DECLARATION FOR THE PhD THESIS

To be inserted as the first page of the thesis

The undersigned

SURNAME	Lebedinski
FIRST NAME	Lara
PhD Registration Number	1287564

Thesis title:

Essays in Labour Economics

PhD in	Economics
Cycle	23
Candidate's tutor	Tito Boeri
Year of discussion	2012

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 (*thesis* title) Essays in Labour Economics;
- by (*candidate's surname and first name)* Lebedinski Lara;

- discussed at Università Commerciale Luigi Bocconi – Milano in (year of discussion) 2012 ;
- 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;
- **only in cases where another declaration has been undersigned requesting a temporary embargo**: the thesis is subject to a temporary embargo for (indicate duration of the embargo) ..... months;

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) **choose hypothesis 7a or 7b indicated below**:

X 7a) 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;

Or

7b) that the thesis meets one of the temporary embargo hypotheses included in the declaration "**TEMPORARY EMBARGO REQUEST OF THE PhD THESIS**" undersigned elsewhere.

Date _31/01/2011_

Signed (write first name and surname) Lara Lebedinski

Essays in Labour Economics

Lara Lebedinski (1287564)
Bocconi University

Thesis Committee:

Prof. Tito M. Boeri, Department of Economics, Bocconi University (Chair)

Dr. Simone Ghislandi, Department of Policy Analysis and Public Management, Bocconi University

Dr. Andreas Madestam, Department of Economics, Bocconi University

Dissertation in partial fulfillment of the requirements for the academic degree of
Doctor of Philosophy in Economics (XXIII cycle).

Contents

Acknowledgments

It is a pleasure to thank those who made this thesis possible. I am particularly grateful to my advisors Prof. Tito Boeri and Dr. Andreas Madestam for their continous encouragement and support. I am also thankful to Prof. Eliana La Ferrara who was my advisor in the first stage of my dissertation.
I am indebted to my co-author Marianna Battaglia, Dr. Michele Pellizzari, seminar participants at Bocconi development and labour reading group, AMID LSE summer school and Bocconi Edge 2011 Jamboree.
Lastly, I would to thank the Serbian Ministry of Education, in particular Božidar Jovanović and Tünde Kovacs-Cerović, for their cooperation. Ipssos Strategic Marketing Agency, especially Dragiša Bjeloglav and Tatjana Jovanov, provided invaluable help in the data collection process.

Introduction

This dissertation has been written as part of my effort to complete the PhD programme in Economics at Bocconi University in Milan, Italy. The dissertation is in the field of labour economics with a special focus on the Roma minority group. Roma are the largest ethnic minority in Europe. They demonstrate high poverty levels, low educational attainment 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 thesis. The dissertation consists of three chapters. The first two chapters examine the effect of a remedial education programme, called Roma Teaching Assistant Programme, in the Serbian context on schooling outcomes and educational aspirations. The third chapter studies the impact of residential segregation on labour market outcomes.

The Roma Teaching Assistant programme, a remedial education programme, is the main programme aiming at improving schooling outcomes of the Roma ethnic minority in Europe. In the first chapter, together with my co-author Marianna Battaglia, I study the causal effect of this programme on schooling outcomes using different difference-in-difference approaches. For the purpose of this study, we have constructed a unique data set comprising grades for Roma and Non Roma pupils for 38 primary schools for four scholastic years. We find that the programme is effective in reducing absences for children in the lower four grades of primary school. We find also an impact on grades of first graders because they are the ones with whom the assistants mostly work.

The second chapter, also co-authored with Marianna Battaglia, probes whether the Roma Teaching Assistant programme is effective in raising aspirations of parents about their children's education. We argue that Roma Teaching Assistants are perceived by parents as role models and that their presence leads parents to believe in greater social mobility. We use three different measures of aspiration: expectation of the likelihood of the child to find a job once adult, expected salary and highest expected educational level the child will achieve. To study the effect on aspirations, we conducted a survey

in Belgrade with 300 households. We use a simple-difference approach and control for a large set of individual controls. Our results suggest that parents of pupils in treated schools expect them to achieve higher education level, while job market perspectives remain unchanged.

The third chapter of the thesis looks at the effect of segregation on labour market performance of Roma. My measure of segregation is the percentage of Roma in a census tract. I combine World Bank's Living Standard Measurement Survey together with the census data on segregation. I find suggestive evidence that women benefit from living in more segregated areas. A 10 percentage points increase in segregation improves the probability of being employed for women by 6 to 7 percentage points. One reason for this result could be that women are more enganged in jobs where personal contacts and hence the size of the co-ethnics network matter. My second finding is that less employable Roma sort into census tracts characterised by a relatively high percentage of Roma.

CHAPTER 1: Equal Access to Education: An Evaluation of the Roma Teaching Assistant Programme in Serbia*

Marianna Battaglia† Lara Lebedinski‡

Abstract

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 dropout. 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. We

*We would like to thank Tito Boeri, Eliana La Ferrara, Andreas Madestam, Michele Pellizzari and participants at AMID summer school 2010, Bocconi University 2011 Development Reading Group and Bocconi University PhD seminar. We are grateful to the personnel of the Serbian Ministry of Education, especially Božidar Jovanović, Tünde Kovacs-Cerović, for giving us the authorisation for this project and for answering all our questions. We are indebted to Nikola Duvnjak from OSCE for answering all our questions about the pilot phase of the programme. All opinions expressed are of the authors, all errors are our own.

†Department of Economics, Bocconi University. Email:marianna.battaglia@phd.unibocconi.it

‡Department of Economics, Bocconi University. Email:lara.lebedinski@phd.unibocconi.it

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 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.

Keywords: programme evaluation, primary education, Roma

JEL Classification: I21, I25, J13, D04

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 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]; significantly lower family permanent incomes, also due to greater household size and lower incidence of home ownership, and lower wages, given the overrepresentation among low skilled jobs. They often lack access to credit and property ownership and are overdependent on social benefits.

[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. Nowdays, however, they are mainly attended by pupils who are late at enroling and by pupils who decided to return to school after dropping out.

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 grades 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. A high percentage of Roma (40%-50%) is indeed younger than 18 years old and focusing on children and young people is broadly recognized a crucial step towards Roma inclusion: higher enrolment rates and better achievement at school are expected to lead to persistent effects in the labor 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 mainly in the range of 40% to 60%, still really low. Moreover, students may enrol at the beginning of the year, but may not actually attend school: the percentage of completion rates of primary school are 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 try to 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.[3] 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. Recently, especially in the European Union, they have started to consistently attract media

[3]All data are primary data, collected by the authors in the Summer-Autumn 2010.

attention. The visa liberalisation and the adhesion to the Union of countries like Romania and Bulgaria, in which the percentage of Roma population is high, have indirectly led to significant migration flows towards Western countries. 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 and increased the interest of the European Union on 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 this programme is useful not only for these countries but also for the receiving ones, where a minority group becomes also a migrant group.

The Roma Teaching Assistant Programme[4] is the main programme targeting Roma inclusion in education in South Eastern Europe.[5] It began in Serbia in 2002 as a pilot programme carried out by different NGOs and from 2007 to 2009 was led by the Organisation for Security and Cooperation in Europe (OSCE). Since 2009 the Serbian 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 chil-

[4]The Serbian name of the programme is *Romski Asistenti - Pomoć u Nastavi*.

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

dren from marginalised groups.[6] Nonetheless, the Ministry of Education expects that mainly Roma children will benefit from this programme.

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. Our results suggest that, on average, marks have improved by almost 0.36 standard deviations and dropouts have reduced by 6.6 percentage points for those children exposed to the programme in their first grade. The programme was successful in reducing absences by 0.14 standard deviations.[7] 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 seems to be especially the case for female and migrants, 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 under-

[6]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.*

[7]This translates to 3 to 4 days more spent at school.

performing 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 enroled 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 conducted 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.[8] Test scores 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 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 test score 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

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

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.[9] 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 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 busing (Angrist and Lang, 2004) or alter the neighborhoods 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. 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 test score gap disappears in reading and decreases by 85% in mathematics.

The third strand of the literature is the literature on programmes aiming at achieving better schooling outcomes of the poor. This literature suggests that conditional cash transfers, modeled after the Mexican programme PROGRESA,

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

are successful in improving enrolment and attendance in many developing countries. However, policies that promote school enrolment may not promote learning: early contributions indicate that programmes which are effective at reducing absence from school often do not have an impact on test scores 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 test scores, 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 test scores 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).

Within this framework and despite the difficulties of dealing with a not randomly assigned programme, we are able to identify the effects of a remedial education programme targeting a minority group suffering high poverty levels.

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 The Education System and Roma in Serbia

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.[10]

Primary school consists of 8 years. In the first four grades pupils have one teacher who teaches all compulsory subjects except English, while in the upper four years of primary school pupils have one teacher per subject. In the first grade children get 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 has at least one insufficient in the lower four grades 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 schools to reduce repetition rates, especially in the lower four grades.

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.[11] 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 Context

Data on Roma in Serbia are inaccurate and scarce. Official census data from 2002 suggest that in Serbia there are 108,000 Roma, although estimates put forward a number of somewhere between 350,000 to 500,000 or approximately 6% of the

[10]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 enroled 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.

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

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, 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 (70%). Nonetheless, the participation of females is only around 30% and therefore considerably lower than the national average (50%). 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 children do not attend school whereas this is the case for only 1% of the overall population of primary school age. 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 schools started enroling children with incomplete documents, but there still is 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 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 have adequate clothing for winter months and live in overcrowded homes where they do not have adequate conditions to pursue their studies. Plus, 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 face often discrimination from teachers and other pupils: in class 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. 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.

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 now 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 79 schools starting from November 2010.

Based on when the programme started in a school, we divide the schools in two groups: schools which have started with the programme in September 2009 (*Early Enrollees*) and schools which were assigned a Roma assistant starting with 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 79 *Late Enrollees*.

Both schools and potential Roma assistants had to apply in order to participate to the programme. Among 78 schools which applied in 2009 OSCE chose 26 *Early Enrollees* schools based on the following two criteria: first, the percentage of Roma students between 5% and 40% and then, preferably, the availability of preschool programme in the school.[12] The requirements for Roma assistants were knowledge of Romani, secondary school diploma and experience in working with children.

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 79 schools out of 254 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 reason 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

[12] 64 out of 78 schools which applied had a percentage of Roma beetwen 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.

an assistant from her entry in the school anyhow.[13] Selection criteria for now pedagogical assistants remained unchanged. Parents may have likely not been aware of the existence of the programme before enroling 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.[14] Therefore, we are confident that our analysis is not affected by possible selection of children into schools.

Every school receives only one assistant. Schools which entered the RTA programme received a description of the assistant's duties, but they are free to decide how to allocate the time of the assistant depending on the need of the school. In general, Roma assistants participate in regular lessons where they provide additional help to Roma pupils who have difficulties in following classes. Moreover, they organise additional lessons and help them with their homework and assignments. They mainly work with lower grades, especially the first. One day per week assistants dedicate to visiting pupils' parents. In most cases Roma live in segregated settlements so that assistants can go to the settlement and visit several families. Usually they visit parents of children who have been absent from school and inform other parents on how children are doing at school. Their objectives are indeed: making sure that children go to school; preventing them from dropping out and helping them to succeed at school.

[13]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.

[14]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 enroling at school for the first time reduced between the two years and it did it proportionally in both types of schools.

4 Preliminary Analysis

4.1 Data and Trends of the Variables

We use first hand collected data. They come from administrative records of 23 schools among *Early Enrollees* and 15 schools among *Late Enrollees*.[15] We select the 15 *Late Enrollees* schools out of 79 which got an assistant in 2010 according to the following criteria: firstly, they have to be in the same district of a *Early Enrollees* school[16]; 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.[17]

Table 1: Programme timeline

	2009 Early Enrollees	2010 Late Enrollees
Number of schools applying to the programme	78	254
Number of schools joining the programme	26	79
Number of schools in our sample	23	15

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.[18] Figure 1 reports the distribution of schools from which we have collected data. In

[15] 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 the central area of the country and one in the South.

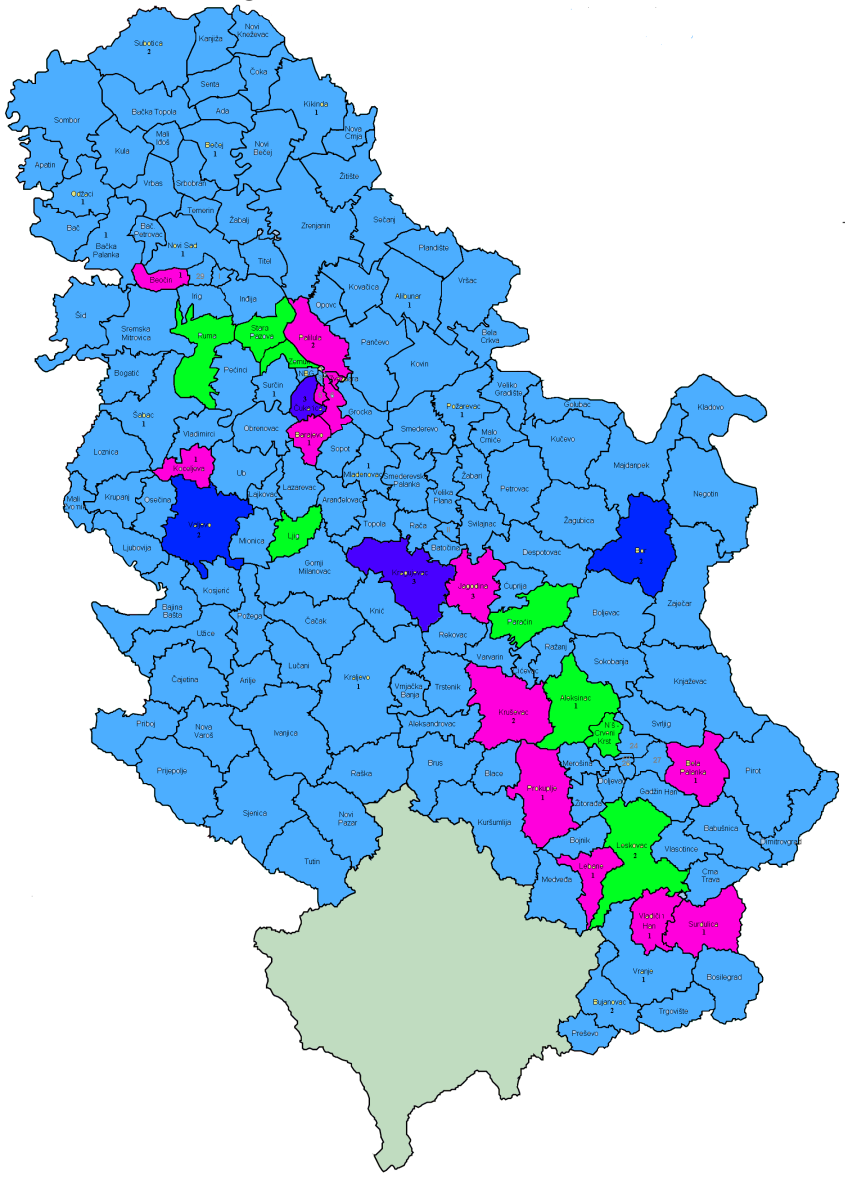
[16] A district is made by more municipalities. In Serbia there are 24 districts and 160 municipalities.

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

[18] 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 municipality 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.

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*.

Figure 1: Location of the schools with assistants



The data set contains information on 4 scholastic years, that is from 2006/2007 until 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.[19] 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 and 3 summarise respectively the averages of the control variables and main outcomes of interest for Roma and Non Roma children in the pre- and treatment year, 2008/2009 and 2009/2010 respectively.[20]

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 (columns 1). The 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.[21]

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

[19]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-Yugoslavia. Many Roma refugees in Serbia, for instance, come from Kosovo.

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

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

Table 2: Averages of control variables in pre- and treatment year

	Pretreatment year			Treatment year		
	Early Enrollees (1)	Late Enrollees (2)	Diff. (1-2) (3)	Early Enrollees (1)	Late Enrollees (2)	Diff. (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 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%.

4.33 for Non Roma. This gap is even more explicit when we look at getting an insufficient average (at least one insufficient grade): on average among Roma almost 20% of pupils get an insufficient grade; the corresponding percentage for Non Roma is not even 1%. Moreover, dropouts seem to be almost exclusively of Roma children: in 2008/2009 in *Late Enrollees* schools 1.9% among Roma

Table 3: Averages of outcomes in pre- and treatment year

	Pretreatment year			Treatment year		
	Early Enrollees (1)	Late Enrollees (2)	Diff. (1-2) (3)	Early Enrollees (1)	Late Enrollees (2)	Diff. (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)
Insufficient average:						
Roma	0.24	0.19	0.05	0.168	0.164	0.04
			(0.02)			(0.016)
Non Roma	0.008	0.01	-0.002	0.007	0.0068	0.0002
			(0.002)			(0.002)
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%.

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 schooling days, a Non Roma child is absent from school approximately 7 to 8 days in a year, while a Roma child

misses school on average somewhere between 23 and 25 days in a year.

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 both a slight improvement in all marks and a decrease in the percentage of students with an insufficient average. 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 be 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.

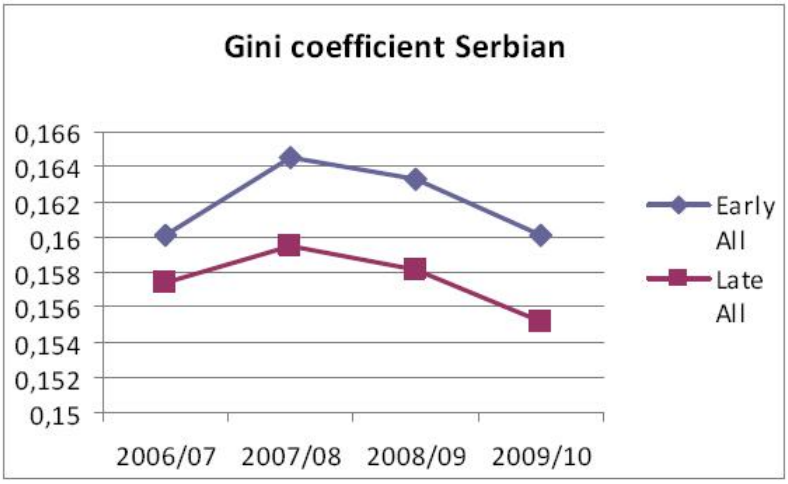
Figure 2 shows the trends of the Gini coefficient over the four year period − from 2006 to 2010 − in Serbian and Mathematics.

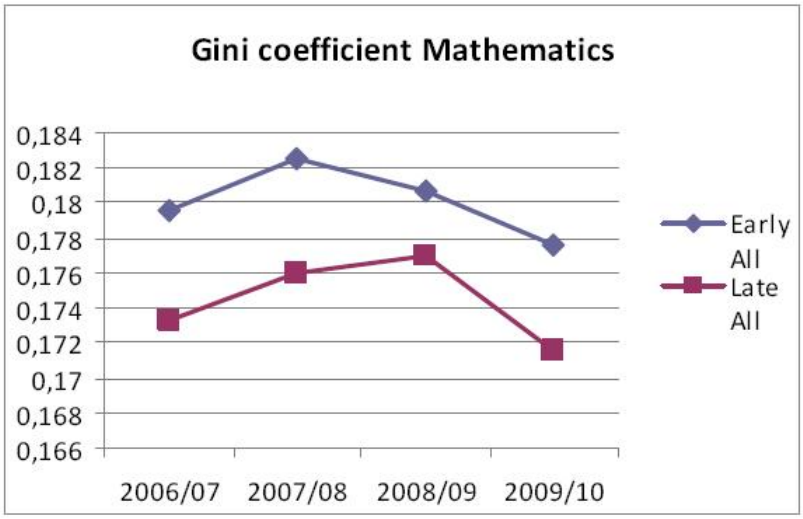
The graph suggests that inequality is lower in *Late Enrollee* schools for both subjects. An encouraging fact emerging from the graph is that inequality is

[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.

Figure 2: Gini coefficient of inequality in Serbian and Mathematics





decreasing in both *Late* and *Early Enrollee* schools over the period. Nonetheless, we do not see a sharp reduction in the inequality index among *Early Enrollee* schools compared to *Late Enrollee* 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.

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.[24] 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 selection of the schools assigned 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 are 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 starting the programme in the two different years may differ because schools had to apply in order to get selected for the programme. 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.[25] We do know, though, that observable characteristics do not differ between those schools which applied in the first year and those in the second

[24]The dataset reduces here to only Roma children. Non Roma children are not included in the following analysis.

[25]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.

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] Moreover, we 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 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 interested and motivation (application) in the same way, in both years.

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 from *Early Enrollees*-treated schools.

