the existence of barriers of access to education. Roma people often lack of required ID and face financial constraints. On average, in Serbia costs associated with school (books and other school material) correspond to almost 2% of yearly household income (LSMS 2003). In our sample of Roma people, they correspond to almost 6% of yearly household income.[12] Some children face difficulties at school due to language barriers, they are engaged in child labour and suffer of discrimination from teachers and pupils.[13] However, there are many reasons to believe that the lack of goals and aspirations is an important factor influencing the educational decision of Roma people. First, a large percentage of them live in segregated settlements. Since they are isolated from the mainstream society, they do not often have different models to which they can relate to in their immediate neighbourhood (this idea is line with the

[12]For 10% of them these costs even ranged between 12% to 25% of yearly household income.

[13]Some children have a limited knowledge of Serbian: in a survey conducted by UNICEF - Multiple Indicator Cluster Survey, 2006 - only 10% of Roma declare Serbian to be their mother tongue. Moreover, Roma pupils may face discrimination from teachers and other pupils in schools: they are often seated in the last row, teachers do not read their homework and do not encourage them in their studies. Frequently they are also sent to special schools with consequences in future employment opportunities.

argument of Wilson (1987)). According to the 2002 Census, in Serbia 83% of self-declared Roma live in census tracts with at least 7% of the Roma population. Second, it is extremely rare that Roma people perform jobs for which high levels of education are required. For instance, in Serbia there are usually no teachers of Roma origin working in schools. They can barely be found in any public office.[14] Third, there is evidence that the mean earnings of Roma workers are lower than those of Non-Roma workers, especially for higher levels of education. Figure 3.1 reports descriptive data for the city of Belgrade. We needed to use two sources of data because the Serbian Statistical Office only collects data for jobs in the formal sector usually not performed by Roma. There are no official data on wages of Roma people given they mainly perform informal activities. Data for Non-Roma are obtained from the Serbian Statistical Office and data for Roma from our sample. The figure intends to simply provide a picture of the context and the differences do not pretend to be indicative of the exact amounts.

[insert FIGURE 3.1 here]

Nonetheless, we argue that a policy intervention would be successful in increasing Roma aspirations because they underestimate the outcomes of investing in education. Among Roma there are high differences in average earnings between one education level and another. In our sample, for instance, average

[14] In our sample only 36 women (out of 487) and 27 men (out of 427) in working age perform jobs with a full time contract, in the formal sector.

wages with secondary education are 27% higher than with primary education for boys and 21% for girls, and average wages with primary education are 29% higher than with uncompleted primary for boys and 21% for girls.[15] Moreover, the higher is the education level achieved, the better are the job market perspectives, both in terms of type of contract and place where to perform the job. Data from LSMS (2003) reported in Figure 3.2 suggest that these differences are substantial for Roma living in Serbia. The top panel presents their types of contract by education level. 90% of boys and 80% of girls with a secondary education level have got a written contract, while almost none works without a contract. Conversely, among those with only primary school almost 30% perform their activities without a contract. The bottom panel reports the places where jobs are performed. The percentage of those who work in the street or in flee market reduces drastically with the level of education. This is even more evident for girls. The pattern for office/factory as the place where to work is the opposite.

[insert FIGURE 3.2 here]

Therefore, conditional on the fact that Roma people's earnings are lower than Non-Roma's ones, there is still room for improvement based on education among Roma. The more one studies the higher are the wages one gets and the better are the conditions to perform the job. It is crucial to understand whether

[15]For Non-Roma the average wages with secondary education are 49% higher than with primary for boys and 60% for girls. For Non-Roma we use 2011 data for the city of Belgrade obtained from the Serbian Statistical Office.

parents are aware of the actual returns to schooling. We could not conduct a baseline survey before the introduction of the programme. Thus, we need to look at data of parents not affected by the programme. Figure 3.3 reports the distributions of expected returns to education for parents whose children attend the schools that received the assistant in the second year of the RTA. Their averages are given by the solid lines. The dashed lines correspond to average wages of people in our sample by education. There are few women who completed primary and, especially, secondary school. Thus, results for females are less trustable. Official data do not provide this information. The first panel reports the expected wage distributions, conditional on not having achieved any level of education. These distributions are more concentrated on the right of the dashed lines of actual average returns. Parents expect for their children higher returns when no level of education is achieved. The second and third panels of the figure report the expected wage distributions, conditional on having achieved a primary and a secondary level of education, respectively. For male these distributions tend to be more concentrated on the left of the dashed line of actual average returns. Parents expect for their male children less than what people with that education levels actually earn. There is limited or imperfect information and this likely fosters low aspirations for Roma people.

[insert FIGURE 3.3 here]

3.2.2 The Roma Teaching Assistant Programme

The Roma Teaching Assistant Programme is the main programme in Central and Eastern Europe targeting Roma inclusion in education.[16] After the initial pilot phase, the programme took off on a larger scale in the scholastic year 2009/2010. In Autumn 2009, 26 schools (*Early Enrollees*) entered the RTA programme. In the following year an additional 77 schools (*Late Enrollees*) joined. Each school is assigned one teaching assistant. On average, the number of Roma per school is 86 (13% of total pupils enrolled) and those directly helped by the assistant are 27 - almost one third - especially among younger grades. Although schools are somewhat free in allocating the time of the assistant, her major tasks are helping children during regular classes and organising after-school extra classes. One day per week the assistant visits parents of children who are not going to school and informs other parents about their children's progress.

The RTA programme is not a randomized experiment: schools and assistants had to apply in order to be part of the programme. Schools were chosen on the following two criteria: (1) a percentage of Roma pupils between 5% and 40%, and (2) preferably the availability of a preschool programme in the school.[17] The

[16]For a more extensive description of the programme see Battaglia and Lebedinski (2011).

[17]Information on the existence of a preschool programme are available only for the 78 schools applying in 2009. For the 252 schools applying in 2010 this information was not required anymore. In that year 50 assistants were assigned to kindergartens which offer themselves preschool programmes. Schools which were not offering the preschool programme could have then been close to kindergartens offering it. The Roma pupil would have been

requirements for assistants were the following: (1) secondary school attainment, (2) knowledge of Romani and (3) preferred experience in working with children.[18] It is not explicitly stated that the assistant needs to be Roma: only the knowledge of their language is required. However, all of them are of Roma origin. The selection criteria remain the same in both years and schools and assistants which applied in the first year and do not get selected could also apply in the second year. Observable characteristics do not differ between schools which applied before and schools which applied later. The percentage of Roma pupils is 13.6% in *Early Enrollees* and 13.7% in *Late Enrollees* schools.[19] Unfortunately we do not know what motivates some schools to apply before others and whether these motivations are related to differences in the principle or in school quality.[20] This might arise a problem of selection bias. Our estimates can be overesti-

helped by an assistant from her entry in the school anyhow. Since 2007 the attendance of at least 6 months of a cost free preschool programme is compulsory; in 2010 its length has been extended to 9 months.

[18]In 2009 among 158 people applying, 26 were selected; in 2010, among 329 people, 77 got the job (and 50 more became assistants in kindergartens). All assistants live in the same municipality of the school they work for. Among those belonging to the same municipality, detailed criteria, based on level of education attained, motivation and experience in working with children, were used to rank assistants.

[19]This is the only information we have got, together with their size, for schools that applied and did not get selected. *Early Enrollees* schools count on average 792 pupils and *Late Enrollees* 894.