The advantages and disadvantages of both control groups need to be mentioned. The first control group consists of schools which enroled 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 which 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, as mentioned before, is that we are not able to control for unobservable differences which have led some schools to enter the programme before other schools. Using older cohorts in the treatment schools as a control group obviously 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 have a regular trend over the years. A possible way to better take into account strengths and weaknesses of both approaches may be therefore

[26]Schools which applied in the first year could also apply in the second year. However, only 34% of schools which applied in 2009 did it again in 2010 and only half of them got selected in the second year (13 out of 24 schools applying in both years). Among these selected schools in 2010, only 2 schools, corresponding to 15%, is present in our subsample.

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

combining the *Early - Late Enrollees* analysis with the cohort one: we look at the triple difference between cohorts of treated and control schools.

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 Enrollees* schools with *Late Enrollees* schools in the years 2008/2009 - when there was no programme - and 2009/2010 - when the programme got introduced.

5.1.1 Average treatment approach

Our specification (1) is a difference-in-difference model with school fixed effects:

$$Y_{ijt} = \beta_0 + \delta_t + \rho_j + \beta_1 treatment_j * post_t + \beta_2 X'_{ijt} + \varepsilon_{ijt} \quad (1)$$

The outcome variables Y_{ijt} are final marks in Serbian and Mathematics, probability to get an insufficient average, 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. Our 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 s/he 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.

The results of the regressions for the different outcomes of interest are re-

ported in Table 4.[28] For all outcomes we estimate the regressions with controls (columns 1). We then split our sample by gender (columns 2 and 3). 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 4 to 7). We expect that pupils from 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. We choose to investigate further the impact of the programme on two vulnerable groups by differentiating between migrants and non migrants (columns 8 and 9) and between late starters and non late starters (columns 10 and 11). The reason is that we do not have data on whether and how much the assistant work with each child, but we know that assistants are expected to work more with the weakest children. Migrants, often refugees from Kosovo[30], and late starters are likely to experience more difficult conditions with respect to other children and are expected to perform worse in schools.

Overall, the results suggest that the programme had only a statistically significant impact on hours of absences which fell by 0.14 standard deviations.[31] This is especially the case for male, whose reduction in absences is of 0.22 standard deviations. Marks in Mathematics and Serbian, and the probability to get an

[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 is 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.

[30]21% of migrants in our data was born in Kosovo.

[31]Pupils exposed to the programme were on average 3 to 4 days less absent from school than pupils not exposed to it. On average, Roma pupils are absent from school 28 days in a year.

Table 4: Average treatment approach

	(1) with controls	(2) female	(3) male	(4) first grade	(5) second grade	(6) third grade	(7) fourth grade	(8) migrant	(9) non migrant	(10) late starter	(11) non late starter
MATHEMATICS											
post	0.065	0.096	0.041	-0.076	0.040	0.208*	0.168*	-0.087	0.103	0.036	0.040
	(0.062)	(0.080)	(0.056)	(0.175)	(0.126)	(0.117)	(0.085)	(0.063)	(0.069)	(0.076)	(0.060)
treatment*post	0.030	0.015	0.053	0.213	-0.085	0.040	-0.156	0.187	-0.012	-0.072	0.064
	(0.077)	(0.091)	(0.085)	(0.221)	(0.192)	(0.146)	(0.158)	(0.123)	(0.083)	(0.131)	(0.081)
SERBIAN											
post	0.046	0.079	0.027	-0.048	0.099	0.222***	0.109	-0.108	0.091*	-0.021	0.031
	(0.048)	(0.055)	(0.050)	(0.132)	(0.133)	(0.079)	(0.076)	(0.073)	(0.051)	(0.072)	(0.047)
treatment*post	0.012	-0.035	0.058	0.121	-0.130	-0.110	-0.075	0.265**	-0.048	-0.012	0.026
	(0.066)	(0.075)	(0.080)	(0.177)	(0.207)	(0.097)	(0.177)	(0.098)	(0.068)	(0.119)	(0.062)
No. observations	3961	1916	2045	989	1111	988	873	591	3370	878	3083
INSUFFICIENT											
post	-0.039	-0.053	-0.027	-0.041	-0.007	-0.090	-0.044	-0.011	-0.050*	-0.074	-0.017
	(0.028)	(0.035)	(0.031)	(0.079)	(0.062)	(0.055)	(0.028)	(0.035)	(0.028)	(0.060)	(0.021)
treatment*post	-0.031	0.013	-0.075	-0.045	0.014	-0.020	-0.038	-0.051	-0.021	0.010	-0.040
	(0.039)	(0.041)	(0.051)	(0.099)	(0.098)	(0.069)	(0.078)	(0.061)	(0.040)	(0.090)	(0.031)
DROPOUT											
post	0.015**	0.001	0.027**	0.067***	-0.014*	0.011***	-0.007	0.025	0.011**	0.033	0.012*
	(0.006)	(0.010)	(0.012)	(0.016)	(0.008)	(0.004)	(0.011)	(0.016)	(0.004)	(0.028)	(0.007)
treatment*post	-0.006	0.003	0.028*	-0.018	-0.037	0.027*	-0.004	0.003	0.006	0.037	-0.003
	(0.009)	(0.014)	(0.014)	(0.025)	(0.013)	(0.011)	(0.020)	(0.027)	(0.006)	(0.038)	(0.008)
No. observations	4039	1951	2088	1005	1140	1009	885	615	3424	905	3134
ABSENCES											
post	32.853***	22.456***	42.034***	64.693**	19.900	21.515*	-0.007	35.488*	29.331***	26.388	37.998***
	(9.078)	(10.797)	(10.764)	(28.341)	(23.608)	(17.049)	(0.020)	(18.866)	(5.012)	(27.384)	(4.675)
treatment*post	-16.679*	-4.713	-26.119**	-28.336	-19.406	-10.048	0.045**	-39.673	-9.294	-12.740	-17.783**
	(9.078)	(10.797)	(10.764)	(28.341)	(23.608)	(17.049)	(0.020)	(25.205)	(8.355)	(28.282)	(7.474)
No. observations	3868	1871	1997	945	1084	980	885	576	3492	853	3015

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

insufficient grade would suggest that the programme had a positive impact on Roma pupils but coefficients do not show to be significant. Only migrants in *Early Enrolles* schools see their marks in Serbian to significantly increase by 0.24 standard deviations. Late starters are not affected by the programme.

5.1.2 Intensity of treatment approach

It is worth investigating the intensity of the programme. Each school has only one assistant implying that the higher 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 four quartiles, depending on the number of Roma children per school: schools in the fourth quartile have the highest number of Roma, while schools in the first quartile have the lowest.[32] The main difference to the prior model is that include dummies for the quartiles and that we interact these dummies with treatment and time. The omitted category is the first quartile.

The intensity of treatment is modeled as follows:

[32]The average number of Roma between the two years - pre- and treatment year - is used in order to define the quartiles. 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 in each school and we obtain comparable results.

$$Y_{ijt} = \beta_0 + \delta_t + \beta_1 treatment_j + \beta_2 treatment_j * post_t + \sum_{k=2}^{4} \alpha_k quartile.Roma_{kjt} + \tag{2}$$

$$+ \sum_{k=2}^{4} \gamma_k quartile.Roma_{kjt} * post_t + \sum_{k=2}^{4} \delta_k quartile.Roma_{kjt} * treatment_j +$$

$$+ \sum_{k=2}^{4} \phi_k quartile.Roma_{kjt} * treatment_j * post_t + \varepsilon_{ijt}$$

We use the same control variables as in the previous regressions. The coefficient β_2 tells us the impact of the programme on the first quartile. Coefficients ϕ_k are the coefficients of interest for the other three quartiles, whereas the effect of the programme on these quartiles is $\beta_2 + \phi_k$. The regressions results are in Tables 5 and 6. Again, we have the main specification with controls in column (1) and we look at the impacts by gender, grade, migrant status of the child and whether s/he is late or no late starter.

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. Note, however, that schools with few and many Roma are probably different, i.e. they have different characteristics, because the number of Roma per school is not randomly assigned. For instance, it is possible that Roma in schools with less Roma are children of more educated parents who are also more integrated and that children in schools with more Roma have a worse socioeconomic background. Thus from our results we are not able to infer whether the programme could have the same effect in schools with more Roma if it was replicated with the same intensity as in schools with few Roma.

Turning now to results, we find that absences, for instance, reduce by 17 days in a year when the school is in the first quartile of schools' distribution; when the school is in the second quartile, this reduction is almost nullified. This is

Table 5: Intensity of treatment - A

	(1) with controls	(2) female	(3) male	(4) first grade	(5) second grade	(6) third grade	(7) fourth grade	(8) migrant	(9) non migrant	(10) late starter	(11) non late starter
MATHEMATICS											
treatment*post	0.521***	0.803***	0.204	1.880**	0.965***	-1.271***	1.264**	1.714***	0.275	0.510	0.548***
	(0.171)	(0.173)	(0.233)	(0.707)	(0.348)	(0.290)	(0.480)	(0.411)	(0.237)	(0.461)	(0.144)
quartile_2*treatment*post	-0.325*	-0.464**	-0.116	-1.497**	-1.180***	2.237***	-1.552***	-1.845***	-0.036	-0.228	-0.422**
	(0.190)	(0.199)	(0.251)	(0.728)	(0.375)	(0.458)	(0.526)	(0.533)	(0.250)	(0.528)	(0.167)
quartile_3*treatment*post	-0.500**	-0.815***	-0.137	-1.883**	-0.666	1.456***	-1.574**	-1.559***	-0.265	-0.751	-0.446**
	(0.231)	(0.269)	(0.288)	(0.811)	(0.474)	(0.440)	(0.615)	(0.452)	(0.289)	(0.500)	(0.203)
quartile_4*treatment*post	-0.636***	-1.022***	-0.220	-1.823**	-1.252***	0.983***	-1.349**	-1.640***	-0.418	-0.479	-0.689***
	(0.190)	(0.196)	(0.251)	(0.719)	(0.414)	(0.354)	(0.539)	(0.445)	(0.252)	(0.499)	(0.177)
SERBIAN											
treatment*post	0.358	0.450*	0.242	1.080	0.763**	-0.848***	0.801**	1.983***	0.009	0.894**	0.285
	(0.231)	(0.253)	(0.273)	(0.697)	(0.363)	(0.303)	(0.374)	(0.324)	(0.217)	(0.436)	(0.201)
quartile_2*treatment*post	-0.234	-0.209	-0.216	-0.433	-0.883**	1.719***	-1.627**	-1.972***	0.132	-0.695	-0.231
	(0.256)	(0.303)	(0.291)	(0.733)	(0.434)	(0.458)	(0.641)	(0.399)	(0.244)	(0.516)	(0.251)
quartile_3*treatment*post	-0.277	-0.469	-0.050	-0.871	-0.307	0.491	-0.784	-1.578***	0.015	-0.961**	-0.137
	(0.271)	(0.310)	(0.323)	(0.716)	(0.463)	(0.383)	(0.602)	(0.360)	(0.264)	(0.448)	(0.236)
quartile_4*treatment*post	-0.467*	-0.648**	-0.268	-1.081	-1.092**	0.517	-0.722	-1.963***	-0.133	-0.753	-0.452**
	(0.241)	(0.260)	(0.286)	(0.697)	(0.431)	(0.317)	(0.436)	(0.347)	(0.231)	(0.481)	(0.210)
No. observations	3961	1916	2045	989	1111	988	873	591	3370	878	3083

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

Table 6: Intensity of treatment - B

	(1) with controls	(2) female	(3) male	(4) first grade	(5) second grade	(6) third grade	(7) fourth grade	(8) migrant	(9) non migrant	(10) late starter	(11) non late starter
INSUFFICIENT											
treatment*post	-0.137***	-0.208***	-0.053	-0.562***	-0.102	0.040	0.003	-0.292	-0.082***	-0.246*	-0.092**
	(0.036)	(0.063)	(0.080)	(0.108)	(0.060)	(0.128)	(0.068)	(0.207)	(0.025)	(0.136)	(0.040)
quartile_2*treatment*post	-0.002	0.107	-0.120	0.248*	0.288**	-0.456***	0.084	0.571**	-0.110*	0.076	-0.019
	(0.069)	(0.077)	(0.125)	(0.141)	(0.117)	(0.160)	(0.181)	(0.278)	(0.063)	(0.229)	(0.057)
quartile_3*treatment*post	0.051	0.214**	-0.126	0.516**	-0.161	-0.085	-0.015	0.087	0.014	0.262	-0.020
	(0.082)	(0.082)	(0.142)	(0.217)	(0.172)	(0.197)	(0.170)	(0.243)	(0.078)	(0.186)	(0.071)
quartile_4 **treatment*post	0.147**	0.263**	0.019	0.650***	0.106	0.088	-0.190	0.207	0.109**	0.200	0.130**
	(0.059)	(0.099)	(0.094)	(0.129)	(0.118)	(0.179)	(0.118)	(0.218)	(0.050)	(0.180)	(0.053)
DROPOUT											
treatment*post	-0.007	-0.027	0.016	-0.160*	0.095	-0.000	0.065	-0.043	-0.001	0.079	-0.013
	(0.039)	(0.040)	(0.043)	(0.090)	(0.065)	(0.066)	(0.056)	(0.152)	(0.018)	(0.169)	(0.021)
quartile_2*treatment*post	0.015	0.141*	-0.091	0.094	-0.058	0.025	-0.023	0.233	-0.011	-0.036	0.013
	(0.041)	(0.059)	(0.056)	(0.104)	(0.074)	(0.070)	(0.073)	(0.170)	(0.024)	(0.181)	(0.021)
quartile_3*treatment*post	0.002	0.034	-0.033	0.102	-0.104*	0.004	-0.019	0.001	0.001	-0.077	0.005
	(0.040)	(0.042)	(0.048)	(0.099)	(0.060)	(0.065)	(0.058)	(0.153)	(0.022)	(0.174)	(0.025)
quartile_4*treatment*post	0.008	0.035	-0.02	0.121	-0.066	-0.007	-0.05	0.033	0.004	-0.077	0.011
	(0.040)	(0.045)	(0.045)	(0.094)	(0.069)	(0.069)	(0.059)	(0.153)	(0.022)	(0.176)	(0.025)
No. observations	4039	1951	2088	1005	1140	1009	885	615	3424	905	3134
ABSENCES											
treatment*post	-86.799**	-105.188***	-69.135*	-202.805*	-114.608	-50.972	-20.640	-298.926***	-27.393	-232.652***	-51.766
	(34.779)	(35.008)	(39.464)	(102.579)	(68.565)	(53.117)	(56.768)	(94.120)	(23.535)	(68.128)	(33.304)
quartile_2*treatment*post	85.352**	177.351***	10.457	126.853	189.609**	15.556	-8.544	481.240***	2.247	260.127***	54.448
	(36.654)	(45.228)	(41.894)	(107.047)	(79.331)	(70.835)	(113.733)	(137.503)	(28.152)	(77.132)	(35.658)
quartile_3*treatment*post	96.056**	126.206***	67.658	188.665*	69.964	73.242	34.052	265.233***	44.936	312.349***	41.565
	(35.623)	(39.147)	(42.771)	(108.743)	(73.882)	(56.059)	(70.644)	(97.551)	(27.328)	(78.981)	(34.237)
quartile_4*treatment	77.140**	99.709***	56.312	197.019*	93.017	82.396	-77.249	289.510***	18.310	212.908**	40.988
	(36.359)	(35.929)	(41.663)	(106.602)	(73.063)	(56.186)	(61.571)	(100.346)	(25.927)	(84.849)	(34.373)
No. observations	3868	1871	1997	945	1084	980	885	576	3492	853	3015

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

especially the case for female, for whom being in a school with a lower number of Roma seems to be more favourable: on average, if exposed to the programme, not only they are absent from school 21 days less, but also their marks in Mathematics and Serbian increase respectively by 0.72 and 0.39 standard deviations, and their probability to get an insufficient average reduces by 20 percentage points.[33] Striking are the results on migrants and late starters in schools of the first quartile in terms of absences. Migrants also show to perform much better in Mathematics and Serbian when in a *Early Enrollees* first quartile school. Again, these results are almost nullified starting from the second quartile.

5.1.3 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 run the same regressions (regression (1) and (2)) for the years 2006/2007 versus 2007/2008 and for the years 2007/2008 versus 2008/2009 (Table 10 in the Appendix). 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.

[33] On average, female are absent from school 28 days in a year. Their average mark in Mathematics is 2.45 and in Serbian 2.66. The probability to get an insufficient average corresponds to 16%.

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 select the first grade because assistants are expected to work mainly with them. We first estimate the following regression for *Early Enrollees*:

$$Y_{ijt} = \beta_0 + \beta_1 young_i + \beta_2 post_t + \beta_3 young_i * post_t + \varepsilon_{ijt} \quad (3)$$

where Y_{ijt} are again final marks in Serbian and Mathematics, probability to have an insufficient average, 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.

The same regression (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:

$$\begin{aligned} Y_{ijt} = \beta_0 + \beta_1 young_i + \beta_2 post_t + \beta_3 young_i * post_t + \gamma_1 treatment_j * post_t + \\ + \gamma_2 young_i * treatment_j + \gamma_3 young_i * post_t * treatment_j + \varepsilon_{ijt} \end{aligned} \quad (4)$$

The regressions are estimated with controls as in previous specifications. We also look at the impacts by gender, grade, migrant status of the child and whether s/he is a late or no late starter. Results are in Tables 7, 8 and 9.

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 Enrollees* school increases marks in Mathematics and Serbian

Table 7: Cohort regression - A

	with controls			female			male		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MATHEMATICS	T	C	ALL	T	C	ALL	T	C	ALL
young	-0.141	-0.104	-0.077	-0.141	-0.063	-0.105	-0.119	-0.161	-0.043
	(0.150)	(0.145)	(0.132)	(0.134)	(0.224)	(0.215)	(0.251)	(0.230)	(0.171)
post	0.083	0.106	0.132**	0.116**	0.165*	0.190**	0.067	0.072	0.098*
	(0.050)	(0.062)	(0.054)	(0.049)	(0.087)	(0.071)	(0.070)	(0.062)	(0.058)
young*post	0.082	-0.241	-0.291*	-0.019	-0.344	-0.426*	0.180	-0.205	-0.239*
	(0.116)	(0.143)	(0.152)	(0.132)	(0.234)	(0.241)	(0.155)	(0.142)	(0.135)
treatment*post			-0.059			-0.080			-0.045
			(0.074)			(0.088)			(0.086)
young*treatment			-0.102			-0.041			-0.157
			(0.198)			(0.240)			(0.250)
young*post*treatment			0.381*			0.428			0.412*
			(0.194)			(0.279)			(0.207)
SERBIAN									
young	-0.073	-0.011	0.035	-0.175	-0.035	-0.060	0.040	0.004	0.130
	(0.118)	(0.154)	(0.119)	(0.117)	(0.160)	(0.145)	(0.192)	(0.278)	(0.202)
post	0.051	0.077	0.118**	0.037	0.104	0.159**	0.082	0.079	0.106**
	(0.056)	(0.050)	(0.045)	(0.051)	(0.072)	(0.063)	(0.081)	(0.049)	(0.041)
young*post	0.079	-0.255**	-0.300***	0.052	-0.284	-0.352**	0.101	-0.295*	-0.328**
	(0.102)	(0.104)	(0.101)	(0.113)	(0.173)	(0.170)	(0.161)	(0.154)	(0.139)
treatment*post			-0.079			-0.131			-0.043
			(0.074)			(0.086)			(0.086)
young*treatment			-0.147			-0.123			-0.168
			(0.138)			(0.154)			(0.222)
young*post*treatment			0.382**			0.416*			0.423*
			(0.149)			(0.215)			(0.212)
No. observations	2395	1567	3962	1180	736	1916	1215	831	2046

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

by respectively 0.36 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 2) 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.[34] Participation of assistants in regular lessons, organisation

[34]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.

Table 8: Cohort regression - B

	with controls			female			male		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
INSUFFICIENT	T	C	ALL	T	C	ALL	T	C	ALL
young	0.284***	0.236***	0.266***	0.281***	0.257***	0.288***	0.280***	0.211**	0.240***
	(0.065)	(0.043)	(0.046)	(0.049)	(0.080)	(0.084)	(0.097)	(0.081)	(0.059)
post	-0.071**	-0.045	-0.039*	-0.055*	-0.083**	-0.054*	-0.094**	-0.019	-0.029
	(0.031)	(0.033)	(0.022)	(0.028)	(0.033)	(0.029)	(0.045)	(0.045)	(0.032)
young*post	-0.003	0.028	0.011	0.067	0.051	0.036	-0.068	0.020	0.003
	(0.052)	(0.068)	(0.067)	(0.063)	(0.111)	(0.100)	(0.063)	(0.081)	(0.084)
treatment*post			-0.031			-0.002			-0.058
			(0.039)			(0.041)			(0.054)
young*treatment			-0.005			-0.024			0.016
			(0.077)			(0.098)			(0.097)
young*post*treatment			-0.014			0.031			-0.074
			(0.088)			(0.119)			(0.109)
DROPOUT									
young	-0.002	0.011	0.002	-0.002	0.013	0.009	-0.005	0.008	-0.008
	(0.006)	(0.014)	(0.007)	(0.013)	(0.017)	(0.017)	(0.010)	(0.015)	(0.009)
post	0.015**	-0.002	-0.003	0.021*	-0.019	-0.016	0.009	0.010	0.007
	(0.006)	(0.010)	(0.006)	(0.010)	(0.012)	(0.010)	(0.009)	(0.015)	(0.014)
young*post	0.014	0.079***	0.080***	0.031	0.071***	0.077***	-0.003	0.087***	0.083***
	(0.013)	(0.019)	(0.018)	(0.020)	(0.019)	(0.018)	(0.012)	(0.028)	(0.028)
treatment*post			0.018**			0.038**			0.003
			(0.009)			(0.015)			(0.016)
young*treatment			0.005			-0.003			0.015
			(0.006)			(0.016)			(0.012)
young*post*treatment			-0.066***			-0.047*			-0.087***
			(0.022)			(0.027)			(0.029)
No. observations	2438	1601	4039	1200	751	1951	1238	850	2088
ABSENCES									
young	3.123	37.687*	18.656	-8.373	39.502	26.379	12.221	32.323	8.644
	(8.355)	(20.554)	(17.503)	(11.240)	(23.527)	(21.174)	(14.033)	(23.473)	(17.397)
post	10.022	30.724***	20.728**	7.655	21.190*	8.570	13.007	37.766***	30.488***
	(5.975)	(9.912)	(7.856)	(7.172)	(10.707)	(8.771)	(8.865)	(11.015)	(9.396)
young*post	23.579*	57.493**	54.639**	40.707**	67.618**	67.603**	9.740	52.423**	49.480**
	(11.537)	(21.263)	(22.235)	(16.000)	(28.938)	(26.469)	(14.055)	(22.036)	(24.287)
treatment*post			-10.409			-0.785			-17.036
			(10.354)			(11.898)			(12.891)
young*treatment			-0.868			-19.597			16.848
			(16.384)			(19.622)			(16.766)
young*post*treatment			-31.867			-28.524			-40.337
			(24.945)			(30.772)			(27.852)
No. observations	2336	1532	3868	1152	719	1871	1184	813	1997

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

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.[35] However, we find that the increase in absences is smaller in

[35]In treated schools, absences increase by 1 day (29 days of average absences); in control schools by 2 days (37 days of average absences).

Table 9: Cohort regressions by subgroups

	migrant			non migrant			late starter			non late starter		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	T	C	ALL	T	C	ALL	T	C	ALL	T	C	ALL
MATHEMATICS												
young*post	0.095	-1.328***	-1.280***	0.093	-0.005	-0.060	-0.218	0.055	-0.074	0.164	-0.198	-0.271
	(0.340)	(0.246)	(0.262)	(0.115)	(0.136)	(0.140)	(0.230)	(0.329)	(0.344)	(0.137)	(0.151)	(0.183)
young*post*treatment			1.424***			0.168			-0.138			0.447*
			(0.425)			(0.182)			(0.407)			(0.227)
SERBIAN												
young*post	0.328	-1.209***	-1.141***	0.060	-0.032	-0.089	0.177	-0.321	-0.412	0.101	-0.132	-0.213
	(0.336)	(0.258)	(0.280)	(0.100)	(0.124)	(0.107)	(0.184)	(0.327)	(0.290)	(0.130)	(0.125)	(0.135)
young*post*treatment			1.482***			0.157			0.563			0.317
			(0.417)			(0.146)			(0.341)			(0.189)
INSUFFICIENT												
young*post	0.076	0.210*	0.205*	-0.012	-0.019	-0.038	0.084	0.041	0.049	-0.042	-0.004	-0.014
	(0.190)	(0.105)	(0.105)	(0.047)	(0.071)	(0.071)	(0.066)	(0.133)	(0.124)	(0.055)	(0.075)	(0.073)
young*post*treatment			-0.134			0.023			0.045			-0.027
			(0.217)			(0.088)			(0.146)			(0.091)
DROPOUT												
young*post	-0.029	0.055	0.056	0.020	0.080***	0.082***	0.027	0.181**	0.160**	0.012	0.056***	0.059***
	(0.036)	(0.059)	(0.059)	(0.014)	(0.023)	(0.023)	(0.047)	(0.065)	(0.060)	(0.009)	(0.019)	(0.018)
young*post*treatment			-0.087			-0.064**			-0.125			-0.047**
			(0.069)			(0.027)			(0.075)			(0.020)
No. observations	311	304	615	2127	1297	3424	568	337	905	1870	1264	3134
ABSENCES												
young*post	-9.963	66.857	68.441	26.788**	50.431*	46.649*	29.451	-23.600	-33.643	18.882	61.719***	63.210***
	(40.296)	(67.480)	(66.753)	(11.412)	(23.996)	(24.361)	(31.309)	(66.445)	(63.947)	(11.187)	(20.571)	(18.327)
young*post*treatment			-80.342			-22.217			68.614			-44.595**
			(77.954)			(26.872)			(72.658)			(21.744)
No. observations	294	282	576	2042	1250	3292	545	308	853	1791	1224	3015

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

Early Enrollees than in *Late Enrollees* schools, confirming the results obtained by the average treatment approach also in the magnitude of the difference (5/6 school days). Although the coefficient γ_3 of column 3 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 the same school.

Consistent results with the average treatment approach are also obtained for migrants: marks of treated first graders increase considerably, by 1.29 standard deviation in Mathematics and by 1.34 standard deviation in Serbian, on average. Late starters confirm not to be affected by the programme.

5.2.1 Placebo Regressions

Again, we need to control for the robustness of our results by running regressions (regression (3), (4)) for the years 2006/2007 versus 2007/2008 and for the years 2007/2008 versus 2008/2009 (see Table 11 in the Appendix). The two placebo tests suggest that our results are robust.

5.3 Spillover effects

It remains to investigate whether this programme affected also 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. Detailed results are provided in Tables 12 and 13 in the Appendix.[36]

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.

[36]Placebo tests for the pre-treatment year (2007/2008) are here reported.

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.

The comparison of Roma and Non Roma schooling outcomes suggest that the current differences in marks between Roma and Non Roma are enormous. Moreover, Roma children are much more absent from school than their Non Roma colleagues. The stark differences in all main schooling outcomes underline the necessity in this context for government programmes aiming at helping Roma pupils.

Results of our analysis suggest that the programme had 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 3 to 4 days more to school. 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 female and migrants, for whom being in a school with a lower number of

Roma turns out to be more favourable. We also find that, on average, marks have improved by almost 0.36 standard deviations and dropouts have reduced by 6.6 percentage points for those children exposed to the programme in their first grade.

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 has looked 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. It would be worth rethinking to assign more than one assistant to schools with a large number of Roma.

References

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), 159–217.

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*.

A Tables

Table 10: 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
INSUFFICIENT						
treatment*post	-0.031	-0.035	-0.025	0.051	0.034	0.058
	(0.027)	(0.032)	(0.038)	(0.037)	(0.028)	(0.052)
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 in parentheses:
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: 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
INSUFFICIENT						
young*post	-0.090	-0.082	-0.075	0.092	0.117*	0.119*
	(0.097)	(0.068)	(0.068)	(0.076)	(0.066)	(0.068)
young*post*treatment			-0.033			-0.015
			(0.117)			(0.101)
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 in parentheses:
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Average treatment approach - Non Roma

Treatment (2008/2009 and 2009/2010)				
	(1)	(2)	(3)	(4)
	Mathematics	Serbian	Insufficient	Absences
post	0.011	-0.011	-0.002	5.025***
	(0.018)	(0.031)	(0.003)	(1.185)
treatment*post	0.029	0.054	0.001	-1.586
	(0.025)	(0.034)	(0.003)	(1.725)
No. observations	14981	14982	15085	14686
Placebo 2007/2008 and 2008/2009				
post	0.058**	0.052**	-0.002	-0.518
	(0.027)	(0.025)	(0.002)	(2.874)
treatment*post	-0.043	-0.029	0.001	3.749
	(0.033)	(0.029)	(0.002)	(3.106)
No. observations	15345	15345	15448	15052

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

Table 13: Cohort regression - Non Roma

Treatment (2008/2009 and 2009/2010)	Mathematics			Serbian			Insufficient			Absences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	T	C	ALL	T	C	ALL	T	C	ALL	T	C	ALL
young*post	0.044	-0.062	-0.070	0.033	-0.056	-0.066	-0.002	0.009	0.009	0.658	3.178	3.259
	(0.058)	(0.085)	(0.079)	(0.072)	(0.077)	(0.069)	(0.006)	(0.008)	(0.008)	(2.682)	(3.274)	(3.057)
young*post*treatment			0.113			0.099			-0.010			-2.491
			(0.098)			(0.100)			(0.010)			(4.138)
No. observations	8232	6749	14981	8232	6750	14982	8288	6797	15085	8099	6587	14686

Placebo 2007/2008 and 2008/2009	Mathematics			Serbian			Insufficient			Absences		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	T	C	ALL	T	C	ALL	T	C	ALL	T	C	ALL
young*post	-0.099*	-0.056	-0.053	-0.118**	-0.089	-0.087	0.005	0.007	0.009	0.610	-1.718	-2.440
	(0.050)	(0.104)	(0.099)	(0.047)	(0.101)	(0.097)	(0.007)	(0.011)	(0.011)	(2.847)	(3.922)	(3.935)
young*post*treatment			-0.049			-0.037			-0.004			2.787
			(0.111)			(0.107)			(0.012)			(4.986)
No. observations	8573	6772	15345	8573	6772	15345	8631	6817	15448	8435	6617	15052

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

CHAPTER 2: The Curse of Low Aspirations: Does a Remedial Education Programme Change Perceived Returns to Education of Roma People?*

Marianna Battaglia† Lara Lebedinski‡

Abstract

We use first hand collected data to examine how parents of children exposed to a remedial education programme change expectations about their future. We are interested in changes in expectations on the likelihood of the child to find a job once adult, on her expected salary and on the highest level of education she would achieve. We argue that these changes are likely to occur through a role model mechanism. In the Roma Teaching Assistant Programme we study, all the assistants are Roma and from the same social background of pupils they help. Our data are obtained by a survey conducted in 5 municipalities of Belgrade one year after its introduction: we targeted Roma households with primary school aged children enroled

*We would like to thank Tito Boeri, Eliana La Ferrara, Andreas Madestam and Michele Pellizzari for continous encouragement and very helpful discussions throughout this project. We are especially indebted to the personnel of the Ipssos Strategic Marketing Agency, in particular Dragiša Bjeloglav and Tatjana Jovanov, for the successful data collection. This study would not have been possible without the support of Božidar Jovanović, Tünde Kovacs-Cerović and the other personnel of the Ministry of Education. All opinions expressed are of the authors, all errors are our own.

†Department of Economics, Bocconi University. Email:marianna.battaglia@phd.unibocconi.it

‡Department of Economics, Bocconi University. Email:lara.lebedinski@phd.unibocconi.i

in schools exposed to the programme in its two subsequent phases. Results suggest that parents of pupils in treated schools expect them to achieve higher education levels, while job market perspectives remain unchanged.
Keywords: aspirations, perceived returns to education, programme evaluation, Roma
JEL classification codes: I25, J13, D04

1 Introduction

Recent research in economics of education emphasises how aspirations for the future wellbeing affect consistently the incentive to invest in the present and how the social environment plays a role in shaping one's aspirations. The aim of this paper is to examine the impact of a remedial education programme on aspirations of children and parents belonging to the marginalised Roma minority group. There are several barriers of access to education for Roma, the largest ethnic minority in Europe: absence of documents, financial constraints, child labour, discrimination from teachers and pupils and language barriers.[1] Costs as compared to the discounted stream of expected future benefits are likely perceived as too high by them. In the recent years, at least in Serbia, schools started enroling children with incomplete documents and all first grade pupils got free text books. Nonetheless, with the exception of Hungary, which has a very high enrolment rate in primary school among Roma (95%), in most countries they are still in the range of 40% to 60%.[2] The completion rates of primary school are even lower: only 30% to 40% of Roma adults have completed compulsory primary

[1]Roma pupils often 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. Some children may also face difficulties at school due to 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.

[2]Roma 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.

education.[3] Certainly not enough effort has been done in order to reduce such constraints: costs are still pretty high for many Roma families.[4] Nonetheless, besides that, much of the low completion rates can be explained by the little value attached to schooling and education by parents.[5] School is not expected to give enough future opportunities: there is often discrimination in the formal job market for people coming from minority groups and the informal one do not often require education. They would there collect rubbish, sell goods on the market or do for low skilled jobs. Perceived benefits of going to school may be so low compared to the respective costs that there is no reason to invest. A vicious cycle of low aspirations arises. Roma are poorer than other population groups and more likely to fall into poverty and remain poor. They suffer severe social exclusion in terms of overrepresentation among low skilled jobs and no participation in the political and cultural life and this is persistant over time. Trying to stop these mechanisms would be beneficial for the societies where they mainly live and in turn for the current destination countries where their recent migration flows have often lead to the appearance of informal settlements, increased number of unemployed and inadequacy of the education system in receiving new foreign pupils.[6]

[3]There is reason to believe that even 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 then count as finished primary schools although the requirements in these schools are much lower. 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 they are mainly attended by pupils who are late in enroling and by pupils who decided to return to school after dropping out.

[4]On average, in Serbia costs associated with school (books, other school material, excursion) 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. For 10% of them these costs even ranged between 12% to 25% of yearly household income.

[5]A majority of Roma parents has low educational attainment: they do not even complete primary school, corresponding to 8 schooling years. In our sample, mother's years of schooling completed are, on average, 5 while father's years of schooling completed are 7.

[6]Recently, especially in the European Union, Roma have started to consistently attract media attention. The visa liberalisation and the adhesion to the Union of countries like Romania and Bulgaria, in which the percentage of Roma population is very high, have indirectly led

Our measures of aspirations are the expected probability that the child will find a job with primary or secondary school, expected returns to secondary education and expectations of parents' about the highest level of education their children will attain. We use a simple difference approach by exploiting the gradual implementation of the programme to identify its impact on aspirations. We find that parents whose children were exposed to the programme are more likely to expect their children to attain a higher level of education. An examination of heterogeneous effects suggests first, that our results on highest expected level of education are driven by responses from non-Muslim parents. Second, parents may revise their expectations in response to the programme for both younger (6 to 10 years) and older boys (11 to 15 years), while in the case of girls only the expectations for the younger generations change. Job market perspective seem to remain unchanged.

We argue that these changes are likely to occur through a role model mechanism. In the Roma Teaching Assistant Programme we study, all the assistants are Roma and from the same social background of pupils they help. Our paper is mainly related to the contributions of Ray (2004), Genicot and Ray (2010) and Nguyen (2008). 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. The existence of unrealistic aspirations induces frustration. On the contrary, "the "best" sort of aspirations are those that induce a reasonable distance between one's current living standards and where one wants to be" (Genicot and Ray (2010),p.1). There are reasons to believe that the lack of goals and aspirations is an important factor influencing the educational decision in the case of Roma. First, a large percentage of Roma live in segregated settlements. According to the 2002 Census, in Serbia

to significant migration flows towards the Western countries. The extraordinariness of the phenomenon has led to hot discussions within the European countries and civil society and increased the interest of the European Union on those countries which will likely enter the Union in the future and where a high percentage of Roma population resides, e.g. Serbia.

83% of self-declared Roma live in census tracts with at least 7% of the Roma population.[7] Since the majority of Roma is isolated from the mainstream society, they do not have often role models to which they can relate to in their immediate neighbourhood (this ideas is line with the argument of Wilson (1987)). Secondly, it is extremely rare that Roma people are doing any skilled jobs and that children and their parents can relate to other Roma outside of their environment. For instance, there are no teachers of Roma origin working in schools. They can also not be found in any public offices.

The Roma Teaching Assistant (RTA) Programme mainly aims to improve schooling outcomes of the disadvantage Roma children. It consists of assigning one person of Roma origin, the so called teaching assistant, to a school that is being attended by at least 5% of Roma children. Depending on the needs of the children and the school, the assistant helps children during regular classes or she organises additional classes. Moreover, one day per week is dedicated to work with children's parents. This work consists of visiting parents whose children are not coming to school as well as informing other children's parents about the progress their children are making at school.[8] The RTA also provides children and parents with a role model which should lead to a betterment of their aspirations.

For the purpose of this analysis, we have conducted an extensive survey in 300 Roma households with both parents and their children in 12 different settlements of Belgrade in 2010. The RTA programme was introduced in 2009 and we look at its impact after a year of introduction.

The main econometric challenge of our study is the identification. The introduction of Roma teaching assistants was not randomised and we construct a control group of parents and children who attend schools which received teaching assistants at a later point of time. We use a large set of control variables to control for observable confounders and for all outcomes we separately estimate

[7] According to the Census, which underestimates the Roma population, there were only 1.4% self-declared Roma in Serbia in 2002. Estimates suggest instead they are approximately 4−6% of the overall population.

[8] In a companion paper (Battaglia and Lebedinski, 2011) we evaluate the effect of the programme on schooling outcomes.

the regressions for boys and girls.

Related literature. 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. In reality, however, we often observe that individuals underinvest in education given its high returns and low direct costs. The literature has so far examined various factors such as credit constraints, high discount rates and low school quality which could contribute to an underinvestment in education.[9] More recently, several works have questioned the assumption of perfect information of the returns to education and emphasised the importance of subjective expectations (Manski, 1993; Jensen, 2010; Nguyen, 2008; Kaufmann and Attanasio, 2009). It is the returns perceived by individuals that affect schooling decisions, and these perceptions may be inaccurate, due to limited or imperfect information. This gap can be filled by providing additional information through statistics (Jensen, 2010) or through a role model − an actual person sharing his/her success story (Nguyen, 2008). These tools turn out to be most cost−effective solutions than incentives, like cash transfers or private school vouchers.

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 in the US (Dee, 2004). However, to the best of our knowledge there are no studies 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 and elicit directly from children their role models, whereas we argue that assistants of the RTA Programme may be perceived as role

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

models. The paper is therefore trying to fill this gap in the literature and, together with its companion paper (Battaglia and Lebedinski, 2011), 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 and there are few attempts to investigate how to improve life circumstances of Roma in general and Roma kids in particular. We also contribute to the existing literature by providing primary data in a context where data are scarce and inaccurate.

In the background of our work is the literature on residential segregation and neighbourhood effects, which studies the relevance of neighbourhoods and ones peers in influencing socioeconomic outcomes.[10] 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). The neighbourhood were one lives can clearly affect ones labour market (see for instance Clark and Drinkwater (2002); Edin et al. (2003); Bayer et al. (2008); Boeri et al. (2011)) and educational outcomes (see for instance 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 follow. Section 2 gives some information on the background and the way the survey has been designed. Section 3 describe our data and provide some descriptive statistics. Section 4 presents the estimation strategy and our results. Section 5 discusses findings and concludes.

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

2 Background: Context and Roma Teaching Assistant Programme

2.1 Roma in Serbia

Official data on Roma in Serbia are scarce and inaccurate.[11] The reason for the absence of data is that there is generally very little awareness of the importance of data collection in Serbia among the policy makers. The reason for the inaccuracy of data is that Roma often do not declare themselves as Roma in surveys. Most Roma consider themselves both Roma and Serbian and the question of nationality allows only one answer.[12]

The 2002 Census puts forward a number of 108.000 Roma or 1.44% of the total population. Estimates, on the other hand, suggest a number between 350.000 and 500.000 or approximately 4-6% of the overall population (Open Society Institute, 2007). The Living Standard Measurement Survey (LSMS) from 2003 provides rich information on the living conditions of the Roma population. 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 survey are alarming. Two out of three Roma households were poor implying that their average consumption was below the poverty line.[13] Almost half of the Roma population (42%) is younger than 18 years. Only 56% of children from Roma settlements aged 7 to 14 attend school.[14] Among the adults, 25% have no schooling and another 36% have not finished primary school. The employment rate among the males is very similar to the non Roma population (70%), but the female employment rate is very low with

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

[12] 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.

[13] The percentage of the extremely poor among the interviewed Roma was 11.9%.

[14] The percentage of children who attend school at the age of 7 to 10 is 65% (of whom 8% attend special schools for mentally handicapped children), whereas at the age of 11 to 14 this percentage is 46%.

35%. The LSMS confirms that Roma live in difficult conditions and that they constitute a marginalised minority.

2.2 The Roma Teaching Assistant Programme

The Roma Teaching Assistant programme is the main programme in Central and Eastern European countries targeting Roma inclusion in education.[15] After the initial pilot phase, the programme took off in 2009/2010 on a larger scale. In 2009, 26 schools (*Early Enrollees*) entered the RTA programme and in the following year an additional 79 schools (*Late Enrollees*) joined.