[20]In 2009/2010 the programme was advertised in newspapers *Politika* and *Prosvetni Pregled*, the last being a newspaper for people working in the education sector; in 2010/2011 schools' directorates - one directorate may be responsible for more than a municipality - were encharged to send applications directly to schools.

mated: parents' aspirations can be correlated with the quality of the school. If children are going to better schools, parents may reasonably expect for them better educational achievement and labour market perspectives, unconditional to the programme.[21] However, some schools which applied in 2009 did not apply anymore in the year after.[22] Thus, if they really were more motivated and of better quality that those applying later, it is hard to understand why they did not want to be part of the programme anymore in 2010. Other schools that applied and meet the requirements in both years got selected only in 2010. They should not differ from those selected in 2009. In our sample of Belgrade, among the 4 schools which got the assistant in 2010, one did also apply the year before. We believe that the selection mechanism does not bias substantially our results. Furthermore, schools selected in the first year are not different in observable characteristics from schools selected later. The same holds for the assistants. Table 1 reports the data in our sample.[23] We collected data from 9 schools in Belgrade: 5 schools received the assistant in 2009; 4 schools in 2010.[24]

[21] Average marks of previous years in *Early Enrollees* schools do not suggest they are better schools. Yet, we cannot exclude differences in the principle's motivation.

[22] In the whole country less than half of schools and assistants which applied in the first year and did not get selected applied again in 2010.

[23] The same holds for the whole sample of schools involved in the programme in Serbia. In the RTA the schools selected in the first year are not different in observable characteristics from the schools selected later. The same holds for the assistants (Battaglia and Lebedinski, 2011).

[24] In Belgrade 6 schools got selected in the RTA programme in the first year of its implementation. One school did not provide us the list of students so it is excluded from our sample. 8 schools were selected in 2010. We have

[insert TABLE 3.1 here]

Early Enrollees and *Late Enrollees* schools count a similar number of Roma per class, 4.43 and 5.23, and a similar percentage of Roma per school, 20% and 22%. The sex composition among Roma is the same: 52% of students is female in *Early Enrollees* schools and 46% in *Late Enrollees* schools. 40% of Roma are born in Belgrade in *Early Enrollees* schools and 32% in *Late Enrollees* schools. Schools slightly differ only in class size: *Early Enrollees* schools have smaller classes, with 22.24 students versus 25.63 of *Late Enrollees* schools. The characteristics of the assistants in the two types of school are also comparables. Almost all of them are female with experience in NGO. In *Early Enrollees* schools 40% of the assistants got a university degree; in *Late Enrollees* schools 33%.[25]

3.3 Data and Descriptive Statistics

3.3.1 The Survey Design

We use first-hand collected data obtained through a survey conducted with 300 Roma households in 5 municipalities of Bel-

got detailed administrative data from a subsample of 4. For the remaining 4 we only know the percentage of Roma per schools. They are comparable to those in our subsample.

[25]Among assistants in *Late Enrollees* schools there is one missing value for the information on the maximum level of education. This explains why the categories *secondary school* and *university* do not sum to 1.

grade.[26] The survey took place in Autumn 2010, one year after the implementation of the programme in *Early Enrollees* and before *Late Enrollees* schools received the assistant. In 2010 schools received the assistant in November/December. The households in our sample have children who were enrolled in both types of schools. Pupils were randomly selected from lists of students provided by the same schools.[27] Figure 3.4 displays a map of Belgrade with the 12 settlements were the survey was carried out. Settlements with the numbers 1 to 5 are the ones where assistants were introduced in 2009/2010, that is the settlements with children from *Early Enrollees* schools. Settlements 6 to 13 had assistants starting from 2010/2011.

[insert FIGURE 3.4 here]

We are interested in the effect of the RTA programme on children from the lower four grades, given that the assistants mainly work with them. Our sample is constructed in such a way that all households have at least one child in the lower four grades of primary school in the scholastic year 2009/2010.

Three sets of questionnaires are administered in the survey: a household questionnaire providing information on the household and community characteristics, a questionnaire for the mother or caretaker and a questionnaire for the children. The *mother questionnaire* consists of an extensive series of questions about

[26] The five municipalities are Voždovac, Zvezdara, Zemun, Palilula, and Čukarica.

[27] The response rate is 93.46%: 321 households have been contacted and 300 answered. Households were not compensated for their participation.

the education of the children aged 6 to 15 living in the household. In this section some questions on child labour are also asked. Children who attended first to fourth grade of primary school in 2009/2010 are asked information about their school and teachers. The *child questionnaire* also contains quick tests on children's abilities to read and write and to do some mathematics.

3.3.2 The sample

Our sample is divided in two groups. The first group consists of 122 households with children enrolled in 5 schools which got a Roma teaching assistant in 2009/2010. These households are selected randomly from lists of students provided by schools and correspond to the treated group.[28] The 178 remaining households were randomly selected from settlements in Belgrade close to 8 schools which received the RTA programme in 2010/2011 and they are our control group.[29] The number of households selected from each settlement is proportional to the size of settlement.

We consider the whole household to be treated if at least one child goes to a school with an assistant in the first year of the implementation of the programme. We do expect that parents'

[28]The lists contain pupils who are enrolled in the schools in year 2009/2010. Students who dropout are therefore included in the sample.

[29]In Belgrade 6 schools received the assistant in 2009/2010 and 9 schools in 2010/2011. In our sample we have 5 schools out of 6 which received the assistant in 2009/2010 and 8 out of 9 among those which received her in 2010/2011. We were not allowed to collect data in the schools excluded. These schools are not different in characteristics from those belonging to the sample.

aspirations are created at the household level: once a child is exposed to the programme, expectations on future opportunities change for all children of the same household. Table 3.2 reports the characteristics of treated and control groups.

[insert TABLE 3.2 here]

They are comparable in terms of observable characteristics. Their differences in means are not statistically significant in most of the cases. Wealth, monthly income, educational attainments and household composition do not differ between groups.[30] Households are equally located in rural and urban areas.[31] 32% of households in the control group and 31% in the treated group have at least one member working in the informal sector. However, treated households are more in only Roma settlements (28% versus 16%) and among non-treated households there are significantly more Muslim (80% versus 57%). It would be worthy to investigate whether the programme impacts differently depending on the religion and the type of settlement (see section 4.3 on heterogeneous effects).[32]

Identification requires the absence of selective sorting into treatment. Field analysis suggests that parents were not aware

[30] Rank among siblings is significantly higher among treated households. Nonetheless, we do not believe this would be problematic given that household composition does not differ between the two groups.

[31] We define urban area a municipality with more than 35,000 inhabitants.

[32] Overall, the characteristics of our sample are in line with official data. Only, somewhat surprising with respect to them, few households have both parents with unfinished primary school (7%) and a relatively large share of households has at least a parent with finished secondary school (19%).

of the existence of the programme before enrolling their children at school. Data also confirm that *Early Enrollees* were not attracting more Roma students than *Late Enrollees* in the first year of the programme.[33] Therefore, we are confident that our analysis is not affected by possible selection of children into schools.

3.3.3 Outcome variables

We use three different sets of questions to understand whether the programme is effective in changing parents' aspirations about their children. We focus on the expectations of parents because we believe that at such a young age (6 to 15) the aspirations of parents are more relevant for a child's educational attainment and more reliable for expected returns to education than her own aspirations.[34]

The first and second set of questions relates to expected re-

[33]Roma pupils joining *Early Enrollees* schools in the pretreatment year - 2008/2009 - corresponded to 2.4% of all Roma enrolled in these schools. In *Late Enrollees* they were 2.1%. In 2009/2010 these percentages were respectively 1.6% and 1.3%. The number of Roma pupils enrolling at school for the first time reduced between the two years and it did it proportionally in both types of schools.