The schools had to apply and they 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 (this was only the case for schools in 2009[16]). The requirement for assistants were the following: (1) secondary school attainment, (2) knowledge of Romani and (3) preferable experience in working with children.[17]

Each school was assigned only one teaching assistant, and although the schools were somewhat free in allocating the time of the assistant, their major tasks in the school were to help children during regular classes and to organise after-school extra classes. One day per week assistants visit parents of children who are not coming to school and inform other parents about their children's progress.

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

[16]In 2010, 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 followed 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.

[17]We argue that there is not a problem of selection bias in this framework: schools applying the first year are not different in observable characteristics from schools applying later. The selection criteria remain the same and the committee which decided the schools selected rated them based on their shown interest and motivation in the same way, in both years. Unfortunately 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. However, schools which applied in the first year could also apply in the second year and at least a quarter of them did it.

In a companion study on the impact of the programme on schooling outcomes in its first year of implementation (Battaglia and Lebedinski, 2011), we find that the programme had a positive effect. There is evidence that children exposed to the programme went more to school and that, on average, marks have improved and dropouts have reduced 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 especially for females and migrants.

3 Data and Descriptive Statistics

3.1 The Survey Design

We use first hand collected data obtained through a survey conducted with 300 Roma households in 12 settlements in 5 municipalities of Belgrade.[18] The survey took place in autumn 2010.[19] The households in our sample have either children who were enroled in an *Early Enrollee* school or in a *Late Enrollee* school. Figure 1 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 Enrollee* schools. Settlements 6 to 13 had assistants starting from 2010/2011.

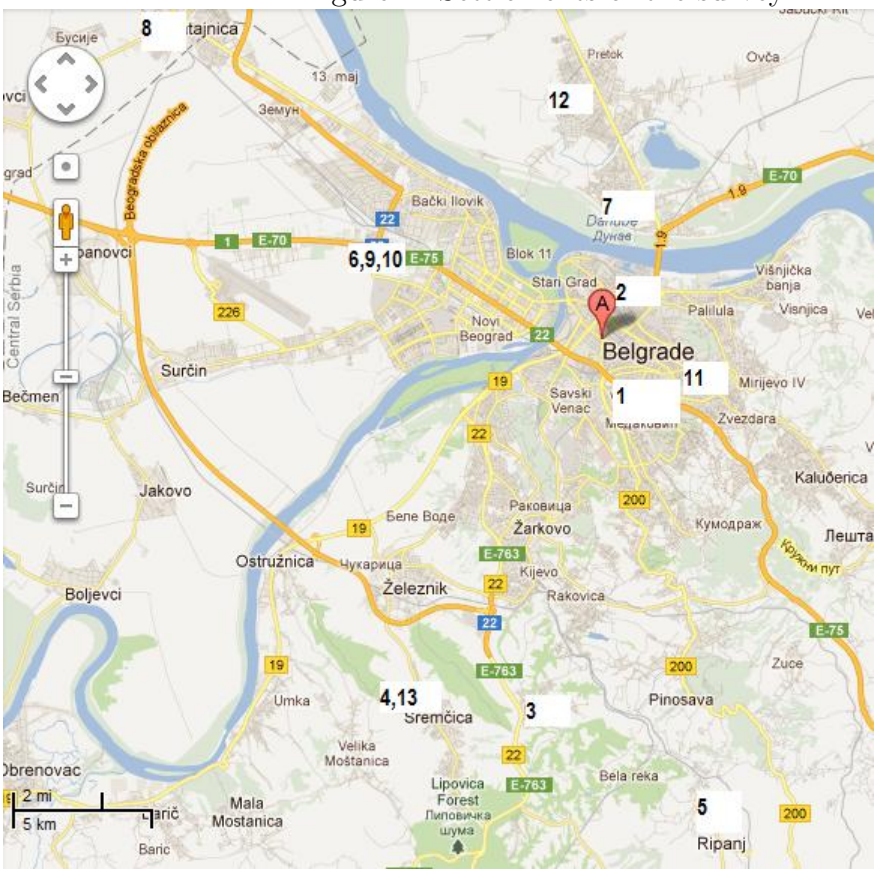
We are interested in the effect of the RTA programme on children from lower four grades, given that the assistants mainly work with them, and hence our sample was constructed in such a way that all households have at least one child in the lower four grades of primary school in the scholastic year 2009/2010.

Three sets of questionnaires were 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

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

[19]In 2010 schools received the assistant in November/December, after the survey was conduced.

Figure 1: Settlements of the survey



children. *Mother*'s questionnaire consisted of an extensive series of questions about the education of the children aged 6 to 15 years living in the household. Some questions on child labour are also asked in this section. Children who attended first to fourth grade of primary school in 2009/2010 were asked information about the school and teachers. The *child* questionnaire also contains a quick test on their abilities to read and write and doing some mathematics.

3.2 The Sample

Our sample can be divided in two groups. The first group consists of 122 households with children exposed to the programme in 5 schools which got a Roma

teaching assistant in 2009/2010. These households were selected randomly from lists of students provided by schools and correspond to the treated group. The 178 remaining households were randomly selected from settlements in Belgrade close to the 8 schools which received the RTA programme in 2010/2011 and are our control group.[20] 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' expectations on aspirations are created at the household level: once one is exposed to the programme, expectations on future opportunities change for all children of the same household. However, there is only one assistant per school and not every Roma child in the treated school is actually followed by her. We can therefore explore another definition of being treated beside the main one. Our second definition considers a household to be treated only if at least one child is in an *Early Enrollee* school and there is evidence from the responses that the assistant has in fact worked with her. A household is in this case treated if either (1) parents stated that there is someone in the school who helps the kid with his/her homework or s/he is following additional classes at school, or (2) there is someone from the school who ever come to her place or called he because of the kid, or (3) the child responded that there is someone else (besides the teacher) who helps him/her in class.[21]

Treated and control groups are comparable in terms of observable characteristics: their differences are not statistically significant in most of the cases (see Tables 1 and 2). Wealth, income, educational attainments and household

[20] 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.

[21] We decided not to ask explicitly the parents whether the school of their children is in the RTA programme, because it was not clear to us whether the parents are aware of the name of the programme and how they perceive the teaching assistant, i.e. as assistants, teachers, etc.. Parents' and children's answers do rarely differ. In those cases, parents' information are used. We also consider children's responses for the creation of the treatment variable because of the presence of missing values in the parents' responses.

composition do not differ between groups.[22] However, among non-treated households there are significantly more Muslim households, suggesting that it would be worthy to investigate the impact of the programme also taking into account differences in religion (see section 4.3 on heterogeneous effects).

Overall, some interesting observations emerge from the summary statistics of our sample (Table 3). First, a substantial proportion of households (32%) have at least one person working in the informal sector. Second, somewhat surprising with respect to official data, only few households have someone with unfinished primary school (7%) and a relatively large share of households has at least a person with finished secondary school (19%).

3.3 Outcome Variables

We use three different sets of questions to understand whether the programme was effective in increasing parents' aspirations for 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/his own aspirations.

The first and second set of questions relates to expected returns to education. They could be asked to either mother or father (or caretaker), however the mother was the main interviewee in 92% of cases. They were asked for the oldest boy and the oldest girl. These outcomes are at the household level.[23]

The first set of questions investigates the impact of the programme on the likelihood of getting a job with primary school as the highest degree achieved and likelihood of getting a job with secondary school as the highest degree achieved.[24]

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

[23]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.

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

Table 1: Means of control variables in treated and control households - Variables at the household level

Variables at the household level	Treatment	Control	Difference
Wealth[a]	-0.14	0.22	-0.36
			(0.27)
Monthly Total income (in dinars)[b]	28224.39	29453.33	-1228.94
			(2144.78)
Informal (=1)[c]	0.31	0.32	-0.01
			(0.05)
Urban (=1)	0.47	0.53	-0.06
			(0.06)
Only Roma in settlement (=1)[d]	0.28	0.16	0.12**
			(0.05)
No schooling (=1)[e]	0.04	0.04	0.00
			(0.02)
Unfinished primary school (=1)[e]	0.02	0.02	0.00
			(0.01)
Finished primary school (=1)[e]	0.67	0.76	-0.09
			(0.05)
Finished secondary school (=1)[e]	0.23	0.16	0.07
			(0.05)
Muslim (=1)	0.57	0.80	-0.23***
			(0.05)
Number of children under 5	0.75	0.70	0.05
			(0.10)
Number of female children between 6 and 18	1.73	1.59	0.14
			(0.14)
Number of male children between 6 and 18	1.77	1.80	0.10
			(0.13)
Number of adults	2.46	2.44	0.02
			(0.12)

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 a household member.

Table 2: Means of control variables in treated and control households - Individual level characteristics

Variables at the individual level	Treatment	Control	Difference
Children characteristics			
Male (=1)	0.50	0.54	-0.04*
			(0.04)
Age of child	10.11	9.74	0.37
			(0.20)
Rank among siblings	2.60	2.35	0.25***
			(0.09)
Demeaned grade in mathematics[a]	-0.01	0.01	-0.02
			(0.09)
Demeaned grade in Serbian[a]	-0.02	0.01	-0.03
			(0.09)

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

[a] The grade has been demeaned from the average grade of the school.

Though we are mainly interested in the probability of finding a job with secondary school, we use for comparison reasons both outcomes. We ask "Assume that your oldest boy(girl) has finished primary school (and that is his(her) highest degree) and he(she) is 25-30 years old: how certain are you that he(she) will have any kind of job?". The same question has been asked for "finished secondary school". The responses to this question come from a five point Likert scale and they are "Absolutely sure", "Quite sure", "Maybe", "Unlikely" and "No, he/she 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 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, despite this drawback, we decided to offer responses because of the low educational level of our respondents. 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 his/her

Table 3: Summary statistics: Control variables

Variables at the household level	Mean	Std. dev.	Obs.
Wealth[a]	0.08	2.17	290
Monthly Total income (in dinars)[b]	28949.47	18931.88	300
Informal (=1)[c]	0.32	0.47	284
Urban (=1)	0.51	0.50	300
Only Roma in settlement (=1)[d]	0.21	0.41	300
No school/unfinished primary school (=1)[e]	0.07	0.25	300
Finished primary school (=1)[e]	0.74	0.44	300
Finished secondary school (=1)[e]	0.19	0.39	300
Muslim (=1)	0.71	0.45	300
Number of children under 5	0.72	0.80	300
Number of female children between 6 and 18	1.65	1.21	300
Number of male children between 6 and 18	1.75	1.03	300
Number of adults	2.44	1.00	300
Treated household 1[g]	0.41	0.49	299
Treated household 2[h]	0.24	0.43	299
Variables at the individual level			
Children characteristics			
Male (=1)	0.52	0.50	673
Age of child	9.89	2.56	673
Rank among siblings	2.20	1.17	673

[a] The wealth index is 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 a household member.
[f] The grade has been demeaned from the average grade of the school.
[g] At least one of the children goes to a school which has an assistant.
[h] At least one of the children goes to a school which has an assistant and either the child or both parents declare that there is someone helping that child at school.

child will not find a job for a given 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 third of the respondents answered "Unlikely" and "No, he/she will not find a job" in the case of primary school and "Absolutely sure", "Quite sure" and "Maybe" in the case of secondary (see Figures 2 and 3). Nonetheless, ordered logit analysis suggests that some categories may not be collapsed (see Table 4). 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. Both estimates, with a dummy outcome and a categorical outcome, are provided.

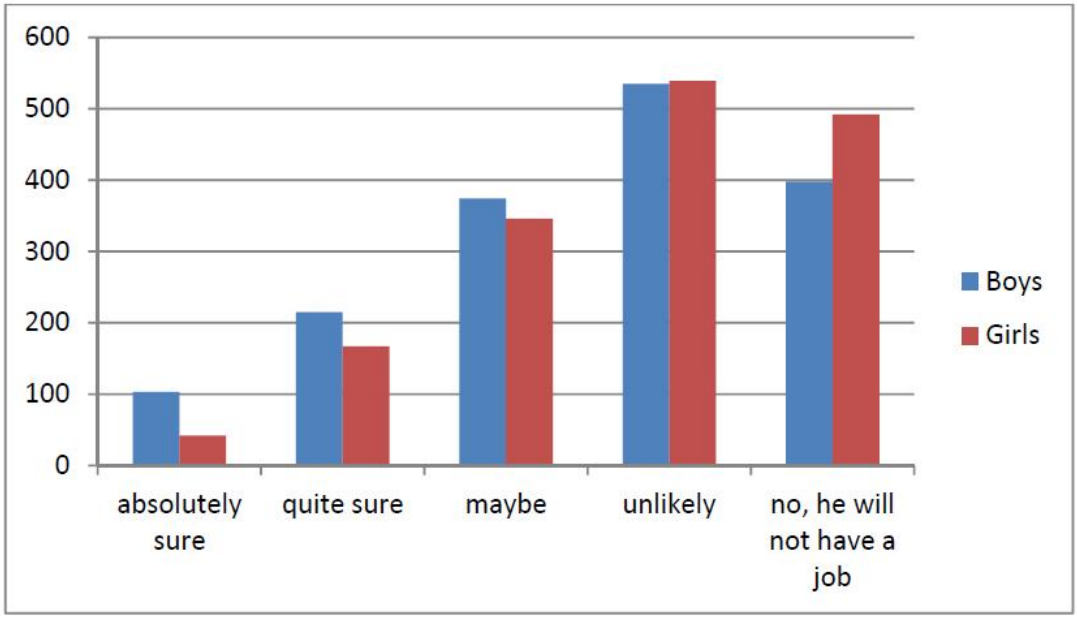
Table 4: Ordered Logit - thresholds among categories

	Primary		Secondary	
	Boys	Girls	Boys	Girls
	(1)	(2)	(3)	(4)
treatment	0.114	0.134	-0.322	0.200
	(0.276)	(0.311)	(0.278)	(0.294)
controls	yes	yes	yes	yes
cut1				
constant	-2.959***	-3.201***	-0.307	0.334
	(0.650)	(0.842)	(0.802)	(1.036)
cut2				
constant	-1.488**	-1.346*	1.389*	2.509**
	(0.613)	(0.806)	(0.804)	(1.062)
cut3				
constant	-0.077	0.237	2.744***	3.902***
	(0.598)	(0.791)	(0.814)	(1.066)
cut4				
constant	1.550***	1.796**	4.874***	6.334***
	(0.595)	(0.794)	(0.926)	(1.157)
No. observations	263	255	263	254

Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The second set of questions elicits minimum and maximum amount parents expect that their children will earn. We ask "Assume that your oldest boy has

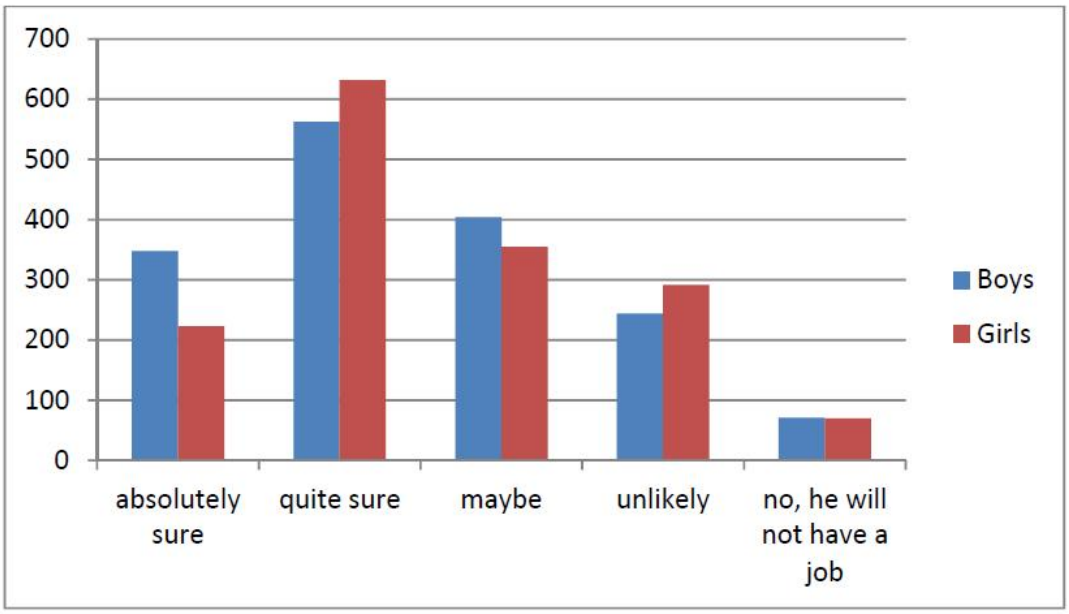
Figure 2: Likert scale for the probability to find a job with primary school by gender



finished primary school (or secondary) and this is his highest degree and he 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. Secondary school earnings are however associated with regular type of jobs. The same questions have been asked for the girl. We elicited the minimum and maximum earnings and we use their (log) average as our measure of expected earnings.

The third relevant outcome is the highest expected level of education of the child. The exact question was: "What level of formal education do you think that *(name)* will complete?" This question was asked for each child between 6 to 15 years old. We created the dummy "secondary as the highest level of education" which takes the value 1 when it was answered "secondary (or more)" and 0 otherwise. Summary statistics for the outcome variables in our sample are

Figure 3: Likert scale for the probability to find a job with secondary school by gender



reported in table 5.

As expected, the probability to find a job and its expected salary increase with education for both boys and girls. Roma expectations show to be in line with their current returns to education and much lower than the average in Belgrade. Figure 4 reports the average monthly earnings according to the Serbian statistical office (2011) and the average Roma earnings in our sample by gender. Differences in expectations between treated and control groups are also presented.

In treated households expected future earnings are higher than in control households for both primary school and secondary school. The fact that expected monthly salary are higher in treated households may suggest that the program was successful in raising expected returns to education. This difference, however, does not turn out to be statistically significant in most of the cases. Descriptive statistics show that the expected mean earnings are higher in treated households only for girls (Table 6).

Table 5: Summary statistics: Outcome variables

Variables at the household level	Mean	Std. dev.	Obs.
Probability to find a job: Boys[a]			
With primary school (=1)	0.42	0.49	295
With secondary school (=1)	0.82	0.38	296
Probability to find a job: Girls[a]			
With primary school (=1)	0.35	0.48	287
With secondary school (=1)	0.79	0.41	285
Expected mean log earning: Boys			
With primary school	9.91	0.34	124
With secondary school	10.21	0.30	241
Expected mean log earning: Girls			
With primary school	9.82	0.32	99
With secondary school	10.14	0.30	224
Variables at the individual level			
Expected to finish : Boys			
Secondary school[b]	0.61	0.49	299
Expected to finish : Girls			
Secondary school[b]	0.63	0.48	275

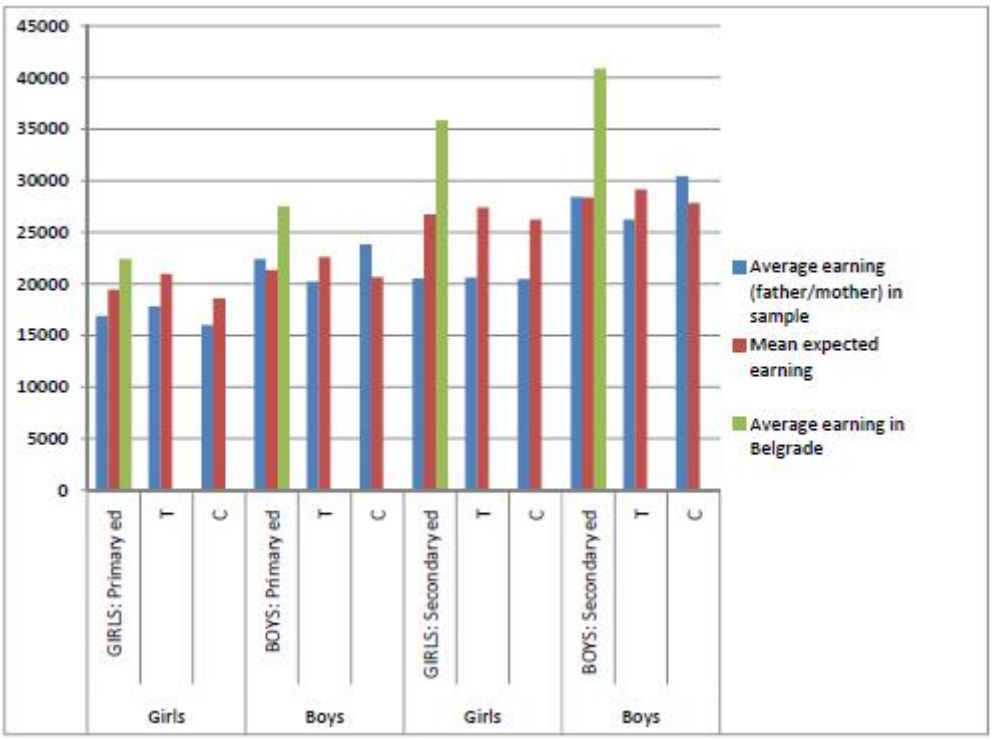
[a] Respondent expects the child to find a job with a given education level.

[b] Parents' expectation on child's highest educational level.