[34]However we ask pupils which is the highest expected level of education they expect to achieve. Two-thirds of parents answer as their children, whereas one-third expect their kids to achieve a lower education level than children expect. Only in few cases parents expect more than their children and this happens mainly when children perform well at school. Discrepancy in answers between parents and children is mainly found among poorer and larger families, living in *only Roma* settlements, Muslim and with lower levels of education. Pupils also perform worse at school than their classmates. Results are not reported, but they are available upon request.

turns to education. They are asked to either mother or father (or caretaker), however the mother is the main interviewee in 92% of cases. They are asked for the oldest boy and the oldest girl in the household.[35]

The first set of questions considers expectations about the likelihood of getting a job given a certain level of education achieved. We ask "Assume that your oldest boy has finished primary (or secondary) school - and that is his highest degree - and he is 25-30 years old: how certain are you that he will have any kind of job?". We ask the same question for the oldest girl. Although we are mainly interested in the probability of finding a job given a secondary education level, we use for comparison reasons also the probability of finding a job given a primary school education level.[36] The responses to this question come from a five point Likert scale and they are "Absolutely sure", "Quite sure", "Maybe", "Unlikely" and "No, s/he will not find a job". The Likert scale has a disadvantage: different respondents can interpret the scale differently so that other factors such as optimism or education affect the response. Alternatives such as

[35]In the pilot survey we asked the questions for each child but we realised that there was no variation in the responses between the children of the same sex. As a consequence we decided to pose this question only for the oldest male and for the oldest female child. In only 6% of cases the oldest child may be older than 15 and thus not enrolled in a compulsory school. However, we believe that this is not a problem in our setting given the hypothetical nature of the question and the fact that the treatment is considered at the household level. In the worst case this would imply that our results are underestimating the real effect.

[36]The same question is asked for the case in which the child does not finish primary school. Results are not reported and available upon request.

explaining probabilities to interviewees and asking them to express their expectations using a cardinal scale are suggested by the literature (Delavande et al., 2009). However, due to the low educational level of our respondents, this drawback could not be overcome and we decided to offer them the possibility to choose among five different responses.[37] For the purpose of our analysis, we converted the five Likert scale outcomes to a dummy variable. If the respondent declared that it is unlikely or that her child will not find a job given a certain education level, we set the probability to zero. In the other three cases – "Absolutely sure", "Quite sure", "Maybe" –, we set the probability to one. We believe that by aggregating the categories to a dummy we do not loose important information: almost two-thirds of respondents answered "Unlikely" and "No, s/he will not find a job" in the case of primary school and one third in the case of secondary.[38]

[insert FIGURE 3.5 here]

The second set of questions elicits minimum and maximum amounts parents expect that their children will earn once employed. We ask "Assume that your oldest boy has finished primary school (or secondary) and this is his highest degree and he

[37] Between 1% (male) to 5% (female) of households did not answer these questions.

[38] Ordered logit analysis suggests that some categories may not be collapsed (see Table 3.9 in Appendix). For instance, while for secondary school "Absolutely sure" and "Quite sure" can be clearly collapsed, they should not be in the case of primary school. For this reason, we also keep the variable as categorical. Estimates with categorical outcomes are available upon request. They confirm the results reported.

is 25-30 years old. Think about the kinds of jobs he might be doing in this case. What do you think is the minimum amount he can earn per month? And the maximum amount?" The interviewees have been asked explicitly to take into account both regular and irregular types of income. The same questions have been asked for girls. We obtained the minimum and maximum earnings and we use their (log) average as our measure of expected earnings.[39]

The third relevant outcome is the highest expected education level of the child. The exact question is: "What level of formal education do you think that *(name)* will complete?" We create the dummy "secondary as the highest level of education" that takes the value 1 when it was answered "secondary (or more)" and 0 otherwise. The question is asked for each child between 6 and 15 years old.[40] For consistency we also estimate the impact with a reduced sample, corresponding only to the oldest boy and the oldest girl in the household.[41] Summary statistics for the outcome variables in our sample are reported in Table 3.3 and suggest a possible positive impact of the programme on both expected salary and level of education achieved. In treated households expected future earnings are higher than in control households for both primary and secondary level of education.

[39] 3% of households did not answer in the case of questions referring to a male child, while these percentage is around 15% for a female child. Estimates with minimum and maximum earnings are available upon request.

[40] The median number of children aged 6 to 15 per household is 2. There are many missing values for this outcome of interest. This explains why our sample is as big as with the other outcomes.

[41] Results are not reported, but they are available upon request.

Respondents in treated households are also more likely to expect their children to finish secondary school. This difference is only significant for boys.

[insert TABLE 3.3 here]

3.4 Estimation Strategy and Results

In the Roma Teaching Assistant Programme all the assistants are Roma and from the same social background of the pupils they help. They can act as role models for the kids they work with. The presence of a person sharing her successful story should affect children's and their parents' aspirations about their future in two ways. First, we expect that treated parents think that also their children can succeed. They would be more likely to find a job and a better job with higher salaries, conditional on achieving a specific education level. We do expect the highest impacts the higher is the education level attained. The better jobs are obtained with a higher education level. It is worth investing in education. Second, as a consequence they would more likely expect their children to finish secondary school.

3.4.1 Early Enrollees versus Late Enrollees

We estimated the impacts of the RTA programme on returns to education with the following specification:

$$Y_j = \alpha_0 + \alpha_1 treatment_j + \alpha_2 X_j' + \varepsilon_j \tag{3.1}$$

where Y_j corresponds to the outcomes of interest for the household j: likelihood of finding a job with primary school as the highest degree achieved, likelihood of finding a job with secondary school as the highest degree achieved, (log) mean amount of earnings per month with primary education and (log) mean amount of earnings per month with secondary education. $treatment_j$ equals 1 whether there is at least one child in the household who goes to a treated school. X'_j includes household wealth per capita, whether there is someone in the family who works in the informal sector, whether the household lives in a urban area and in a Roma settlement, whether the household is Muslim, the maximum education level of parents and household composition characteristics. For the outcome "secondary school as the highest expected level of education", a second specification is introduced:

$$Y_{ij} = \beta_0 + \beta_1 treatment_j + \beta_2 X'_{ij} + \nu_{ij} \tag{3.2}$$

$treatment_j$ is defined as above. X'_{ij} also includes age and age squared of the child, her gender, rank among siblings, demeaned mark in Mathematics and Serbian of the previous scholastic year.[42] Standard errors are clustered at the cohort times school level. Regressions are estimated separately for boys and girls because we are interested in the effects for each gender.[43] We

[42]The marks are demeaned from the average school marks (among Roma). For children in their first grade, the average school marks are used.

[43]It is worth investigating whether the gender of the assistant may affect differently boys and girls for our outcomes of interest. Results do not suggest that aspirations change depending on the sex of the assistant. This seems

do also report results with the pooled sample in Table 3.10 in Appendix.

Results for the probability of finding a job, expected earnings and highest expected education level are reported in Table 3.4. For consistency, all the estimates are OLS. Probit estimates for the two dummy outcomes are not reported but they are available upon request.

Columns (1) to (4) show estimates for boys, while columns (5) to (8) for girls. The coefficients for the expected probabilities of finding a job with primary and secondary school as the highest level of education are reported in the top panel of Table 3.4. They document that the direction of the impact is robust to excluding controls, but the inclusion of controls improves the precision of the estimates. Results are not statistically significant in all specifications. However, they suggest a possible positive trend in expectations. For boys the coefficients turn positive with secondary education; for girls they are larger in absolute terms for primary school as highest degree achieved than for secondary school. Given the low statistical significance of the results, however we argue that job market perspective of those exposed to the programme remain substantially unchanged. Having at least one child in a treated school does not change parents' expectations about their children's future opportunities to find a job

to matter only for the likelihood of getting a secondary education level for girls. Their parents expect them to achieve an higher level of education when the assistant is female than when he is male. However, the caveat here is that among assistants only one is male. Results are not reported, but they are available upon request.

compared to having no children in a treated school.[44]

[insert TABLE 3.4 here]

The second part of the table shows the results for the expected (log) mean earnings per month. Parents in treated households expect higher future salaries for both boys and girls. Conditional on having achieved a secondary education level, being in a treated household increases the expected monthly earnings by almost 9.6% for boys and 10.5% for girls, on average.[45] This increase corresponds to almost 26 Euro with respect to an average expected earning in households not exposed to the programme of 271 Euro for boys and 255 Euro for girls. Thus, although treated parents do not expect higher employment perspectives for their children, they do expect higher salaries once they obtain a job. This suggests that they likely expect them to get better jobs. For boys this is the case also conditional on having achieved a primary education level: being in a treated household increases the expected monthly earnings by almost 11%, on average. The regression results for secondary education as the highest expected level of education are reported in the bottom part of the table. We find that parents in treated households are more likely to expect their children to finish secondary school. The impact is statistically significant only for boys. On average, parents of

[44] The coefficients of controls are not reported, but they are available upon request.