The table 6, which compares all outcomes between *Early Enrollee*−treated and *Late Enrollee*−comparison households, also highlights that respondents in treated households are more likely to expect their male and female children to finish secondary school. This difference is only significant for boys.[25]

[25]Figure 5 in the Appendix shows the percentage of respondents who expect their child to finish (1) no school, (2) at most primary school, (3) at most secondary school and (4) more than secondary school in treated and control schools for both genders. It demonstrates that there is a larger share of parents who expect their child to finish at most secondary school among treated schools compared to control schools. This difference in highest expected level comes from a shift from primary to secondary education level.

Figure 4: Comparison of real and expected returns to education



4 Estimation Strategy and Results

The provision of a role model, through the Roma Teaching Assistant Programme, affects parents' aspirations about their children in three different ways. First, we expect that treated parents think that their children are more likely to find a job with secondary education. Second, we expect that parents expect higher salaries conditional on having a secondary school degree. And third, we expect that parents believe it is more likely that their child will finish secondary school.

The first specification is the following:

$$Y_j = \alpha_0 + \alpha_1 treatment_j + \alpha_2 X'_j + \varepsilon_j \quad (1)$$

where Y_j are the outcomes of interest for the household j: likelihood of finding

Table 6: Means of outcome variables in treated and control households

Variables at the household level	Treatment	Control	Difference
Probability to find a job: Boys[a]			
With primary school (=1)	0.35	0.48	-0.13**
			(0.06)
With secondary school (=1)	0.82	0.82	0.00
			(0.05)
Probability to find a job: Girls[a]			
With primary school (=1)	0.31	0.39	-0.08
			(0.06)
With secondary school (=1)	0.74	0.82	-0.07
			(0.05)
Expected mean log earning: Boys			
With primary school	9.97	9.87	0.10
			(0.06)
With secondary school	10.24	10.18	0.06*
			(0.19)
Expected mean log earning: Girls			
With primary school	9.90	9.78	0.12*
			(0.07)
With secondary school	10.18	10.11	0.07*
			(0.04)
Variables at the individual level			
Expected to finish : Boys			
Secondary school (=1)	0.67	0.57	0.10*
			(0.06)
Expected to finish : Girls			
Secondary school (=1)	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 with a given education level.

a job with primary school as the highest degree, likelihood of finding a job with secondary school as the highest degree; (log) mean amount of earning per month depending on the degree expected to be obtained.

For the outcome "secondary school as the highest degree expected", a second specification is introduced and standard errors are clustered at the household level:

$$Y_{ij} = \beta_0 + \beta_1 treatment_{ij} + \beta_2 X'_{ij} + \nu_{ij} \quad (2)$$

$treatment_j$ and $treatment_{ij}$ stand for whether there is at least a child of the household who goes to a treated school and X'_{ij} 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, the maximum level of education in the household and household composition characteristics. In the individual level regression we also add age, age squared, the gender of the child, ranking among siblings, demeaned mark in Mathematics and Serbian.[26]

4.1 Main Results

Results for the probability to find a job, expected earnings and highest expected level of education are reported respectively in Table 7, Table 8 and Table 9. For consistency, all the estimates are estimated with ordinary least squares. Probit estimates for the two dummy outcomes are presented in Table 13 in the Appendix.

The coefficient for the expected probability to find a job is negative and not significant in almost all specifications both for boys and girls. It is larger in absolute terms for primary school as highest degree than secondary school, suggesting a possible positive trend in expectations. Given the low statistical significancy of the results, however we would better argue that job market perspective of those exposed to the programme remain unchanged. Having at least a child in a treated school does not change parents' expectations for their children's future opportunities compared to having no children in a treated school. Higher possibilities in terms of future employment are expected by wealthier households. Being in a

[26]The grade has been demeaned from the average grade of the school.

urban and segregated settlement impacts negatively on the expected probability to find a job with secondary education mainly for girls.

Table 8 shows the results for the expected (log) mean earnings per month. Overall, they are higher in treated households for both girls and boys. The coefficients are significant only for the secondary school level: on average, being in a treated household increases the expected monthly earnings by almost 8% for boys and 9% for girls, conditional on having achieved secondary school. Again, wealthier families have higher expectations and urban households less. Not surprisingly, the highest level of education in the household matters for expected returns to education. However, while higher education levels in the household has a differential impact on expected earnings for boys, it does not for girls. In their case having someone with secondary school in the household instead of adults with at maximum only primary school does not make a difference in terms of expected earnings. Interestingly, Muslim households expect on average higher earnings conditional on the education level for both genders.

The regression results for secondary education as the highest level attained are reported in Table 9. For both boys and girls we find that parents in treated households are more likely to expect their children to finish secondary school. Yet, the effect on completion of secondary education is significant only for boys and it is relatively large. On average, parents exposed to the programme are 11.6 percentage points more likely to expect their male children to finish secondary education than parents not exposed to the programme.

4.2 A Different Definition of Treated Households

In our main specification we define a household as treated if at least one child from that household goes to a school with a Roma teaching assistant. The school does not keep track of the names of the children with whom the assistant interacted, but our survey allows us to know whether the child in the *Early Enrollee* school was actually followed by the assistant in the first year of the programme. Both effects are indeed of interest. If a child is in a treated school but not followed

Table 7: Probability to find a job with primary/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)
treatment	-0.124**	-0.075	-0.007	0.003	-0.075	-0.112*	-0.077	-0.042
	(0.058)	(0.063)	(0.046)	(0.047)	(0.057)	(0.064)	(0.050)	(0.051)
wealth		-0.022		0.028**		0.009		0.035***
		(0.016)		(0.012)		(0.015)		(0.012)
informal		0.092		-0.053		0.065		0.027
		(0.063)		(0.054)		(0.065)		(0.050)
urban		-0.265***		-0.060		-0.292***		-0.096*
		(0.061)		(0.052)		(0.059)		(0.051)
only Roma in settlement		-0.081		-0.023		0.116		-0.150**
		(0.087)		(0.064)		(0.081)		(0.073)
finished primary school		-0.018		-0.048		-0.131		0.056
		(0.115)		(0.091)		(0.121)		(0.099)
finished secondary school		0.097		0.016		-0.084		0.080
		(0.131)		(0.104)		(0.142)		(0.117)
number of children under 5		0.045		-0.028		0.005		-0.022
		(0.042)		(0.035)		(0.041)		(0.036)
number of male children between 6 and 18		-0.014		0.044		-0.050		-0.008
		(0.037)		(0.031)		(0.035)		(0.032)
number of female children between 6 and 18		-0.016		0.029		-0.015		-0.008
		(0.036)		(0.026)		(0.032)		(0.031)
number of adults		-0.023		-0.084**		0.035		-0.046
		(0.034)		(0.033)		(0.035)		(0.034)
muslim		0.262***		0.029		0.083		0.030
		(0.066)		(0.055)		(0.068)		(0.061)
constant	0.474***	0.474***	0.824***	0.982***	0.383***	0.599***	0.825***	0.972***
	(0.038)	(0.144)	(0.029)	(0.120)	(0.038)	(0.150)	(0.030)	(0.136)
No. observations	295	270	296	270	287	262	285	261
R2	0.015	0.138	0.000	0.059	0.006	0.121	0.009	0.118

Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Expected log mean earnings with primary/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)
treatment	0.102	0.081	0.070*	0.079*	0.118*	0.074	0.073*	0.091*
	(0.063)	(0.075)	(0.038)	(0.043)	(0.065)	(0.084)	(0.039)	(0.049)
wealth		0.010		0.012		0.028*		0.027**
		(0.017)		(0.010)		(0.017)		(0.011)
informal		0.029		0.055		0.025		0.062
		(0.073)		(0.048)		(0.081)		(0.049)
urban		-0.180***		-0.109***		-0.144		-0.098**
		(0.068)		(0.042)		(0.090)		(0.043)
only Roma in settlement		0.039		0.036		0.037		0.100
		(0.090)		(0.051)		(0.088)		(0.068)
finished primary school		0.165		0.136**		0.181		0.163**
		(0.120)		(0.067)		(0.143)		(0.073)
finished secondary school		0.226		0.181**		0.227		0.163*
		(0.149)		(0.087)		(0.181)		(0.096)
number of children under 5		-0.064		-0.023		-0.058		-0.019
		(0.043)		(0.024)		(0.051)		(0.027)
number of male children between 6 and 18		0.037		0.007		-0.000		-0.011
		(0.044)		(0.024)		(0.037)		(0.025)
number of female children between 6 and 18		0.048		0.033		0.023		0.022
		(0.035)		(0.022)		(0.041)		(0.026)
number of adults		0.031		-0.033		-0.024		-0.018
		(0.048)		(0.025)		(0.037)		(0.026)
muslim		0.051		0.126**		0.073		0.133**
		(0.089)		(0.052)		(0.124)		(0.062)
constant	9.871***	9.543***	10.179***	9.994***	9.778***	9.644***	10.116***	9.895***
	(0.039)	(0.164)	(0.026)	(0.088)	(0.040)	(0.215)	(0.028)	(0.110)
No. observations	124	115	241	220	99	93	224	209
R2	0.020	0.155	0.013	0.117	0.032	0.167	0.014	0.138

Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Secondary school as the highest expected level of education by gender

Gender	Boys		Girls	
	(1)	(2)	(3)	(4)
treatment	0.100	0.116*	0.105	0.081
	(0.071)	(0.065)	(0.071)	(0.084)
wealth		0.046**		0.045**
		(0.018)		(0.021)
informal		-0.211***		-0.049
		(0.073)		(0.078)
urban		-0.070		0.023
		(0.069)		(0.076)
only Roma in settlement		-0.051		-0.167
		(0.092)		(0.113)
finished primary school		0.046		0.175
		(0.169)		(0.212)
finished secondary school		0.044		0.232
		(0.187)		(0.219)
number of children under 5		0.003		0.127
		(0.071)		(0.079)
number of male children between 6 and 18		0.064		-0.055
		(0.059)		(0.058)
number of female children between 6 and 18		-0.022		-0.052
		(0.049)		(0.057)
number of adults		0.012		-0.006
		(0.050)		(0.046)
age of child		0.057		0.123
		(0.146)		(0.181)
age of child squared		-0.003		-0.003
		(0.006)		(0.008)
rank among siblings		-0.071		0.052
		(0.053)		(0.054)
demeaned grade in mathematics		0.108**		0.044
		(0.042)		(0.057)
demeaned grade in Serbian		0.070		0.080
		(0.044)		(0.058)
muslim		-0.191**		-0.226***
		(0.080)		(0.082)
constant	0.569***	0.486	0.585***	-0.343
	(0.045)	(0.849)	(0.047)	(1.037)
No. observations	299	212	275	198
R2	0.010	0.343	0.011	0.251

Robust standard errors corrected for clustering at the household level in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

by an assistant she would however be aware of her existence: the role model mechanism would work, although likely with a lower intensity. We therefore create an additional variable *assistant* which is equal to 1 if the child is in fact helped by the assistant, and 0 otherwise. The variable is equal to 1 only for treated households and individuals. This discussion gives us two additional specifications, one for the regressions at the household level, regression (3) and the other for regressions at the individual level, regression (4):

$$Y_j = \theta_0 + \theta_1 treatment_j + \theta_2 treatment_j * assistant_j + \theta_3 X'_j + \upsilon_j \qquad (3)$$

$$Y_{ij} = \gamma_0 + \gamma_1 treatment_{ij} + \gamma_2 treatment_{ij} * assistant_{ij} + \gamma_3 X'_{ij} + u_{ij} \qquad (4)$$

Y_j and Y_{ij} are the outcomes of interest for the individual i in the household j, namely likelihood of finding a job with primary school as the highest degree, likelihood of getting a job with secondary school as the highest degree; (log) mean amount of earnings per month depending on the degree expected to be obtained; secondary school as the highest degree expected.

The effect of the RTA programme on households (individuals) who have interacted with the assistant equals the sum of the coefficients $\theta_1 + \theta_2$ $(\gamma_1 + \gamma_2)$, whereas the effect of the programme on households treated who have not directly interacted with the assistant is θ_1 (γ_1). The baseline group is the control group: pupils in *Late Enrollee* schools. θ_2 (γ_2) alone captures the effect on aspirations for treated children in *Early Enrollee* compared to "untreated" children in *Early Enrollee*. Table 10 summarises the results for the specifications (3) and (4).

The results for the expected probability to find a job are slightly larger in absolute terms than the results from the main specification, but they are still not significantly different from 0. The expected earnings are significantly higher for households whose children have not interacted with the assistant than for households whose children have interacted with the assistant. Especially for boys,

Table 10: All outcomes with primary/secondary school by gender for a different definition of treatment

Gender	Boys				Girls			
Max. level of	Primary school		Secondary School		Primary school		Secondary school	
education	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Probability to find a job with primary/secondary school by gender								
treatment*assistant	-0.116	-0.119	-0.063	-0.040	-0.131	-0.042	-0.168**	-0.082
	(0.090)	(0.090)	(0.069)	(0.078)	(0.087)	(0.090)	(0.076)	(0.079)
treatment	-0.054	-0.002	0.036	0.033	0.005	-0.085	0.029	0.012
	(0.080)	(0.081)	(0.057)	(0.067)	(0.080)	(0.082)	(0.059)	(0.065)
controls	no	yes	no	yes	no	yes	no	yes
No. observations	294	269	295	269	286	261	284	260
R2	0.020	0.142	0.003	0.056	0.013	0.120	0.025	0.115
Expected log earnings with primary/secondary school by gender								
treatment*assistant	-0.174*	-0.096	-0.165***	-0.140***	-0.238**	-0.109	-0.117**	-0.081
	(0.097)	(0.100)	(0.052)	(0.053)	(0.096)	(0.107)	(0.053)	(0.050)
treatment	0.189**	0.126	0.162***	0.159***	0.234***	0.133	0.137***	0.138**
	(0.079)	(0.094)	(0.049)	(0.054)	(0.082)	(0.099)	(0.048)	(0.055)
controls	no	yes	no	yes	no	yes	no	yes
No. observations	124	115	241	220	99	93	224	209
R2	0.042	0.161	0.044	0.136	0.085	0.175	0.029	0.144
Secondary school as the highest expected level of education by gender								
treatment*assistant			0.074	0.120			0.138	0.232***
			(0.104)	(0.084)			(0.092)	(0.085)
treatment			0.066	0.061			0.036	-0.030
			(0.091)	(0.082)			(0.093)	(0.091)
controls			no	yes			no	yes
No. observations			299	212			275	198
R2			0.012	0.348			0.020	0.273

Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

parents of pupils treated in *Early Enrollee* schools reduce their expectations of future monthly earnings on average by 1.4% compared to those of untreated pupils in the same schools. The role model mechanism would suggest a positive coefficient for pupils expose to the programme. Nonetheless, assistants mainly work with more disadvantage students. The fact that she is actually treated can lead to a lowering of parents' aspirations for the child herself. Those who are aware of the existence of the assistant but do not have their children treated expect instead that their children would have better returns to education. The coefficient θ_1 is in fact positive.

The results for expected level of the education are similar to our main specification. In particular, they suggest that parents revised their expectations of the highest level of education when their children are in fact treated especially for girls: on average, parents of pupils treated in *Early Enrollee* schools are 23.2 percentage points more likely to expect their female children to finish secondary education than parents of pupils untreated in the same schools.

4.3 Heterogeneous Effects

In this section we examine heterogeneous effects of the programme on Muslim versus non Muslim households and younger (6 to 10 years) versus older children (11 to 15).

4.3.1 Muslim Households Versus non Muslim Households

The results of our main specification are 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 children to attain a lower level of education for both genders when compared to non Muslim households. Moreover, descriptive statistics show that the difference between treated and control group is statistically significant for number of Muslim households: our control group has significantly more Muslim households. In Table 14 in the Appendix we provide a comparison of household and individual characteristics for Non Muslim and Muslim households. We find significant differences between the two groups of households. Muslim households are poorer, they are more likely to live in an only Roma settlement, they have less educational attainment and they have more children.

The specification (6) includes whether the household is Muslim and the interaction of being Muslim (in a Muslim household) and in a treated household as follows:

$$Y_j = \delta_0 + \delta_1 treatment_j + \delta_2 muslim_j + \delta_3 muslim_j * treatment_j + \delta_4 X'_j + v_j \quad (5)$$

$$Y_{ij} = \lambda_0 + \lambda_1 treatment_{ij} + \lambda_2 muslim_{ij} + \lambda_3 muslim_{ij} * treatment_{ij} + \lambda_4 X'_{ij} + \varsigma_{ij} \quad (6)$$

Y_j and Y_{ij} are the outcomes of interest for the individual i in the household j. The coefficient δ_1 (λ_1) of the variable Muslim captures the difference between Muslims and non Muslims for the whole sample. The coefficient δ_2 (λ_2) captures the effect of treatment on non Muslims, and $\delta_2 + \delta_3$ ($\lambda_2 + \lambda_3$) is the effect of treatment on Muslims.

Our results are in Table 11. The coefficients for the probability to find a job and expected (log) earning are not much different than the main specification and remain not significant. Treated non-Muslim households however increase their expectations for secondary school completion a lot, whereas Muslim households respond much less. Muslims who are in *Early Enrollee* schools are on average 28.8 percentage points less likely to expect their female children to finish secondary education compared to non-Muslim in the same schools.

4.3.2 Young Versus Old Kids

In this section we examine whether younger children of the age 6 to 10 responded differently to the programme than 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. In fact, 56.12% of the young children among treated households are actually being helped, whereas this is the case for only 37.5% among older 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

Table 11: All outcomes for primary/secondary school by gender with interaction for Muslim households

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 by gender				
treatment	-0.101	-0.034	-0.065	-0.041
	(0.096)	(0.085)	(0.100)	(0.083)
muslim	0.244***	0.003	0.118	0.031
	(0.086)	(0.073)	(0.091)	(0.075)
treatment*muslim	0.039	0.056	-0.071	-0.001
	(0.130)	(0.102)	(0.129)	(0.104)
controls	yes	yes	yes	yes
No. observations	270	270	262	261
R2	0.139	0.060	0.122	0.118
Expected log earnings with primary/secondary school by gender				
treatment	-0.079	0.011	-0.216	0.006
	(0.137)	(0.080)	(0.176)	(0.104)
muslim	-0.043	0.077	-0.105	0.068
	(0.121)	(0.077)	(0.176)	(0.104)
treatment*muslim	0.206	0.105	0.400*	0.132
	(0.162)	(0.096)	(0.205)	(0.114)
controls	yes	yes	yes	yes
No. observations	115	220	93	209
R2	0.167	0.122	0.226	0.146
Secondary school as the highest expected level of education by gender				
treatment		0.240**		0.283**
		(0.098)		(0.125)
muslim		-0.092		-0.058
		(0.108)		(0.130)
treatment*muslim		-0.197		-0.288*
		(0.137)		(0.149)
controls	yes	yes	yes	yes
No. observations		212		198
R2		0.350		0.264

Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ones.

We estimate separately the regression (1) for younger and older children by genders. We have individual outcomes only for expected level of education and we estimate only this outcome. The results are shown in Table 12.

Most of our coefficients are not significant due to the small sample size of each

Table 12: Secondary school as the highest expected level of education for young and old cohorts by gender

Gender	Boys		Girls	
Cohort	young	old	young	old
	(1)	(2)	(3)	(4)
treatment	0.127	0.100	0.100	-0.030
	(0.095)	(0.093)	(0.100)	(0.135)
controls	yes	yes	yes	yes
No. observations	104	108	105	93
R2	0.369	0.401	0.394	0.340

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

subgroup, but the magnitude and direction are still suggestive of the effect. We find that there is little difference between younger and older boys: the programme appears to affect the probability to finish secondary school for both groups. We find a similar effect for younger girls, but this effect is not present for older girls, those who are actually closer to the possibility to reach the goal.

5 Conclusion

Aspirations for the future wellbeing affect current investments in education. In this paper we examine the impact of a remedial education programme on aspirations of children and parents belonging to the marginalised Roma minority group. Our measures of aspirations are the expected probability that the child will find a job with primary or secondary school, her expected earnings conditional on a level of education achieved and the expected highest level of education she will attain. In The Roma Teaching Assistant Programme all the assistants are Roma and from the same social background of pupils they help. The changes occurring as a consequence of the exposure to the programme are thus explained through a role model mechanism.

We use a simple difference approach by exploiting the gradual implementation

of the programme to identify its impact on aspirations. Our data, collected one year after the first implementation, suggest that job market perspectives remain unchanged: parents of children exposed to the programme do not change their expectations on returns to education of their children. Nonetheless, they are more likely to expect them to attain a higher level of education. On average, parents in *Early Enrollee* schools are 11.6 percentage points more likely to expect their male children to finish secondary education than parents in *Late Enrollee* schools. For girls, higher expectations for the highest level of education achievable are obtained only for the pupils who are in fact helped by an assistant: on average, parents of pupils treated in *Early Enrollee* schools are 23.2 percentage points more likely to expect their female children to finish secondary education than parents of pupils untreated in the same schools. 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 second that parents revise their expectations in response to the programme for both younger (6 to 10 years) and older boys (11 to 15 years), while in the case of girls only the expectations for the younger generations change.

Overall, these results appear to be encouraging in terms of highest expected levels of education. The presence of a person of the same social background who showed to be successful motivates parents to believe their children can achieve the same results. Nonetheless, not higher possibilities of employment are expected in the future for even the most educated kids. The role model mechanism shows to work only partially, in a short-time horizon. This may be due to individual difficulties of forecasting - ten to fifteen year-time is too long to understand the possible implications of current choices for these people - or by a persistent belief that the situation can not change. One year of a remedial education programme may be not enough to free Roma from the curse of low aspirations.

References

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.

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.

A Tables

Table 13: Dummy outcomes - probit estimates

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 by gender				
treatment	-0.084	-0.002	-0.120*	-0.069
	(0.069)	(0.048)	(0.068)	(0.054)
controls	yes	yes	yes	yes
No. observations	263	263	255	254
Secondary school as the highest expected level of education by gender				
treatment		0.222**		0.141
		(0.089)		(0.104)
controls		yes		yes
No. observations		208		191

Marginal effects; Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Means of control variables in Non Muslim and Muslim households

Variables at the household level	Non- Muslim	Muslim	Difference
Wealth[a]	0.55	-0.12	0.67**
			(0.28)
Monthly Total income (in dinars)[b]	31308.05	27986.1	3321.94
			(2142.39)
Informal (=1)[c]	0.32	0.32	0.00
			(0.06)
Urban (=1)	0.44	0.54	-0.1
			(0.06)
Only Roma in settlement (=1)[d]	0.06	0.28	-0.22***
			(0.05)
No schooling/unfinished primary school (=1)[e]	0.03	0.06	-0.03
			(0.03)
Finished primary school (=1)[e]	0.60	0.77	-0.18***
			(0.06)
Finished secondary school (=1)[e]	0.36	0.12	0.23***
			(0.05)
Number of children under 5	0.59	0.77	-0.19*
			(0.10)
Number of female children between 6 and 18	1.4	1.75	-0.34 **
			(0.15)
Number of male children between 6 and 18	1.66	1.78	-0.13
			(0.13)
Number of adults	2.34	2.47	-0.09
			(0.13)
Variables at the individual level			
Children characteristics			
Male (=1)	0.56	0.51	0.05
			(0.04)
Age of child	10.23	9.75	0.54
			(0.22)
Rank among siblings	2.25	2.19	0.06
			(0.11)
Demeaned grade in mathematics[f]	0.11	-0.04	0.14
			(0.10)
Demeaned grade in Serbian[f]	0.13	-0.05	0.18*
			(0.10)

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.

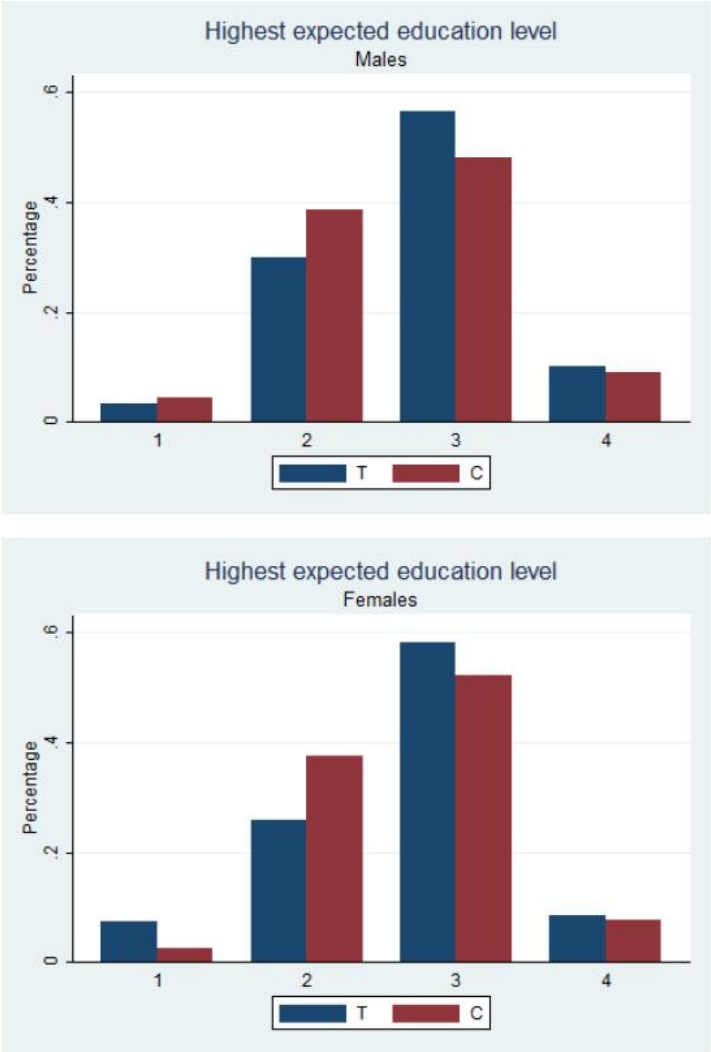
[d] =1 if the respondent declared that the household lives in an exclusively Roma neighbourhood.

[e] It refers to the highest level of education obtained by a household member.

[f] The grade has been demeaned from the average grade of the school.

B Figures

Figure 5: Highest expected education between treated and control groups by gender



1: Unfinished primary, evening or special school
2: Finished primary school
3: Finished secondary
4: More than secondary

CHAPTER 3: Residential Segregation and Labour Market Outcomes of Roma in Serbia*

Lara Lebedinski†
Bocconi University

Abstract

I estimate the impact of residential segregation on employment of Roma in Serbia. Residential segregation is measured as the percentage of Roma in a census tract. I combine data from World Bank's Living Standard Measurement Survey with the census data on segregation. I exploit the within district variation in the percentage of Roma and I use past rates of Roma in a census tract to instrument for present segregation. My results show that Roma in more segregated localities are more likely to have a job and this effect is fully driven by female workers. A one standard deviation increase or equivalently an increase of 10 percentage points in residential segregation raises the probability of a female worker to be employed by 6 to 7 percentage points. My instrument suggests that less employable workers sort into more segregated areas.

Keywords: residential segregation, minorities, Roma

JEL classification codes: J15, J61, R23.

*I am indebted to Tito Boeri and Andreas Madestam for their guidance and support throughout this project.

†Email: lara.lebedinski@phd.unibocconi.it

1 Introduction

Persistent poverty in inner city districts inhabited by black people in the US and increasing clustering of ethnic groups and immigrants in cities in the developed world has led many researchers to try to answer the question on the labour market effects of ethnic and racial geographic segregation.[1]

Similar to other ethnic minorities, a common feature of the Roma population is their spatial segregation from the non Roma population. Roma are a marginalised group and there is concern that the segregation undermines even more their economic opportunities. Understanding whether there is an impact of segregation on labour market outcomes and the exact mechanism of this phenomenon has important and relevant policy implications.

This paper aims to estimate the impact of residential segregation on labour market outcomes of the Roma ethnic group in the Serbian context. In order to eliminate the problem of selection into localities by individuals, I instrument current segregation rates with past segregation. For the purpose of the analysis, I use World Bank's Living Standard Measurement Survey from 2003 and combine it with census data from 1971, 1991 and 2002.

I use both the instrumented and non instrumented probit model and I exploit the within district variation in segregation to estimate the effect of residential segregation on my measure of labour market performance. When I use the whole sample to estimate the probit model I do not obtain significant results.[2] When I divide the sample between female and male workers, I find that female workers in more segregated areas are more likely to be employed. In the case of female workers, a one standard deviation increase in segregation raises female employment by 6 percentage points.

In order to exclude the possibility of sorting into localities, I reduce my sample to individuals younger than 25 and individuals younger than 30 and I obtain

[1]The literature on residential segregation is not limited to labour market consequences of segregation, other outcomes have been examined such as health (Katz et al., 2004), juvenile crime behaviour (Case and Katz, 1991), welfare usage (Bertrand et al., 2003), etc.

[2]My IV probit estimates are close to being significant.

very similar results. To exclude the possibility that my results are driven by some omitted census tract characteristic, I add additional controls for neighbourhood characteristics and this does not alter my results. I use an alternative instrumental variable to confirm my findings. The results are further insensitive to excluding from the sample the capital city (22% of the sample) as well as refugees from Kosovo (8%).

Studies of residential segregation of racial and ethnic groups in the US, UK and Italy (e.g. Cutler and Glaeser (1997); Clark and Drinkwater (2002); Boeri et al. (2011)) have found, on average, a negative effect of segregation on the probability of employment. There are, however, other studies such as Edin et al. (2003) or Damm (2009) which find a positive effect of segregation on immigrant earnings in Sweden and Denmark. The latter two studies attribute the positive effect to information spillovers between immigrants living in enclaves. This mechanism could be at place also in the case of Roma women. This is supported by the fact that in the Serbian labour market referrals play a very important role. I provide evidence that a large share of female workers is employed in the agricultural sector as seasonal workers. I suspect that for this type of work the network channel plays a very important role.

The remainder of the paper is organised as follows: Section 2 gives an overview of the literature related to this study. Section 3 provides a description of the historical background of the Roma minority and it describes the Serbian context. Section 4 explains the data sources and comments on the summary statistics. Section 5 provides the econometric framework for analysing the impact of residential segregation on labour market outcomes. Section 6 concludes.

2 Previous Literature

2.1 Theoretical Findings

The literature explores different theories how the segregation of ethnic minorities can have an effect on labour market outcomes. The different theories do not

exclude each other, they are probably all present in most environments, but their importance depends on the given context.

For instance, under the assumption that there are *firms in enclaves*, the inhabitants of these enclaves will not be exposed to discrimination in firms of their co-ethnics (Becker, 1971). Moreover, Holzer and Ihlanfeldt (1998) uncover that minority workers can even be preferred if a firm's customers are mainly of that minority group. This is especially the case for contact jobs.

Then there is the *social networks effect* which works in two ways: through the information and social norm channel. The information channel captures the fact that individuals communicate and hence, exchange information related to the labour market with each other. A contact can help, for instance, to reduce the application cost for social benefits, provide information on job opportunities, etc. On the other hand the social norms channel works through "peer pressure, stigma and social approval" (Bertrand et al., 2003).

The *spatial mismatch theory*, originally proposed by Kain (1968), postulates that the distance of the so-called ghettos or enclaves from employment opportunities is important. Due to this distance, inhabitants of enclaves are less knowledgeable of new job vacancies, face higher commuting costs and do not respond well to demand shifts.

The *human capital channel* has been explored in a series of works by Borjas (1992, 1994, 1995). The underlying notion is, that the human capital of a child depends not only on the parents' human capital but also on the ethnic capital, that is the average human capital of ones ethnic group. The reason that ethnic capital affects a child's human capital accumulation is that the ethnic capital is a good proxy for the socioeconomic environment of a child.

In their very influential work, Alesina et al. (1999) show that the ethnic composition, in particular ethnic fragmentation of US cities, is inversely related to spending on productive local public goods (such as education or public transportation systems). Their finding underlines that segregation can induce communities to enter a long lasting poverty trap because of low levels of spending on

productive public goods.

2.2 Empirical Approaches

Similar to other economic questions, the literature on residential neighbourhood effects has been plagued by the identification problem. The fact that individuals select a neighbourhood and are obviously not randomly assigned to it, complicates the analysis. Unobserved variables which affect the choice of neighbourhood can also have an impact on labour market outcomes, thus different approaches in the literature have been proposed to circumvent the problem of self-selection.

There are several ways to solve the problem of sorting into neighbourhoods. An obvious possibility to solve the problem of self-selection into neighbourhoods is to use policy experiments in which households are randomly assigned to different neighbourhoods. The most famous such experiment is the Moving to Opportunity (MTO).[3] Similar social experiments have been conducted with immigrants in Sweden (Edin et al., 2003) and Denmark (Damm, 2009).

A second possibility is to assume that there is only sorting within larger areas such as metropolitan areas, but not between metropolitan areas and to aggregate at the metropolitan area (Evans et al., 1992; Cutler and Glaeser, 1997; Card and Rothstein, 2007). This approach then compares outcomes between metropolitan areas and eliminates the unobserved characteristics among neighbours through aggregation.

The third possible technique is to disaggregate data below the level of neighbourhood with the aim to eliminate block-level variation in neighbour attributes (Bayer et al., 2008). This strategy relies on the assumption that the interaction of neighbours takes place at the block level and that there is no correlation in unobservables across blocks within block groups. Boeri et al. (2011) improve on this strategy by instrumenting residential segregation with historical data of block characteristics and thus they take into account also unobservable block

[3]This programme consisted in giving vouchers to households from poor neighbourhoods which they could use to move to lower-poverty neighbourhoods.

characteristics.

3 Background Information on Roma in Serbia

Nowadays Roma can be found in most European countries, but the majority of them has settled in Eastern European countries.[4] Though there are some uncertainties about the exact origin and reasons for their migration, there is consensus among historians that Roma are originally from India and that their first migrations to Europe date back to the 12th century.

This section gives a brief historical overview of the life of Roma in Serbia and it draws largely on Bašić and Jakšić (2005).

3.1 Historical Background of Roma in Serbia

Roma have been inhabitants of Serbia since centuries. There is, for instance, evidence from a Census in 1522. that there were 17,191 Roma houses in Rumelia which was a historical region comprising the territories of the Ottoman Empire in Europe and it included Serbia. Most of the Roma back then were located in the todays southern part of Serbia and around Belgrade.[5] During the Ottoman reign in Serbia, which lasted from the end of 14th century until the beginning of the 19th century, Roma enjoyed an equal treatment as the rest of the population. Evidence from Turkish sources suggests that some Roma groups settled and became sedentary already in the 16th century.[6] The majority of them retained their nomadic life style within the boundaries of the Empire or out of its confines until

[4]The countries with a substantial number of Roma in Eastern Europe are 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.

[5]According to this source, Roma were mainly located in Serbia in Resava, Pirot, Novo Brdo, Novi Pazar, Smederevo, Vranje, Prijepolje mountain Rudni, and in Kosovo in Priština, Vučitrn, Prizren and Peć.

[6]Depending on their origin and religion, Roma are often divided in subgroups. The main distinction between them is the one between Muslims and Christians, who are then divided into other subgroups within each community.

the late 19th century. After Serbia regained independence from the Ottomans in the 19th century not much has changed for Roma. As all the Serbian population also Roma suffered during the First World War. However, in the Second World War Roma were proclaimed an inferior race by the Nazis and a large number of them got killed.[7] Anecdotal evidence suggests that after the Second World War genocide, Roma have not declared themselves as Roma for decades.

During the communist regime after the Second World War, there were some assimilation pressures on Roma communities which restricted their nomadism and included as well compulsory schooling (Ivanov et al., 2006). Despite evident efforts, the Roma minority has remained largely poor, uneducated and marginalised in Serbia.

The series of Yugoslav wars in the 90s has not benefited anyone. The Roma were mostly affected by the Kosovo war in 1999, when a large number of them found a refuge in Serbia. A lower bound for the number of internally displaced Roma from Kosovo was 27,419 in 2002 according to the United Nations High Commissioner for Refugees. Roma organisations estimate this number to be 65,000 (Ivanov et al., 2006).

3.2 Present Situation of Roma in Serbia

The 2002 Census puts forward a number of 108,000 of Roma or 1.44% of the total population. Estimates, on the other hand, suggest a number between 350,000 and 500,000 or approximately 4-6% of the overall population(Open Society Institute, 2007). The Living Standard Measurement Survey (LSMS) from 2003 provides rich information on the living conditions of the Roma population. It is important to note that this survey includes only Roma living in segregated settlements, which is according to the 2002 Census, the case for 83% of the Roma population.

The numbers from the LSMS survey are alarming. Two out of three Roma households were poor implying that their average consumption was below the

[7]Hancock (1989) provides a chronology of the genocide of Roma.

poverty line.[8] Almost half of the Roma population (42%) is younger than 18 years. Only 56% of children from Roma settlements aged 7 to 14 attend school.[9] Among the adults, 25% have no schooling and another 36% have not finished primary school. The employment rate among the males is very similar to the non Roma population (70%), but the female employment rate is with 35% very low. The LSMS data confirms that Roma live in difficult conditions and that they constitute a marginalised minority.

3.2.1 Where are Roma located in Serbia?

Figure 1 shows in which Serbian municipalities, according to the 2002 Census, Roma are located and it appears that they are scattered all around Serbia with the exception of some municipalities in central Serbia. Many Roma can be found in the South East, but some of the municipalities with a relatively high percentage of Roma are also located in the Northern part and in the municipalities around Belgrade.

Bašić and Jakšić (2005) have compiled a dataset of all Roma settlements in Serbia with more than 100 inhabitants. They recorded 593 such settlements and their distribution is shown in Figure 2. This figure shows that most of the settlements are in fact located around towns and cities and in particular around Belgrade. Note that from this figure we can only see where the settlements are located, it does not take into account the size of the settlements.

4 Data and Summary Statistics

The main source is the Living Standard Measurement Survey (LSMS) which has been conducted in 2003 and has a boosted sample of 525 Roma households living in Roma settlements. The sample contains only households living in Roma

[8]The percentage of the extremely poor among the interviewed Roma was 11.9%.

[9]The percentage of children who attend school at the age of 7 to 10 is 65% (of whom 8% attend special schools for mentally handicapped children), whereas at the age of 11 to 14 this percentage is 46%.