[45] The regression coefficients can be interpreted as semi-elasticities. 0.092 corresponds to $100 * (e^{0.092} - 1)$; 0.100 corresponds to $100 * (e^{0.100} - 1)$. We estimate the effects also with minimum and maximum earnings. Results are similar. They are not reported, but they are available upon request.

pupils in *Early Enrollees* schools are 12.3 percentage points more likely to expect their male children to finish secondary school.[46]

3.4.2 Remedial education programme and role model

We cannot exclude that the effect of the programme also passes through the remedial education channel. Parents expect their children to go more to school not because of the role model but because they perform better now thanks to the assistant. In our survey we run quick test scores in Mathematics and Serbian. We define *Maths score* equal to 1 when the kid is able to correctly answer both questions – "Please tell me how much is 5+4?"– and – "Please tell me how much is 23+12?"–, and 0 otherwise. We define *Serbian score* equal to 1 when the kid is able to read and write. A child is able to read when she knows to correctly read the sentence written on a card – "Could you please read me the letters, the word and the sentence on this card?"; *Able to read* takes value 0 when she does not know letters, recognises only letters or knows to read the words but cannot read the complete sentence. A child is able to write when she knows to correctly write a proposed sentence – "Please write the following sentence"; *Able to write* takes value 0 when she does not know to write at all or she writes the sentence with mistakes. Results are reported in Table 3.5.

[46] If we consider only the oldest boy and the oldest girl in the household we obtain similar results. The magnitude of the coefficients is even higher. The coefficient of treatment for boys is statistically significant at 10%.

[insert TABLE 3.5 here]

Girls in *Early Enrollees* schools perform better than girls in *Late Enrollees* in both test scores, although impacts are statistically significant only for Serbian. On average, they get 0.31 of a standard deviation more. For boys the results are less consistent. Those in *Early Enrollees* schools perform better in Serbian than pupils in *Late Enrollees*, but they do not in Mathematics. On average, they get 0.35 of a standard deviation more in Serbian.[47]

Nonetheless, if among the worst performers we find that aspirations have increased for those exposed to the programme, we have evidence that the effect of the programme does not passes through the remedial education channel only. We select only those who wrongly answer the Serbian score and the Maths score. Results are reported in Table 3.6.[48]

[insert TABLE 3.6 here]

Previous results are confirmed, although we need to be cautious in the interpretation of the results given the small sample. They suggest a positive trend in expectations for the probability of finding a job. The coefficient with secondary school becomes

[47] We also studied the impact of the programme on schooling outcomes in its first year of implementation in a companion study Battaglia and Lebedinski (2011). We find that the programme had a positive effect. There is evidence that children exposed to RTA went more to school and that, on average, marks have improved and dropouts have reduced for children in their first grade. Higher and more systematic impacts are obtained in schools with a lower number of Roma, especially if female.

[48] For robustness check, we estimate the effects also selecting those who wrongly answer the Serbian score and those who wrongly answer the Maths score separately. Results do not change.

significantly different from 0. Conditional on having achieved a certain education level, being in a treated household does increase the expected monthly earnings. This is especially the case for girls with primary school. Parents revise their expectations about the highest level of education achievable even more than in the previous specification. On average, parents of pupils in *Early Enrollees* schools are 48.6 percentage points more likely to expect their male children to finish secondary school.

Number of Roma per class

We believe that the number of Roma per class and their percentage are important to understand our results. We expect that the higher is the concentration of Roma per class, the more the parents may perceive that they are more segregated. The impact of the role model would be mitigated. In our survey pupils were randomly selected from lists of students provided by schools so we have not got precise information on the number of Roma per class. However, we also collected first-hand administrative data of each school involved in the programme (Battaglia and Lebedinski, 2011). By merging these data with data obtained through the survey, we are able to control for the number of Roma per class when investigating the impact of the programme on parents' aspirations.[49] Results are reported in Tables 3.7.[50]

[insert TABLE 3.7 here]

[49] We get a smaller sample because we collected administrative data on 4 schools out of 8 among the controls.

[50] We also use percentage of Roma per class. Results confirm those obtained here. They are not reported but they are available upon request.

The number (or percentage) of Roma kids per class matters in the creation of aspirations. The higher is their number in treated schools, the lower the expectations to find a job, get a higher salary and achieve a secondary level of education. Few results are statistically significant. Our sample reduces to few observations. However, the trends are suggestive of a negative impact of the number of Roma per class on the outcomes of interest. The inclusion of this information does not change our main results. Overall, the total effects remain positive.[51] Still, the effect of the programme is mitigated when there is a high number of Roma in the same class.

3.4.3 Heterogeneous effects

In this section we examine heterogeneous effects of the programme on Muslim versus Non-Muslim households, households in *only Roma* settlements versus households in less concentrated settlements, and younger (6 to 10 years) versus older children (11 to 15).

Muslim households versus Non-Muslim households

Our main specification is suggestive of the fact that there could be a differential effect of the programme on Muslim households. Parents from Muslim households expect higher earnings conditional on finishing secondary school and they expect their chil-

[51] The only exception is secondary school as the highest level of education expected for girls. Female are not affected by the number of Roma per class and - if they were - they would rather be positively affected by it.

dren to attain a lower level of education for both genders when compared to Non-Muslim households.[52] Moreover, descriptive statistics show that treated and control groups differ in the number of Muslim households: our control group has significantly more Muslim families. We therefore think it would be worthy to investigate whether the programme affects differently this group. We proceed therefore with the following specification (3.3) which includes the interaction of being in a Muslim household and in a treated household:

$$\begin{aligned} Y_{(i)j} &= \delta_0 + \delta_1 treatment_j + \delta_2 muslim_j + \\ &+ \delta_3 treatment_j * muslim_j + \delta_4 X'_{(i)j} + \epsilon_j \end{aligned} \tag{3.3}$$

$Y_{(i)j}$ are the usual outcomes of interest.[53] The coefficient δ_1 captures the effect of treatment on Non-Muslims. The coefficient δ_2 captures the difference between Muslims and Non-Muslims for the whole sample, and δ_3 is the differential impact of interest. Our results are reported in columns (1) and (3) of Table 3.8.

[insert TABLE 3.8 here]

Overall, estimates suggest that the programme does not impact differently Muslim and Non-Muslim in terms of job market perspectives and expected salaries. Nonetheless, Non-Muslim Roma react more in terms of expectations on educational achievement to the presence of a teaching assistant than Muslim Roma. Non-Muslims who are in *Early Enrollees* schools are on average

[52] Results are not reported but they are available upon request.
[53] $X_{(i)j}$ does not include now whether the household is Muslim.

21.4 percentage points more likely to expect their male children to finish secondary education compared to Non-Muslims in control schools. We do not know the religion of the assistants in order to investigate further.

Households in only Roma settlements versus households in less concentrated settlements

Descriptive statistics show that treated households are more in *only Roma* settlements. It is worthy to investigate whether the programme affects differently those in *only Roma* settlements and those in *mixed (Roma and Non-Roma)* settlements. We therefore proceed with specification (3.4) which includes the interaction of being in a *only Roma* settlements and being treated:[54]

$$\begin{aligned} Y_{(i)j} = \phi_0 + \phi_1 treatment_j + \phi_2 onlyroma_j + \\ + \phi_3 treatment_j * onlyroma_j + \phi_4 X'_{(i)j} + \eta_j \end{aligned} \tag{3.4}$$

Our results are reported in columns (2) and (4) of Table 3.8. Overall, estimates suggest that the programme does not impact differently households in *only Roma* settlements and households in *mixed* settlements in terms of job market perspectives and expected salaries. The coefficients are not statistically significant in all the cases. Nonetheless, parents of children living in *mixed* settlements who are in *Early Enrollees* schools are on average 13.6 percentage points more likely to expect their male children

[54] $X_{(i)j}$ does not include now whether the household is in a *only Roma* settlement.

to finish secondary education compared to parents of children living in *mixed* settlements in control schools. Our overall results on expected level of education are mainly driven by responses from parents living in *mixed* settlements.