Figure 1: Distribution of Roma settlements in Serbian municipalities

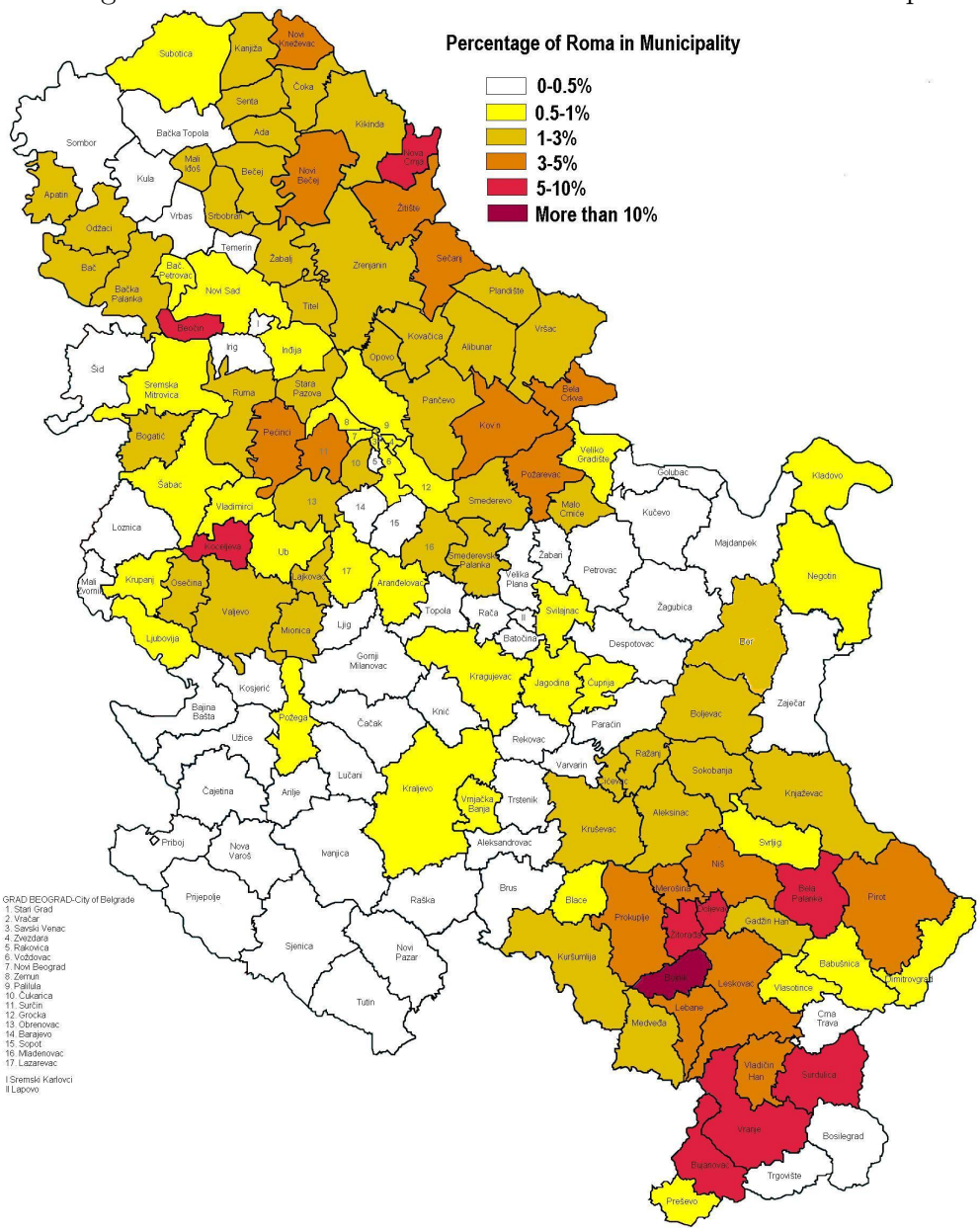
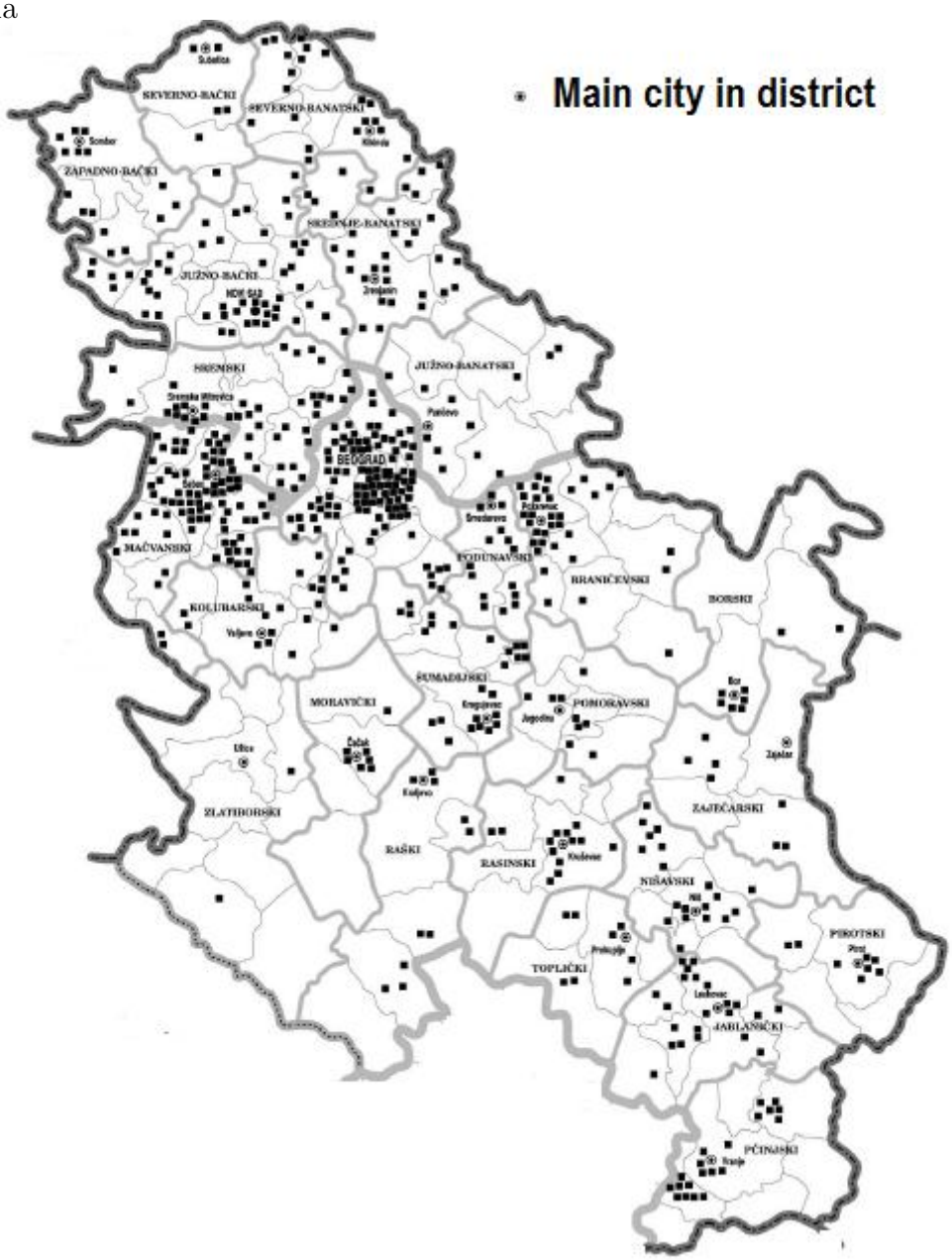


Figure 2: Distribution of Roma settlements with more than 100 inhabitants in Serbia



settlements and is not representative of the Roma population as a whole, but rather of the population living in segregated settlements. The precise data on the number of Roma in Serbia is not available and therefore as the basis of the sampling the 2002 Census has been used. The census tracts for the sample were chosen among all census tracts which had at least 7% of self-declared Roma according to the 2002 Census. In addition, the sample was supplemented with households from towns and cities. The reason is that census tracts in cities are much larger and if the same rule of at least 7% of Roma living in a census tract was followed, Roma in towns and cities would not have been sampled.

I supplement the LSMS data set with additional data taken from the 1971, 1991 and 2002 Census on the number of Roma and the total population in each locality from the sample. I will use locality and census tract as synonyms, because localities are the smallest territorial unit for which data from the census are available.[10]

Municipalities in Figure 3 which are included in the sample are denoted with a blue circle. We can see that the sample is spread throughout the country and that it covers most municipalities which have Roma.

4.1 Summary statistics

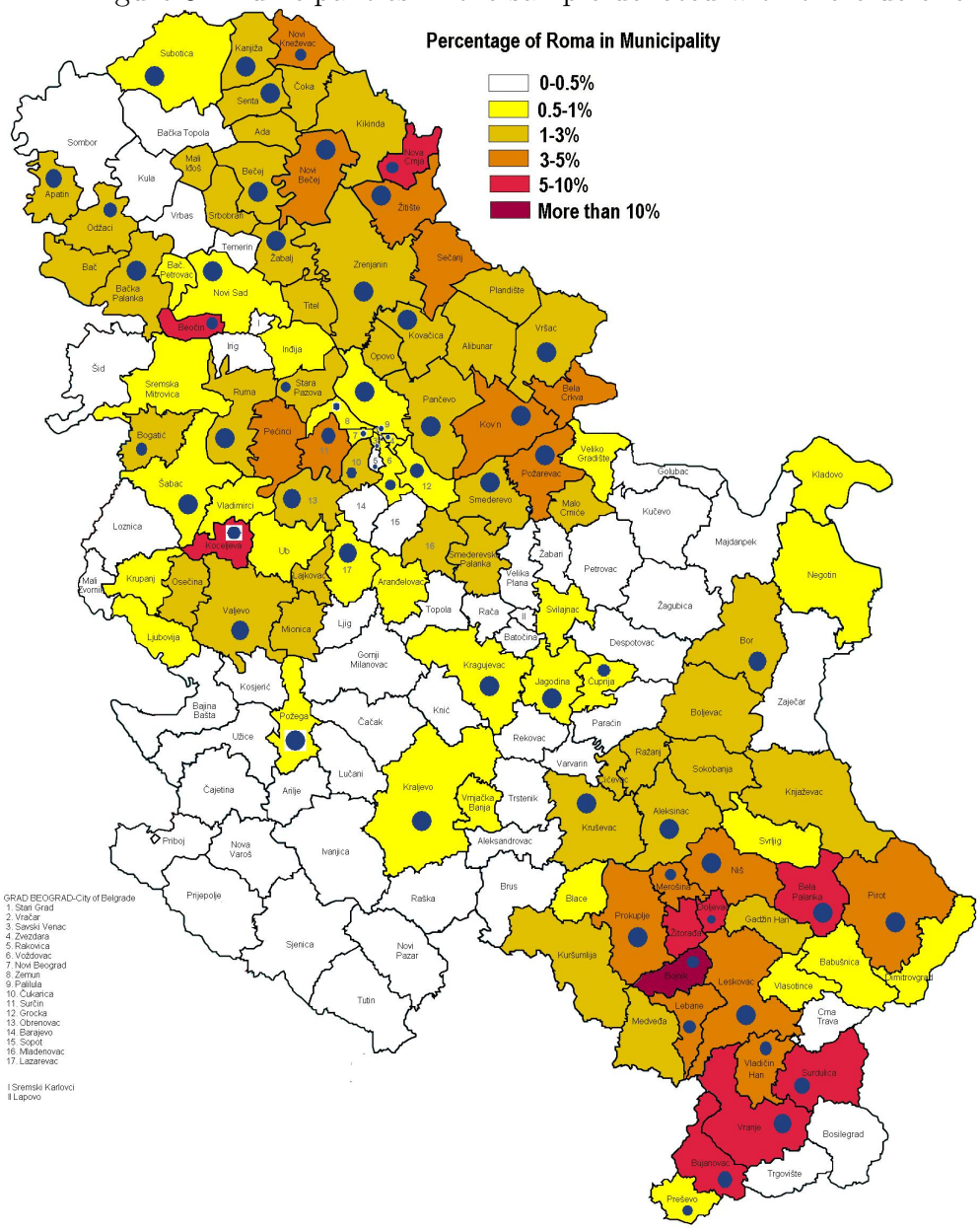
Table 1 summarises characteristics of census tracts in my sample. Total population in a census tract is with 33,000 on average relatively large, but with 8,342 the median is much lower. In the average locality in my sample there are 76.17% Serbs, 8.94% Roma and 14.85% of other ethnic groups.[11]

Table 2 provides descriptive statistics for the outcome and control variables. The main variable of interest is employment. I include in my analysis all workers

[10]Serbia is divided in three levels of administrative division: districts, municipalities and localities. Districts are the highest level of administrative division and there are 24 of them. Below districts are municipalities and their number amounts to 122. The lowest level of administrative division are the census tracts and their number is 4552.

[11]Other ethnic groups recorded in the sample are Montenegrins, Yugoslavs, Albanians, Bosniaks, Bulgarians, Bunjevac, Valachians, Goranac, Hungarians, Macedonians, Muslims, Germans, Rumanians, Russians, Ruthenians, Slovaks, Slovenes, Ukrainians, Croats and Czechs.

Figure 3: Municipalities in the sample denoted with the blue circle



Note: Municipalities in the sample are denoted with a blue circle. The size of the circle does not coincide with the number of households surveyed in the municipality. The placement of the dot is not indicative of where the households were interviewed in that municipality.

Table 1: Summary statistics of census tracts from sample

	Mean	Med.	Std. Dev.	Min.	Max.	No. obs.[a]
Total population	33328	8342	49038	553	217773	84
Serbian population	27953	5302	42674	23	187253	84
Serbian population %	76.17	86.40	24.35	0.75	96.53	84
Roma population	781	395	923	105	4461	84
Roma population %	8.94	4.70	10.67	0.36	48.28	84
Other population	4731	935	10611	1	72568	84
Other population %	14.85	5.14	22.98	0.11	96.00	84

[a] Number of census tracts in sample.

aged 15 to 75 who are not in education. The reason to include both active and inactive workers is that marginal groups, in general, have greater difficulties in finding work and excluding inactive workers, among which are discouraged workers, could lead to a selection bias. 50% of individuals in my sample are employed.

Turning now to control variables, we can see that only a small fraction of individuals speak only Serbian at home. On the other hand, 65% of individuals speak a combination of Romani and Serbian. The households are numerous, on average, all households have 1.6 children younger than 15 and there are 3.67 household members older than 15 years. A very large fraction, 92%, resided in Serbia in 1991, other 7% are refugees from Kosovo. A somewhat higher percentage (61%) of individuals are from urban settings. And lastly, a mean of almost 5 years of education is suggestive of the low education of Roma.

As part of preliminary analysis and with the aim to better understand whether there are significant differences between localities with more and less Roma, I split the sample in two subsamples the bottom 25% and top 25% of residential segregation. The bottom 25% corresponds to individuals living in census tracts with less than 2% and the top 25% are census tracts with more than 11%. I compare the outcome and control variables between the two groups in Table 3.

Table 2: Summary statistics of outcome and control variables

	Mean	Med.	Std. Dev.	Min.	Max.	No. obs.
Outcome variable						
Being employed (=1)	0.50	1.00	0.50	0.00	1.00	1520
Control variables						
Roma population %	8.50	4.50	10.05	0.36	48.28	1520
Wealth[a]	0.23	-0.40	2.56	-4.23	8.66	1490
Serbian spoken at home (=1)	0.09	0.00	0.29	0.00	1.00	1520
Serbian and Romani spoken at home (=1)	0.65	1.00	0.48	0.00	1.00	1520
No. of children under 6	0.78	1.00	0.92	0.00	4.00	1520
No. of children between 6 and 15	0.82	0.00	1.06	0.00	6.00	1520
No. of household members older than 15	3.67	4.00	1.58	1.00	8.00	1520
Age	35.22	32.00	14.23	14.00	74.00	1520
Female (=1)	0.50	1.00	0.50	0.00	1.00	1520
Married (=1)	0.74	1.00	0.44	0.00	1.00	1520
Place of residence in 1991 in Serbia (=1)	0.92	1.00	0.27	0.00	1.00	1520
Urban (=1)	0.61	1.00	0.49	0.00	1.00	1520
Years of education	4.82	4.00	3.56	0.00	16.00	1520
Distance[b]	21.00	15.00	22.00	0.00	21.00	1520

[a] The wealth is calculated using principal component analysis.
[b] Distance from main town in district (in kilometers).

Table 3: Summary statistics for subsample with bottom 25% (1) and top 25% (2) residential segregation

	Bottom 25%[a] (1)	Top 25% (2)	Difference (3) [(1-2)]
Outcome variable			
Being employed (=1)	0.47	0.59	-0.12***
			(0.04)
Control variables			
Wealth	-0.51	0.07	-0.58***
			(0.18)
Serbian spoken at home (=1)	0.13	0.06	0.07***
			(0.02)
Both Serbian and Romani			
spoken at home (=1)	0.61	0.68	-0.07*
			(0.03)
No. of children under 6	1.02	0.48	0.54***
			(0.06)
No. of children between 6 and 15	1.08	0.72	0.36***
			(0.08)
No. of household members			
older than 15	3.46	3.63	-0.17
			(0.11)
Age	33.77	36.35	-2.58**
			(1.02)
Female (=1)	0.51	0.49	0.02
			(0.04)
Married (=1)	0.74	0.74	0.00
			(0.03)
Place of residence in 1991			
in Serbia (=1)	0.85	0.96	-0.11***
			(0.02)
Urban (=1)	1	0.32	0.68***
			(0.03)
Years of education	4.68	4.21	0.47*
			(0.26)
Number of observations	363	410	-47

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

[a] Subsample of individuals living in localities with less than 2% of Roma.

[b] Subsample of individuals living in localities with more than 10.78% of Roma.

Significant differences between individuals living in less and more segregated census tracts are evident and they suggest that sorting into localities is present. Individuals residing in localities with less Roma have, despite their slightly higher level of education, significantly worse labour market outcomes than workers in localities with more Roma. Individuals in less Roma populated localities are wealthier, have significantly more children and are on average 2 years younger.

5 Identification Strategy and Findings

5.1 Identification Strategy

My variable of interest is an indicator variable for being employed and I will comply with the standard in this literature and estimate a probit model. My starting and main regression is:

$$y_{ijk} = \alpha_0 + \alpha_1 RS_{jk} + \alpha_2 X'_{ijk} + \gamma_k + \epsilon_{ijk} \quad (1)$$

where y_{ijk} is employment, my measure of labour market performance, of individual i in census tract j and in district k, RS is residential segregation which I measure as the percentage of Roma in locality j in district k, X'_{ijk} are individual level controls (wealth of the household, dummy for whether only Serbian is spoken at home, dummy for whether the language spoken at home is a combination of Romani and Serbian, number of children younger than 5, number of children younger than 15, number of household members, age, age squared, a dummy for being married, a dummy whether the place of residence before 1991 was in Serbia, a dummy for urban settings, the number of years of education and distance from main city in district) and γ_k are district fixed effects.

Several issues need to be mentioned and discussed about my identification strategy. I am using district level fixed effects and I am exploiting the variation within districts to obtain my estimate of the residential segregation in census tracts. In order to use variation within districts, I need households from at

least two census tracts in my sample to be from the same district because my variable of interest, residential segregation, varies only at the census tract level. For 5 districts I have only one census tract in that district and I cannot use the individuals from these districts in the analysis. For this reason, I exclude these 76 individuals from the sample.[12]

The more important issue is that the coefficient of interest α_1 could be biased due to omitted variables or reverse causality. The ideal experiment for the analysis of residential segregation on labour market outcomes would require a random assignment of Roma households to census tracts.[13] However, this is clearly not the case in my setting. It is reasonable to believe that, when they settled, Roma chose to settle in places where they thought to have good labour market opportunities. As the location decision was based on the labour market opportunities, it is highly likely that the error term ϵ_{ijk} contains unobservable factors which affect both residential segregation and the outcome variable employment. In this case, the coefficient α_1 should be interpreted as a mere correlation between employment and residential segregation and not as a causal effect. I will employ an instrumental variable approach to alleviate the problem of sorting into localities.

Let me rewrite and decompose the error term ϵ_{ijk} from regression (1) and explain more formally the problems of my identification strategy and how I will aim to cope with them:

$$\epsilon_{ijk} = \theta_i + \pi_{jk} + e_{ijk} \tag{2}$$

Consider first the term θ_i that captures individual characteristics which I am not able to control for. If there is some individual characteristic, for instance ability is a well known example, which affects both residential segregation and labour market outcomes, the estimate α_1 will be inconsistent. Assuming that there is negative sorting, meaning that more able and hence more employable individuals

[12]It would certainly be more convincing to use also municipality level fixed effects, but I do not have variation in residential segregation at the municipal level and I cannot include municipality fixed effects in my analysis.

[13]See for instance Edin et al. (2003) or Damm (2009) for this type of approach.

sort into census tracts with less residential segregation, the probit estimate α_1 would underestimate the effect of residential segregation. The second term π_{jk} captures census tract characteristics which could affect both residential segregation and employment characteristics such as infrastructure and public services. The third term e_{ijk} is the usual random term.

To circumvent the problems of omitted variable bias or endogeneity of the standard probit model, I will employ an instrumental variable approach. I will instrument residential segregation in 2002, with residential segregation in 1991. Instrumenting current segregation or other variables with its historical values is not new in the literature. Similar to my study, Altonji and Card (1991) use historical immigration rates to instrument its current values.[14]

My first stage regression is given by:

$$RS_{jk} = \beta_0 + \beta_1 RS1991_{jk} + \beta_2 X'_{ijk} + \gamma_k + u_{ijk} \tag{3}$$

In the first stage regression, I estimate residential segregation in 2002 RS_{jk} with residential segregation $RS1991_{jk}$ in 1991. I include the same control variables as in regressions (1).

My conditional exogeneity condition requires the instrument $RS1991_{jk}$ to be uncorrelated with the error term from the second stage regression 1. In words, there should not be any unobservable characteristics of the individuals or localities which were already present in 1991 and which affect labour market outcomes in 2002.

My identification strategy would thus be invalidated if some individuals sorted into localities before 1991. In order to test whether this is the case and whether my results are driven by sorting before 1991, I will restrict my sample to individuals younger than 25 and individuals younger than 30. The restricted sample of younger individuals together with my identification strategy should allow me to exclude the possibility of sorting into localities before 1991. The individuals of

[14] Another example is Woodruff and Zenteno (2007) who use historical migration rates to instrument for the present connection to the migration network.

the age 25 (or 30) in 2003 were only 13 (or 18) in 1991 and it is unlikely that they have already sorted by 1991 into certain localities. Moreover, to ascertain the sensitivity of my results, I use segregation in 1971 as an alternative instrument for segregation in 2002.

The second problem is that there could be some census tract characteristics also in 1991 which affect both residential segregation and employment opportunities. I will test for this possibility by including a set of control variables which capture the community infrastructure and community facilities.

5.2 Findings

Table 4 provides results for the main specification. Column (1) and column (2) provide the results for non-instrumented and instrumented regressions. In Table 5 I split the sample between males and females. Columns (1) and (2) show the results for male individuals and columns (3) and (4) for female individuals. All regressions are estimated with the probit model.

Let me first discuss how the coefficient of the variable of interest, residential segregation changes among different specification. I find that residential segregation has, on average, a positive impact on employment. This effect is not significant for the sample as a whole. However, when I split the sample between females and males I find a positive impact of residential segregation on labour market outcomes for women. I find that an increase in residential segregation by 10 percentage points increases the probability of a female individual to be employed by 7 percentage points.