Young versus old kids

Younger children aged 6 to 10 may respond differently to the programme from older children aged 11 to 15. There are two reasons to expect this to be the case. First, assistants were explicitly asked to work more with younger children. Second, we know that the gap in knowledge between Roma and Non-Roma children is present already when children enrol in primary school and that it increases over time. Under such circumstances, it might be easier to influence expectations of parents for younger children than for the older ones.

We estimate the following regression by gender. We have individual outcomes only for the expected education level and we estimate only this outcome. *young* is equal 1 if the kid is aged 6 to 10.

$$\begin{aligned} Y_{ij} &= \vartheta_0 + \vartheta_1 treatment_j + \vartheta_2 young_{ij} + \\ &+ \vartheta_3 treatment_j * young_{ij} + \vartheta_4 X'_{ij} + \tau_{ij} \end{aligned} \tag{3.5}$$

The results are shown in columns (5) and (6) of Table 3.8. Our coefficients are not statistically significant when we compare boys in *Early Enrollees* and *Late Enrollees* schools, although the magnitude and direction are still suggestive of the effect. We find

that there is little difference between younger and older boys. The programme affects the probability to finish secondary school similarly for both groups, although the effect is slightly higher for younger kids. We find a different effect for girls instead. Young girls in *Early Enrollees* schools are on average 27.9 percentage points more likely to be expected to finish secondary school than older female schoolmates.

3.5 Conclusion

There are reasons to believe that the lack of goals and aspirations is an important factor influencing the educational decision of Roma people. They often perceive low benefits of going to school compared to the respective costs and underinvest in education. Nonetheless, although there is evidence that the mean earnings of Roma workers are lower than those of Non-Roma workers, among Roma the difference between average earnings from one education level and another are high. The problem is that they are not aware of the situation. The provision of a role model can reduce the information gap between perceived returns to schooling and actual returns which likely fosters low aspirations for Roma people. The Roma Teaching Assistant Programme offers a perfect example. All its assistants are Roma and from the same social background of the pupils they help.

We exploit the gradual implementation of the RTA programme to identify its impact on aspirations. Our data, collected one year after the first implementation, suggest that parents of chil-

dren exposed to the programme expect higher returns to education for their kids. They are not more likely to expect them to find a job, but once employed they are expected to get higher salaries. This suggests that they might expect for them better jobs. On average, being in a treated household increases the expected monthly earnings by almost 9.6% for boys and 10.5% for girls. Moreover, on average, parents of pupils in *Early Enrollees* schools are 12.3 percentage points more likely to expect their male children to finish secondary education than parents of pupils in *Late Enrollees* schools. We cannot exclude that the effect of the programme passes also through the remedial education channel. Parents expect their children to go more to school because they perform better now thanks to the assistant. From our survey we know that those treated do better in test scores. However, if we select those performing badly, our results hold. Furthermore, an examination of heterogeneous effects suggests first that our results on highest expected level of education are driven by responses from Non-Muslim parents and parents of those living in *mixed* settlements. Second, parents revise their expectations in response to the programme mainly for younger kids (6 to 10 years). Especially younger girls are more affected by the presence of an assistant: they are on average 27.9 percentage points more likely to be expected to finish secondary school than older female schoolmates. Overall, these results suggest that the presence of a person of the same social background who showed to be successful motivates parents to believe their children can succeed. Interventions to raise perceived returns may thus be ef-

fective in increasing current investment in education. This study suggests replicable examples in contexts where minorities suffer low attainment rates and social exclusion. It shows the importance of a role model mechanism that works, especially if we consider that we are in short-time horizon. One year of a remedial education programme may not be enough to break the curse of low aspirations, but encouraging results are found in this direction. Investigating the effects of such programmes in the long-run is a central question for future research.

3.6 Bibliography

Alesina, A., R. Baqir, and W. Easterly (1999). Public Goods and Ethnic Divisions. *The Quarterly Journal of Economics 114*, 1243–1284.

Battaglia, M. and L. Lebedinski (2011). Equal Access to Education: An Evaluation of the Roma Teaching Assistant Programme in Serbia.

Bayer, P., S. L. Ross, and G. Topa (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy 116*, 1150–1196.

Beaman, L., E. Duflo, R. Pande, and P. Topolova (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. Sciencexpress.

Blume, L. and S. Durlauf (2006). *Identifying Social Interactions: A Review.* Michael Oakes and Jay Kaufman (eds.), Methods in Social Epidomology, Wiley.

Boeri, T., M. D. Philippis, E. Patacchini, and M. Pellizzari (2011). Moving to Segregation: Evidence from 8 Italian citites. Working Papers 390, IGIER Bocconi University.

Card, D. and J. Rothstein (2007). Racial Segregation and the Black-White Test Score Gap. *Journal of Public Economics 91* (11–12), 2158–2184.

Clark, K. and S. Drinkwater (2002). Enclaves, Neighbourhood Effects and Employment Outcomes: Ethnic Minorities in England and Wales. *Journal of Population Economics 7*(5), 603–628.

Cutler, D. M. and E. L. Glaeser (1997). Are Ghettos Good or Bad? *The Quarterly Journal of Economics, 112*, 827–872.

Dee, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. *The Review of Economics and Statistics 86*(1), 195–210.

Delavande, A., X. Ginè, and D. McKenzie (2009). Measuring Subjective Expectations in Developing Countries.

Durlauf, S. (2004). Chapter 50 Neighborhood Effects. In J. V. Henderson and J.-F. Thisse (Eds.), *Cities and Geography*, Volume 4 of *Handbook of Regional and Urban Economics*, pp. 2173 – 2242. Elsevier.

Edin, P.-A., P. Frederiksson, and O. Åslund (2003). Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment. *The Quarterly Journal of Economics 118*(1), 329–357.

Genicot, G. and D. Ray (2010). Aspirations and Inequality.

Glewwe, P. M. K. (2006). *Handbook of Economics of Education, Chapter 16.* North Holland.

Graham, P. A. (1987). Black Teachers: A Drastically Scarce Resource. *Phi Delta Kappan 68*(8), 598–605.

Jensen, R. (2010). The (perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics 125 (2)*, 515–548.

Kaufmann, K. M. and O. Attanasio (2009, July). Educational Choices, Subjective Expectations, and Credit Constraints. Technical report, NBER Working Paper No. w15087.

Krishnan, P. and S. Krutikova (2010, February). Skill Formation in Bombay's Slums. Technical report, NBER Working Paper No. w15087.

La Ferrara, E. and A. Mele (2006, July). Racial Segregation and Public School Expenditure. Technical report, CEPR Discussion Paper No. 5750.

Ladson-Billings, G. (1994). *The Dreamkeepers: Successful Teachers of African American Children.* Jossy Bass Publishers.

Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies 60 (3)*, 531–542.

Nguyen, T. (2008). *Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar.* Ph. D. thesis, Massachusetts Institute of Technology.

Open Society Institute (2007).

Ray, D. (2004). Aspirations, Poverty and Economic Change. Technical report, Bureau for Research in Economic Analysis of Development (BREAD) Policy Papers No.442.

Wilson, W. J. (1987). *The Truly Disadvantaged*. University of Chicago Press.

3.7 Tables

Table 3.1: Characteristics of the schools and assistants (Belgrade)

	All	Early Enrollees	Late Enrollees	Difference (1-2)	P-value
	(1)	(2)	(3)	(4)	(5)
Characteristics of the schools					
Class size	24	22.24	25.63	-3.39	[0.081]
No. Roma per class	4.89	4.43	5.23	-0.79	[0.759]
No. Roma per class (if at least a Roma)	5.1	4.84	5.32	-0.48	[0.854]
Roma per school (%)	21	20	22	-2	[0.839]
Female					
Roma	0.49	0.52	0.46	0.06	[0.120]
Non-Roma	0.47	0.48	0.47	0.01	[0.551]
Born in the same town					
Roma	0.35	0.40	0.32	0.08	[0.654]
Non-Roma	0.76	0.81	0.73	0.08	[0.209]
Number of schools	9	5[a]	4		
Number of Roma pupils	581	231	350		
Number of Non-Roma pupils	2133	927	1206		
Characteristics of the assistants					
Female	0.875	0.8	1	-0.2	[0.374]
Maximum level of education					
Secondary school	0.5	0.6	0.33	0.27	[0.543]
University	0.375	0.4	0.33	-0.07	[0.877]
Experience with Roma	0.75	1	0.33	0.67	[0.183]
Experience in NGO	1	1	1	0	[.]
Number of assistants	8	5	3[b]		

[a] *Early Enrollees* schools are 6. One school did not provide us the list of students so it is excluded from our sample.

[b] We could not get information about one assistant in the *Late Enrollees* schools.

Table 3.2: Means of control variables in treated and control households

Variables at the household level	All	Treatment	Control	Difference
Wealth[a]	0.08	-0.14	0.22	-0.36
				(0.27)
Monthly Total income (in dinars)[b]	28949.47	28224.39	29453.33	-1228.94
				(2144.78)
Informal (=1)[c]	0.32	0.31	0.32	-0.01
				(0.05)
Urban (=1)	0.51	0.47	0.53	-0.06
				(0.06)
Only Roma in settlement (=1)[d]	0.21	0.28	0.16	0.12**
				(0.05)
No schooling/Unfinished primary school (=1)[e]	0.07	0.07	0.07	0.00
				(0.03)
Finished primary school (=1)[e]	0.74	0.69	0.76	-0.07
				(0.05)
Finished secondary school (=1)[e]	0.19	0.23	0.16	0.07
				(0.05)
Muslim (=1)	0.71	0.57	0.80	-0.23***
				(0.05)
Number of children under 5	0.72	0.75	0.70	0.05
				(0.10)
Number of female children between 6 and 15	1.65	1.73	1.59	0.14
				(0.14)
Number of male children between 6 and 15	1.75	1.80	1.80	0.10
				(0.13)
Number of adults	2.44	2.46	2.44	0.02
				(0.12)
max no. observations	300	122	178	
Variables at the individual level				
Children characteristics				
Male (=1)	0.52	0.50	0.54	-0.04
				(0.04)
Age of child	9.89	10.11	9.74	0.37
				(0.20)
Rank among siblings	2.20	2.33	2.11	0.22**
				(0.09)
Mark in Mathematics[f]	2.77	2.86	2.70	0.16*
				(0.09)
Mark in Serbian[f]	2.85	2.94	2.79	0.15
				(0.09)
max no. observations	673	280	393	

Standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] The wealth index was calculated with principal component analysis. The index ranges between -5.55 and 3.69.

[b] 28950 dinars corresponds to 279 Euro (1 RSD = 0.009626 Euro, November 2011).

[c] =1 if at least one household member works in the informal sector.

[d] =1 if the respondent declared that the household lives in an exclusively Roma neighbourhood.

[e] It refers to the highest level of education obtained by parents.

[f] We use demeaned mark in Mathematics and Serbian. The mark is demeaned from the average school mark.

Table 3.3: Means of outcome variables in treated and control households

Variables at the household level	All	Treatment	Control	Difference
Probability to find a job: Boys				
With primary school (=1)[a]	0.42	0.35	0.48	-0.13**
				(0.06)
With secondary school (=1)[a]	0.82	0.82	0.82	0.00
				(0.05)
Probability to find a job: Girls				
With primary school (=1)[a]	0.35	0.31	0.39	-0.08
				(0.06)
With secondary school (=1)[a]	0.79	0.74	0.82	-0.07
				(0.05)
Expected mean log earning: Boys				
With primary school	9.91[b]	9.97	9.87	0.10
				(0.06)
With secondary school	10.21[c]	10.24	10.18	0.06*
				(0.19)
Expected mean log earning: Girls				
With primary school	9.82[d]	9.90	9.78	0.12*
				(0.07)
With secondary school	10.14[e]	10.18	10.11	0.07*
				(0.04)
Variables at the individual level				
Expected to finish : Boys				
Secondary school (=1)	0.61	0.67	0.57	0.10*
				(0.06)
Expected to finish : Girls				
Secondary school (=1)	0.63	0.67	0.60	0.07
				(0.06)

Standard errors in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Respondent expects the child to find a job given a certain level of education achieved.

[b] The corresponding average earning is 21709 dinars (208 Euro, Nov 2011). For treated households is 22985 dinars (221 Euro, Nov 2011); for control households is 21075 dinars (202 Euro, Nov 2011).

[c] The corresponding average earning is 28654 dinars (276 Euro, Nov 2011). For treated households is 29398 dinars (283 Euro, Nov 2011); for control households is 28141 dinars (271 Euro, Nov 2011).

[d] The corresponding average earning is 19432 dinars (187 Euro, Nov 2011). For treated households is 20915 dinars (201 Euro, Nov 2011); for control households is 18682 dinars (180 Euro, Nov 2011).

[e] The corresponding average earning is 26923 dinars (259 Euro, Nov 2011). For treated households is 27529 dinars (265 Euro, Nov 2011); for control households is 26527 dinars (255 Euro, Nov 2011).

Table 3.4: All outcomes with primary and secondary school by gender

Gender	Boys				Girls			
Max. level of education	Primary school		Secondary School		Primary school		Secondary school	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability to find a job with primary/secondary school								
treatment	-0.109	-0.068	0.004	0.012	-0.067	-0.114	-0.067	-0.037
	(0.087)	(0.078)	(0.057)	(0.055)	(0.083)	(0.078)	(0.055)	(0.054)
controls[a]	no	yes	no	yes	no	yes	no	yes
No. observations	300	276	300	276	294	268	292	267
R-squared	0.012	0.141	0.000	0.067	0.005	0.121	0.007	0.123
Expected log earnings with primary/secondary school								
treatment	0.128	0.107*	0.079	0.092*	0.149*	0.123	0.079	0.100*
	(0.078)	(0.059)	(0.060)	(0.050)	(0.083)	(0.085)	(0.061)	(0.056)
controls[a]	no	yes	no	yes	no	yes	no	yes
No. observations	129	119	246	224	105	98	232	216
R-squared	0.031	0.199	0.017	0.123	0.050	0.241	0.017	0.147
Secondary school as the highest expected level of education								
treatment			0.097	0.123*			0.067	0.003
			(0.073)	(0.066)			(0.080)	(0.086)
controls[b]			no	yes			no	yes
No. observations			299	232			275	221
R-squared			0.009	0.346			0.005	0.230

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults.

[b] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults, age of child, age of child squared, rank among siblings, demeaned mark in Mathematics and demeaned mark in Serbian.

Table 3.5: Test scores by gender

	Maths score		**Serbian score**		**Able to read**		**Able to write**	
Gender	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
treatment	-0.008	0.024	0.176**	0.150*	0.158**	0.036	0.091	0.057
	(0.082)	(0.098)	(0.081)	(0.078)	(0.073)	(0.073)	(0.077)	(0.077)
controls[a]	yes	yes	yes	yes	yes	yes	yes	yes
No. observations	160	141	156	141	160	142	155	142
R-squared	0.237	0.225	0.240	0.188	0.242	0.179	0.254	0.184

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults, age of child, age of child squared, rank among siblings, demeaned mark in Mathematics and demeaned mark in Serbian.