Now I discuss the signs and significance of control variables. The omitted categories of language spoken at home is either only Romani spoken at home or a combination of Romani and another language spoken at home. I would have expected that Serbian spoken at home (language 1) or a combination of Serbian and Romani spoken at home (language 2) has a positive effect on employment. These two variables should capture the fact that individuals speak Serbian better and/or that they are better integrated in the society than the ones who use other

Table 4: Outcome: Being employed (=1)

Model	Probit	IV
	(1)	(2)
Residential segregation	0.006	0.010
	(0.006)	(0.007)
Wealth	-0.024	-0.020
	(0.018)	(0.019)
Serbian spoken at home (=1)	-0.171	-0.151
	(0.163)	(0.164)
Serbian and Romani spoken at home (=1)	-0.098	-0.081
	(0.096)	(0.096)
No. of children under 6	-0.066	-0.068
	(0.045)	(0.046)
No. of children between 6 and 15	0.022	0.009
	(0.037)	(0.038)
No. of adult members older than 15	-0.062**	-0.056**
	(0.025)	(0.025)
Age	0.090***	0.090***
	(0.017)	(0.017)
Age squared	-0.001***	-0.001***
	(0.000)	(0.000)
Female (=1)	-0.971***	-0.957***
	(0.076)	(0.076)
Married (=1)	0.248**	0.262***
	(0.101)	(0.101)
Place of residence in 1991 (=1)	-0.015	-0.019
	(0.149)	(0.149)
Urban (=1)	-0.130	-0.143
	(0.110)	(0.115)
Years of education	0.032***	0.032***
	(0.012)	(0.012)
Distance to main city in district	0.007***	0.006**
	(0.002)	(0.002)
District fixed effects	Yes	Yes
N	1414	1393
First stage F-statistic		342

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

Table 5: Outcome: Being employed (=1) - Sample split between women and men

Model	Probit	IV	Probit	IV
Gender	Male	Male	Female	Female
	(1)	(2)	(3)	(4)
Residential segregation	-0.003	-0.005	0.015*	0.022**
	(0.008)	(0.010)	(0.008)	(0.009)
Wealth	-0.027	-0.019	-0.011	-0.009
	(0.028)	(0.028)	(0.026)	(0.027)
Serbian spoken at home (=1)	-0.002	0.005	-0.302	-0.283
	(0.228)	(0.228)	(0.242)	(0.243)
Serbian and Romani spoken at home (=1)	0.079	0.100	-0.316**	-0.305**
	(0.143)	(0.143)	(0.132)	(0.133)
No. of children under 6	-0.030	-0.036	-0.120*	-0.125*
	(0.072)	(0.072)	(0.064)	(0.065)
No. of children between 6 and 15	0.061	0.051	-0.029	-0.047
	(0.056)	(0.057)	(0.054)	(0.055)
No. of adult members older than 15	-0.021	-0.020	-0.121***	-0.111***
	(0.037)	(0.037)	(0.039)	(0.039)
Age	0.065***	0.064***	0.098***	0.097***
	(0.023)	(0.024)	(0.024)	(0.024)
Age squared	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Married (=1)	0.624***	0.654***	-0.099	-0.106
	(0.150)	(0.152)	(0.135)	(0.136)
Place of residence in 1991 (=1)	-0.077	-0.099	0.031	0.047
	(0.233)	(0.232)	(0.207)	(0.208)
Urban (=1)	-0.042	-0.064	-0.300*	-0.315*
	(0.162)	(0.167)	(0.160)	(0.167)
Years of education	0.046**	0.045**	0.023	0.023
	(0.018)	(0.018)	(0.018)	(0.018)
Distance to main city in district	0.008**	0.008**	0.005	0.004
	(0.004)	(0.004)	(0.003)	(0.003)
District fixed effects	Yes	Yes	Yes	Yes
Number of observations	705	696	709	697
First stage F-statistic		190		162

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

languages in their households. The language spoken at home does not seem to be relevant. The number of adult members in a household, especially for female members, appears to have a negative impact on employment. The probability to be employed increases with age. Somewhat surprising is that Roma from urban localities are less likely to be employed. Education does not seem to significantly affect employment. The reason for this could be that for such low levels of education there are no evident returns. Distance from the main city in the district seems to play a relevant role in all specification. The further away the census tract from the main town in the district, the more likely it is that the individual is employed.

By instrumenting residential segregation in 2002 with residential segregation in 1991, I eliminate the effect of sorting into locations from 1991 to 2002. The fact that the estimate from the instrumented regression is higher than the non-instrumented estimate is suggestive evidence that less employable workers sort into areas with a higher percentage of Roma.

My instrumental variable approach is robust to sorting after 1991, but my results could still be driven by sorting before 1991. In order to understand whether this is the case, I restrict my sample to individuals younger than 25 and individuals younger than 30. I report the results in Table 6 in panels B and C, the results for the whole sample are reported again in panel A. Restricting the sample to younger employees reduces the sample and it increases the standard errors. However, I obtain again significant results of a very similar magnitude for most specifications. The probability of being employed for female workers is still statistically significant. The coefficient is somewhat larger for individuals at most 25 years old in the IV specification (panel B). The reason for the larger coefficient in panel B could be either that the coefficient is imprecisely measured or it could be that being in a more Roma inhabited census tract is even more beneficial for younger Roma women.

My identification strategy relies on the assumption that there are no unobservable characteristics of the localities which were already present in 1991 and

which affect labour market outcomes in 2002. I do not have controls for locality characteristics and to explore the possibility that there are omitted variables I add to my main regression additional control variables which are taken from the survey. I add control variables which capture community infrastructure characteristics and community facilities. To measure community infrastructure I add the following dummy variables (at the community level): electricity network, water supply, sewage, gas pipeline, central heating, telephone network, waste dustbins and asphalt paved streets. I also have information on the presence of some community facilities (children playground, park, sports field, kindergarten, communal premises and food stores). I provide the summary statistics for these variables in the appendix in Table 10. These additional variables are clearly not directly related to employment, but they capture the quality of the neighbourhood and they could be correlated with other locality characteristics which affect employment. I add these control variables to my main regression 1 and I replicate the results for both non-instrumented and instrumented models. The results are reported in panel D in Table 6. With the additional controls my non-instrumented estimates become somewhat smaller, but the IV estimates retain their significance and magnitude.

In order to confirm my results, I use the 1971 census data to instrument for residential segregation in 2002. In general, the older the instrument the more certain one can be that the results are not driven by sorting into census tracts. Table 7 summarises the results. The IV estimate for female workers in column (6) is somewhat smaller than the original IV estimate using the 1991 data, however, the significance remains.

The instrumental variable strategy suggests that *more* employable individuals sort into *less* segregated census tracts. On the other hand, individuals in *more* segregated census tracts are *more* likely to have employment. These two findings together point to the fact that the relationship between employment and residential segregation might be non linear. To understand better the relationship between residential segregation and employment, as suggested by Boeri

Table 6: Outcome: Being employed - Robustness check

Model	Probit	IV	Probit	IV	Probit	IV
Gender	All	All	Male	Male	Female	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Whole sample						
Res. seg.	0.006	0.010	-0.003	-0.005	0.015*	0.022**
	(0.006)	(0.007)	(0.008)	(0.010)	(0.008)	(0.009)
No. obs.	1414	1393	705	696	709	697
Panel B: Individuals at most 25 years old						
Res. seg.	0.013	0.017	0.008	-0.010	0.017	0.038**
	(0.010)	(0.012)	(0.016)	(0.019)	(0.013)	(0.016)
No. obs.	429	419	191	187	202	196
Panel C: Individuals at most 30 years old						
Res. seg.	0.012	0.015	0.012	0.000	0.012	0.025*
	(0.008)	(0.010)	(0.013)	(0.016)	(0.011)	(0.013)
No. obs.	647	637	305	301	317	311
Panel D: Whole sample with additional controls for infrastructure and communal facilities[a]						
Res. seg.	0.003	0.010	-0.006	-0.003	0.010	0.019*
	(0.006)	(0.008)	(0.010)	(0.012)	(0.009)	(0.011)
No. obs.	1414	1393	704	695	708	696

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

Controls included are wealth, Serbian spoken at home (=1), Serbian and Romani spoken at home (=1), number of children under 6, number of children between 6 and 15, number of adult members older than 15, age, age sqaured, Female (=1), married (=1), place of residence in 1991 (=1), urban (=1), years of education and distance to main town in district.

[a] Additional controls: Electricity network, water supply, sewage, gas pipeline, central heating, telephone network, waste, dustbins, asphalt paved streets, children playground, park, sports field, kindergarten, communal premises and food stores. All controls are dummies at the community level.

et al. (2011), I first introduce a quadratic term in regression 1. Then, I create 4 dummies for different levels of residential segregation. The dummies capture whether there are more than 2% (25th percentile), 5% (60th percentile), 10% (73rd percentile) and 15% (83rd percentile) of Roma in the census tract.

Table 7: Outcome: Being employed - Robustness check: Alternative IV: Percentage of Roma in census tract in 1971

Model	Probit	IV	Probit	IV	Probit	IV
Gender	All	All	Male	Male	Female	Female
	(1)	(2)	(3)	(4)	(5)	(6)
IV: Percentage of Roma in census tract in 1971						
Res. seg.	0.006	0.005	-0.003	-0.011	0.015*	0.018**
	(0.006)	(0.007)	(0.008)	(0.010)	(0.008)	(0.009)
No. obs.	1414	1393	705	696	709	697

Robust standard errors in parenthesis: * significant at 10%, ** significant at 5%, *** significant at 1%..
Controls included are wealth, Serbian spoken at home (=1), Serbian and Romani spoken at home (=1), number of children under 6, number of children between 6 and 15, number of adult members older than 15, age, age sqaured, Female (=1), married (=1), place of residence in 1991 (=1), urban (=1), years of education and distance to main town in district.

The coefficients of the dummies are increasing in the case of the probit model in Panel A in Table 8 and that suggests that it becomes more convenient in terms of likelihood of being employed to live in a more segregated locality when the percentage of Roma is 10% or more. When I instrument residential segregation in 2002, I obtain larger point estimates for 2% and 5% and smaller estimates for 10% and 15% than in the non instrumented model (results shown in lower panel B of Table 8). The IV model has very imprecise point estimates and it is somewhat difficult to reconcile the non instrumented and instrumented probit models.

Table 8: Outcome: Being employed - non linearities - %Roma in census tract

	(1)	(2)	(3)	(4)	(5)
Panel A: Probit					
Residential segregation	0.041***				
	(0.015)				
Residential segregation sq.	-0.001**				
	(0.000)				
Residential segregation >2		-0.026			
		(0.133)			
Residential segregation >5			0.052		
			(0.115)		
Residential segregation >10				0.323***	
				(0.114)	
Residential segregation >15					0.251**
					(0.125)
Panel B: Instrumental variable probit					
Residential segregation >2		0.293			
		(0.215)			
Residential segregation >5			0.377*		
			(0.197)		
Residential segregation >10				0.173	
				(0.264)	
Residential segregation >15					-0.192
					(0.301)
Number of observations	1414	1414	1414	1414	

Robust standard errors in parenthesis: * significant at 10%, ** significant at 5%, *** significant at 1%..
Controls included are wealth, Serbian spoken at home (=1), Serbian and Romani spoken at home (=1), number of children under 6, number of children between 6 and 15, number of adult members older than 15, age, age sqaured, Female (=1), married (=1), place of residence in 1991 (=1), urban (=1), years of education and distance to main town in district.

5.3 Additional Robustness Checks

5.3.1 Restricting the Sample

I will check the robustness of my results first by excluding all refugees and then by excluding all individuals from the capital city from my sample. Results from my instrumental variable estimation suggest that less employable individuals sort into localities with more Roma, this leads to a downward bias of my probit estimates. There are 7.05% refugees from Kosovo in the sample who have highly likely come to Serbia in 1999. I exclude these individuals from the sample because I would like to understand how much of the sorting into localities with less Roma is driven only by internal migrants. Panel B in Table 9 shows the results for the sample without Kosovo. When I exclude refugees from Kosovo, the point estimates fall, but they retain their significance. This points to the fact that more employable refugees have sorted into more segregated census tracts.

Serbia is a very centralised country and most of the migration is occurring towards its capital city Belgrade. This fact is also reflected in the sample, 21% of the sample are individuals from Belgrade. To see how much of my results are driven by the dynamics of Belgrade, I exclude all the individuals from Belgrade and I estimate both the non-instrumented and instrumented probit model and the results are summarised in Panel C of Table 9. It appears that Belgrade and the rest of Serbia have a similar dynamic. Individuals who live in census tracts with more Roma are more likely to find a job, this effect is especially strong for female workers.

5.4 Discussion of Results

The first key finding of my study is that individuals living in more segregated census tracts are more likely to have a job. This effect is fully driven by female workers. The theory suggests a series of channels why residential segregation could undermine labour market opportunities. On the other hand, there are only two explanations why segregation could be beneficial. The first explanation

Table 9: Outcome: Being employed - Additional robustness check

Model	Probit	IV	Probit	IV	Probit	IV
Gender	All	All	Male	Male	Female	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Whole sample						
Res. seg.	0.006	0.010	-0.003	-0.005	0.015*	0.022**
	(0.006)	(0.007)	(0.008)	(0.010)	(0.008)	(0.009)
No. obs.	1414	1393	705	696	709	697
Panel B: Sample without refugees from Kosovo						
Res. seg.	0.005	0.008	-0.001	-0.004	0.011	0.017*
	(0.006)	(0.007)	(0.008)	(0.010)	(0.008)	(0.010)
No. obs.	1316	1295	659	650	657	645
Panel C: Sample without capital city						
Res. seg.	0.006	0.009	-0.004	-0.008	0.016**	0.021**
	(0.006)	(0.007)	(0.008)	(0.010)	(0.008)	(0.009)
No. obs.	1092	1071	548	539	544	532

Robust standard errors corrected for clustering at district level in parenthesis: * significant at 10%, ** significant at 5%, *** significant at 1%..
Controls included are wealth, Serbian spoken at home (=1), Serbian and Romani spoken at home (=1), number of children under 6, number of children between 6 and 15, number of adult members older than 15, age, age sqaured, Female (=1), married (=1), place of residence in 1991 (=1), urban (=1), years of education and distance from main town in district.

[a] Additional controls: Electricity network, water supply, sewage, gas pipeline, central heating, telephone network, waste, dustbins, asphalt paved streets, children playground, park, sports field, kindergarten, communal premises and food stores. All controls are dummies at the community level.

would be that there are firms in enclaves, in such firms workers from enclaves could even be preferred because its customers would be from the same ethnic group. Knowing the Serbian context and which type of jobs Roma perform, this is unlikely to be the case. The second possible explanation suggests that there is an information channel at place. In his seminal work Munshi (2003) has found that the size of networks played a very important role in finding a job and earning a higher wage among Mexican immigrants in the US. These results have been confirmed by other studies such as Bayer et al. (2008). It is important

to note that the aforementioned studies use data at a more disaggregated level, however, it is still possible that my more aggregated data and results are driven by local network effects. In particular, in the Serbian context where the majority of jobs is found through networks , the percentage of ones ethnic group in the census tract could be of great relevance for finding a job.

I find only significant results for women and this is certainly puzzling. It is possible that women are engaged in different types of jobs from men where referrals are more important. The question whether the current job was found through a contact was not asked in the questionnaire, and I can only speculate on the importance of referrals for female workers.

The best I can do to understand why women in more segregated census tracts are more likely to have a job is to examine the characteristics of jobs in which women are engaged. I find that a large share of both male and female workers is engaged in seasonal jobs, however, this share is with 46% 10 percentage points larger for women. With this in line is also the fact that 44% of female workers have been employed at their current job for less than a year. For male workers, this number amounts to 34%. The data suggest further that women are 13 percentage points more likely than men to be employed in the agricultural sector. These findings taken together provide some evidence that a large share of women is engaged in seasonal jobs and in particular these jobs seem to be in the agricultural sector. I suspect that the information channel is more important for individuals who work in the agricultural sector as seasonal workers.

The second key finding is that less employable individuals choose to move to more segregated localities.

A weakness of this study is that data is not available at the block level as in Bayer et al. (2008) and Boeri et al. (2011). Disaggregated data would allow me to better examine the interaction at the neighbourhood level. By using census tract data, however, I am capturing the broader consequences of living in a segregated locality than the local neighbourhood interactions.

6 Conclusion

Roma are a marginalised and socially excluded ethnic group which lives mainly in segregated settlements. In order to shape necessary policies for improving their living conditions, it is important to understand whether and how their spatial separation has impacts on economics outcomes. This paper has examined the effect on one very important economic outcome which is the probability of being employed.

In this paper I have provided empirical evidence that women in more segregated census tracts are more likely to be employed than women living in less segregated areas. This effect is not negligible: a one standard deviation or equivalently 10 percentage points increase in residential segregation, increases the probability of being employed by 6 to 7 percentage points. My results are robust to different restriction of the sample as well as to the addition of more control variables. The second finding of this study is that there is negative sorting - more employable workers sort into less Roma populated census tracts, while less employable Roma move to more segregated areas.

I suspect that the positive effect of segregation is driven by the information channel. That is, the higher employment of Roma women living in more segregated environments is occurring due to better and more information provided to them by their co-ethnics.

In terms of policy implications, my results suggest that dispersal policies can be very harmful to Roma as their employment opportunities could depend on the size of their co-ethnics network. In the recent years several settlements have been closed down in Belgrade and their inhabitants have been dispersed in different location in Belgrade. Such policies, if necessary, would have to be accompanied with other policies which would offset the negative spillovers.

References

Alesina, A., R. Baqir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Altonji, J. G. and D. Card (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, Trade and the Labor Market*, NBER Chapters, pp. 201–234. National Bureau of Economic Research, Inc.

Bašić, G. and B. Jakšić (2005). *Umetnost preživljavanja*. Institut za filozofiju i društvenu teoriju.

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*(6), 1150–1196.

Becker, G. S. (1971). *The Economics of Discrimination*. University of Chicago Press.

Bertrand, M., E. F. P. Luttmer, and S. Mullainathan (2003). Network Effects and Welfare Cultures. *The Quarterly Journal of Economics 118*(1), 329–357.

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.

Borjas, G. J. (1992). Ethnic Capital and Intergenerational Mobility. *Quarterly Journal of Economics 107*(1), 123–150.

Borjas, G. J. (1994). Long-Run convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration. *Industrial and Labor relations Review 47*(4), 553–573.

Borjas, G. J. (1995). Ethnicity, neighbourhoods and human-capital externalities. *American Economic Review 85*(3), 365–390.

Card, D. and J. Rothstein (2007). Racial segregation and the black-white test score gap,. *Journal of Public Economics 91*(11–12), 2158–2184.

Case, A. C. and L. F. Katz (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. NBER Working Papers 3705, National Bureau of Economic Research, Inc.

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 7*(5), 603–628.

Damm, A. P. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labour Economics 27*(2), 281–314.

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.

Evans, W. N., W. E. Oates, and R. M. Schwab (1992). Measuring Peer Group Effects: A Study of Teenage Behavior. *The Journal of Political Economy 100*(5), 966–991.

Hancock, I. (1989). Gypsy history in germany and neighboring lands: A cronology leading to the holocaust and beyond. In D. Crowe and J. Kolsti (Eds.), *The Gypsies of Eastern Europe*, pp. 11–30. Armonk: E.C. Sharpe.

Holzer, H. J. and K. R. Ihlanfeldt (1998). Customer Discrimination and Employement Outcomes of Minority Workers. *The Quarterly Journal of Economics 118*(1), 329–357.

Ivanov, A., M. Collins, C. Grosu, J. Kling, S. Milcher, N. O'Higgins, B. Slay, and A. Zhelyazkova (2006). *At Risk: Roma and the Displaced in Southeast Europe.*

United Nations Development Programme Regional Bureau for Europe and the Commonwealth of Independent States.

Kain, J. (1968). Housing Segregation, Negro Employment and Metropolitan Decentralization. *The Quarterly Journal of Economics 82*(2), 175–197.

Katz, L., J. R. Kling, B. J. Liebman, and L. Sanbonmatsu (2004). Moving to opportunity and tranquility: Neighborhood effects on adult economic self-sufficiency and health from a randomized housing voucher experiment. Working Papers 860, Princeton University, Department of Economics, Industrial Relations Section.

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.

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

Woodruff, C. and R. Zenteno (2007). Migration networks and microenterprises in mexico. *Journal of Development Economics 82*(2), 509–528.

A Tables

Table 10: Summary statistics of additional controls

	Mean	Std. Dev.	Min.	Max.	No. obs.
Community infrastructure characteristics					
Electricity network (=1)	0.91	0.29	0.00	1.00	1520
Water supply (=1)	0.76	0.42	0.00	1.00	1520
Sewage (=1)	0.42	0.49	0.00	1.00	1520
Gas pipeline (=1)	0.05	0.22	0.00	1.00	1520
Central heating (=1)	0.00	0.04	0.00	1.00	1520
Telephone network (=1)	0.51	0.50	0.00	1.00	1520
Waste dustbins (=1)	0.32	0.47	0.00	1.00	1520
Asphalt paved streets (=1)	0.45	0.50	0.00	1.00	1520
Community facilities					
Children playground(=1)	0.07	0.25	0.00	1.00	1520
Park (=1)	0.08	0.27	0.00	1.00	1520
Sports fields (=1)	0.09	0.29	0.00	1.00	1520
Kindergarten (=1)	0.19	0.39	0.00	1.00	1520
Communal premises/ cultural clubs (=1)	0.14	0.34	0.00	1.00	1520
Food and grocery stores (=1)	0.58	0.49	0.00	1.00	1520