Table 3.6: Worst performers: all outcomes with primary and secondary school by gender

Gender	Boys		Girls	
Max. level of education	Primary school	Secondary School	Primary school	Secondary school
	(1)	(2)	(3)	(4)
Probability to find a job with primary/secondary school				
treatment	-0.166	0.174*	-0.019	0.051
	(0.168)	(0.100)	(0.182)	(0.104)
controls[a]	yes	yes	yes	yes
No. observations	69	69	67	67
R-squared	0.231	0.128	0.205	0.293
Expected log earnings with primary/secondary school				
treatment	0.047	0.160	0.141*	0.099
	(0.126)	(0.118)	(0.072)	(0.100)
controls[a]	yes	yes	yes	yes
No. observations	42	59	33	57
R-squared	0.496	0.341	0.594	0.221
Secondary school as the highest expected level of education				
treatment		0.486***		0.151
		(0.150)		(0.301)
controls[b]		yes		yes
No. observations		54		27
R-squared		0.482		0.605

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults.

[b] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults, age of child, age of child squared, rank among siblings, demeaned mark in Mathematics and demeaned mark in Serbian.

Table 3.7: Number of Roma per class

Gender	Boys		Girls	
Max. level of education	Primary school	Secondary School	Primary school	Secondary school
	(1)	(2)	(3)	(4)
Probability to find a job with primary/secondary school				
treatment	0.213	0.157	0.351	0.094
	(0.268)	(0.174)	(0.245)	(0.177)
number of Roma per class	0.001	-0.010	0.009	-0.017**
	(0.013)	(0.013)	(0.015)	(0.008)
treatment*	-0.013	-0.007	-0.021	-0.006
number of Roma per class	(0.029)	(0.019)	(0.026)	(0.018)
controls[a]	yes	yes	yes	yes
Total effect	0.200	0.150	0.330	0.088
	(0.242)	(0.16)	(0.221)	(0.16)
No. observations	121	121	120	120
R-squared	0.199	0.189	0.119	0.251
Expected log earnings with primary/secondary school				
treatment	0.262	0.216	0.164	0.229
	(0.269)	(0.148)	(0.335)	(0.159)
number of Roma per class	0.013	0.017**	0.012	0.014
	(0.018)	(0.007)	(0.024)	(0.008)
treatment*	-0.004	-0.027*	0.030	-0.020
number of Roma per class	(0.033)	(0.016)	(0.053)	(0.019)
controls[a]	yes	yes	yes	yes
Total effect	0.258	0.189	0.193	0.209
	(0.242)	(0.130)	(0.288)	(0.142)
No. observations	43	99	34	94
R-squared	0.338	0.256	0.404	0.254
Secondary school as the highest expected level of education				
treatment		0.207		-0.330
		(0.204)		(0.241)
number of Roma per class		0.003		-0.012
		(0.019)		(0.015)
treatment*		-0.028		0.034
number of Roma per class		(0.029)		(0.023)
controls[b]		yes		yes
Total effect		0.179		-0.296
		(0.182)		(0.221)
No. observations		91		100
R-squared		0.441		0.289

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults.

[b] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults, age of child, age of child squared, rank among siblings, demeaned mark in Mathematics and demeaned mark in Serbian.

Table 3.8: Heterogeneous effects: all outcomes for secondary school by gender

Gender	Boys		Girls		Boys	Girls
Max. level of education			Secondary School			
	(1)	(2)	(3)	(4)	(5)	(6)
Probability to find a job with primary/secondary school						
treatment	-0.031	0.057	-0.045	-0.006		
	(0.068)	(0.058)	(0.076)	(0.058)		
muslim	-0.008		0.022			
	(0.067)		(0.078)			
treatment*muslim	0.066		0.012			
	(0.098)		(0.107)			
only Roma in settlement		0.079		-0.068		
		(0.090)		(0.093)		
treatment*only Roma in settlement		-0.200		-0.132		
		(0.135)		(0.155)		
controls[a]	yes	yes	yes	yes		
No. observations	276	276	268	267		
R-squared	0.144	0.076	0.135	0.127		
Expected log earnings with primary/secondary school						
treatment	0.013	0.070	0.001	0.083		
	(0.085)		(0.122)			
muslim	0.077		0.065			
	(0.082)		(0.116)			
treatment*muslim	0.121		0.150			
	(0.109)		(0.137)			
only Roma in settlement		-0.018		0.054		
		(0.087)		(0.106)		
treatment*only Roma in settlement		0.102		0.085		
		(0.112)		(0.123)		
controls[a]	yes	yes	yes	yes		
No. observations	224	224	216	216		
R-squared	0.131	0.127	0.158	0.150		
Secondary school as the highest expected level of education						
treatment	0.214**	0.136*	0.163	0.067	0.093	-0.159
	(0.094)	(0.078)	(0.143)	(0.094)	(0.093)	(0.116)
muslim	-0.134		-0.093			
	(0.093)		(0.131)			
treatment*muslim	-0.146		-0.223			
	(0.138)		(0.158)			
only Roma in settlement		-0.002		0.063		
		(0.103)		(0.133)		
treatment*only Roma in settlement		-0.065		-0.274		
		(0.174)		(0.219)		
young					0.055	-0.378**
					(0.116)	(0.144)
treatment*young					0.056	0.279**
					(0.108)	(0.133)
controls[b]	yes	yes	yes	yes	yes	yes
No. observations	232	232	221	221	232	221
R-squared	0.350	0.347	0.236	0.238	0.348	0.264

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults.

[b] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults, age of child, age of child squared, rank among siblings, demeaned mark in Mathematics and demeaned mark in Serbian.

3.8 Figures

Figure 3.1: Comparison of real returns to education for Roma and Non-Roma (Belgrade)

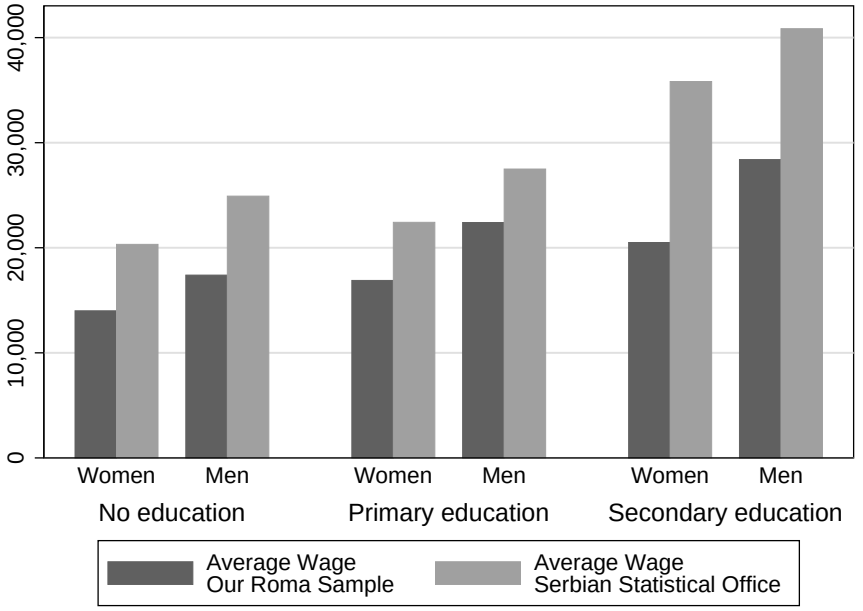


Figure 3.2: Job characteristics by education levels - Roma people

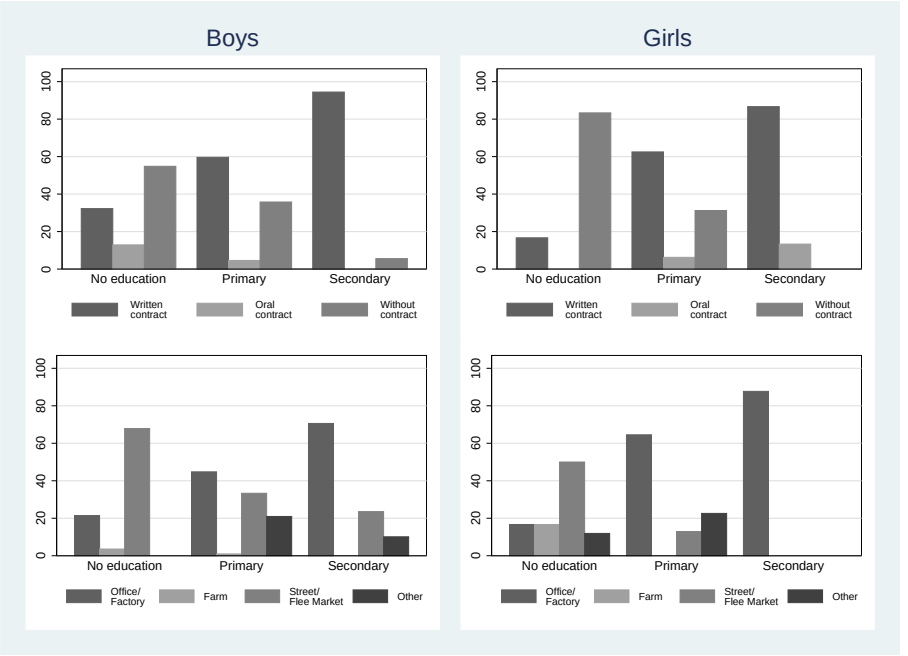


Figure 3.3: Comparison of real and expected returns to education

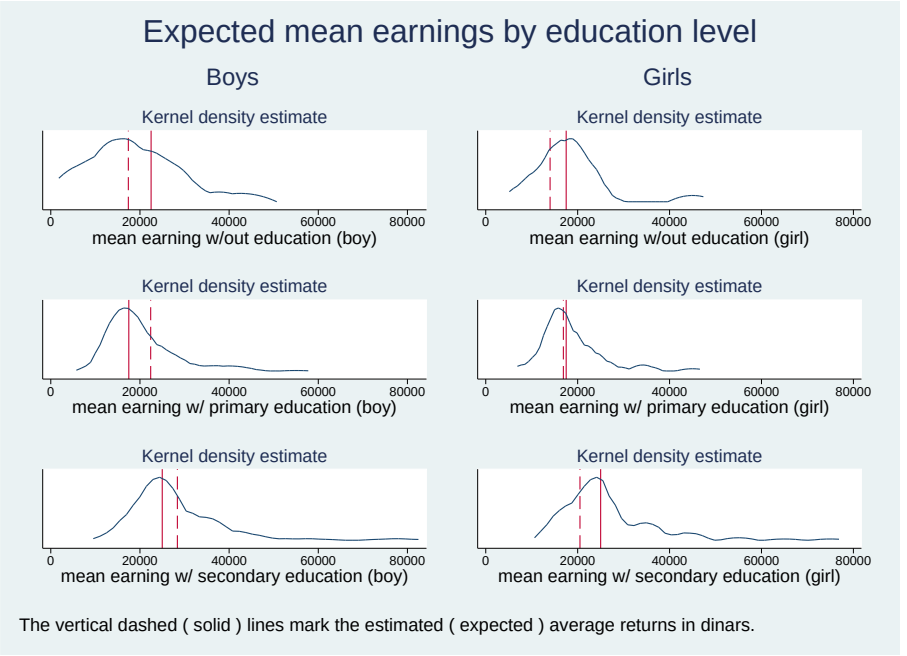
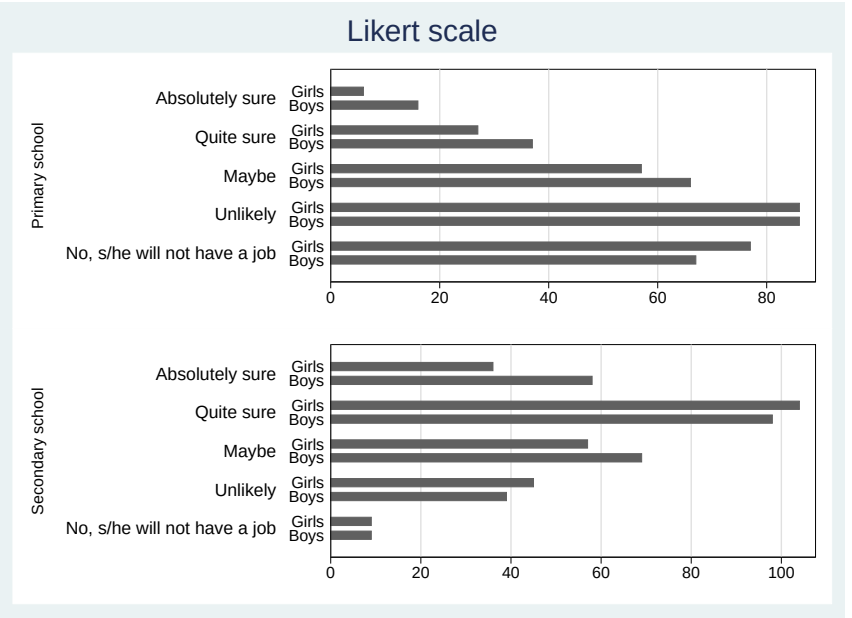


Figure 3.4: Settlements of the survey

Figure 3.5: Likert scale for the probability of finding a job with primary and secondary school by gender



3.9 Appendix

Table 3.9: Ordered Logit - thresholds among categories

	Boys		Girls	
	Primary	Secondary	Primary	Secondary
	(1)	(2)	(3)	(4)
treatment	0.089	-0.422	0.166	0.083
	(0.309)	(0.372)	(0.390)	(0.349)
controls	yes	yes	yes	yes
cut1				
constant	-2.974***	-0.914	-3.457***	-0.338
	(0.661)	(0.704)	(0.720)	(0.864)
cut2				
constant	-1.535***	0.809	-1.573**	1.876**
	(0.611)	(0.691)	(0.684)	(0.871)
cut3				
constant	-0.131	2.131***	0.059	3.234***
	(0.580)	(0.721)	(0.659)	(0.874)
cut4				
constant	1.563**	4.195***	1.647**	5.669***
	(0.616)	(0.888)	(0.674)	(1.032)
No. observations	276	276	268	267

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: Pooled sample: all outcomes with primary and secondary school

Max. level of education	Primary school (1)	Secondary school (2)
Probability to find a job with primary/secondary school		
treatment	-0.109	-0.041
	(0.079)	(0.056)
treatment*male	0.037	0.051
	(0.076)	(0.036)
controls[a]	yes	yes
Total Effect	-0.072	0.010
	(0.076)	(0.052)
No. observations	532	531
R-squared	0.134	0.090
Expected log earnings with primary/secondary school		
treatment	0.109	0.096*
	(0.084)	(0.056)
treatment*male	-0.029	-0.011
	(0.081)	(0.039)
controls[a]	yes	yes
Total Effect	0.079	0.085*
	(0.059)	(0.050)
No. observations	209	431
R-squared	0.223	0.139
Secondary school as the highest expected level of education		
treatment		-0.001
		(0.084)
treatment*male		0.123
		0.081
controls[b]		yes
Total Effect		0.122*
		(0.066)
No. observations		454
R-squared		0.286

Robust standard errors corrected for clustering at the school-cohort level in parentheses: * significant at 10%, ** significant at 5%, *** significant at 1%.

[a] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults.

[b] Control variables included are wealth, informal (=1), urban (=1), only Roma in settlement, finished primary school (=1), finished secondary school (=1), muslim (=1), number of children under 5, number of female children between 6 and 15, number of male children between 6 and 15 and number of adults, age of child, age of child squared, rank among siblings, demeaned mark in Mathematics and demeaned mark in Serbian.