Intersectionality of ethnicity and gender: exploring Romani women's performance in education

Ewa Cukrowska-Torzewska*

Abstract:

This paper examines Romani women multiple disadvantage in education. Three main educational outcomes are analyzed: obtaining secondary or higher education, dropping out from the school and having an ability to read and write. The empirical analysis is based on the data coming from two unique datasets on Roma carried out by the UNDP, i.e. 2004 UNDP Regional Roma Survey and 2011 UNDP/WB/EC regional survey on Roma communities. The results show that variables measuring family background are the key observable factors that explain ethnic gaps in education. On contrary, the gender gaps in education - both among Roma and non-Roma – cannot be explained by the differences in men's and women's characteristics and may be attributed to other unobservable factors such as cultural customs and values. The comparison of the existing gaps over 2004-2011 reveals that some progress in terms of educational performance of Roma has been achieved. Similar patterns are however found for Romani women and men, leaving Romani women still in a more disadvantaged position.

Keywords: gender gap, ethnic gap, education, matching, decomposition, difference-in-difference estimation.
JEL Code: I24, J15, J16.

* Faculty of Economic Sciences, University of Warsaw

Introduction

This paper documents gender and Roma gaps in educational outcomes in the countries of South-Eastern Europe, in which the share of Romani population is significant. The aim of this paper is twofold. Firstly, it seeks to address intersectional status of Romani women and reveal their cumulative disadvantage in education that is due both to ethnicity and gender. Secondly, the overall goal of this research is to initiate a discussion on the need of designing gender sensitive policy intervention aimed at Roma.

The Roma constitute the largest and the most vulnerable ethnic minority in the Central and South Eastern Europe. Ivanov et al. (2006) recognize that Roma's vulnerability is a complex problem consisting of several interrelated aspects, such as education, employment, health and poverty. Recent research shows that despite significant measures to ensure equal access to quality education of vulnerable groups, Roma children still perform in education significantly worse than their non-Roma peers. Kertesi and Kézdi (2011) report that Roma-non-Roma test score gap in Hungary is approximately one standard deviation and it is similar to the gap between African-American and White students in the U.S. in the 1980s. They also find that parenting, family background characteristics, health and school fixed effects are critical factors for explaining the existence of the gap. Recently, Brüggemann (2012) also provided a comprehensive picture of educational performance of Roma by examining Roma-non-Roma gaps in several educational outcomes, such as school attendance, overall and computer literacy or dropout rates.

In this paper previous analyses of Roma performance in education are extended by focusing on the intersectional status of Romani women and examining their multiple disadvantage that is caused by both ethnicity and gender. Gender inequality in the access to education is recognized as one of the main factors curbing economic development and growth (Brummet, 2008). Unequal access to education and differences in educational performance of girls and boys are also found to have long-term implications for the number of development related areas such as labor market participation, family income, early marriages, fertility rates and health awareness (Schultz, 1993, UNESCO, 2010). Data on educational outcomes by gender reveal that in many countries of Central and South-Eastern Europe the percentage of men with at least secondary education still outperforms women.[1] Country-specific analyses of gender gaps in education also show that women are less educated than men and perform in school relatively worse (e.g. Davalos, 2012). Moreover, Romani girls are subject to a higher risk of lower education-attainment due to a higher inference of early marriage and childbirth that make

[1] UNDP (2013). Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World.

them leave the schooling system before its successful completion (Cukrowska and Kóczé, 2013; Ilisei, 2013).[2]

In this paper the analysis of gender and Roma intertwining features in education is divided into three parts that coincide with the specific goals of the research. In particular, these goals include: 1) the identification of the size of the gender and ethnic gaps in educational outcomes; 2) the identification of observable determinants of the gender and ethic gaps in educational outcomes and the degree to which they can explain the existence of the gaps; 3) the identification of the changes in the gaps over the years 2004-2011. The analysis focuses on three major educational outcomes: the probability of obtaining at least secondary education, dropping out from the school and an ability to read and write (literacy rate).

The formal analysis is based on the data coming from the UNDP data collections on Roma community. More specifically, two comparable datasets on Roma are used: one from 2004 and a second from 2011. More attention is devoted to the analysis of the situation in recent period with the use of the 2011 United Nations Development Programme (UNDP), World Bank (WB) and European Commission (EC) regional survey on Roma communities, while the 2004 UNDP Regional Roma Survey is used only to assess whether any changes occurred over the 2004-2011 time framework.

In general, the findings show that Romani women suffer from lower educational attainment, spend fewer years in education and have higher dropout and lower literacy rates – both in comparison to non-Romani women (higher disparities) and Romani men. Family background characteristics are the key factors that explain the ethnic gap in education, but are not enough to explain the gap that persist between males and females. The gender gap in education is thus most likely caused by other structural factors that are not accounted for and that may include social customs and values or perceived returns to women's education. The comparison of the gaps over 2004-2011 shows that changes that appeared over that time were mostly positive, but comparable progress has been achieved by Romani men and Romani women, so that Romani women are still left in a more disadvantaged situation.

The remainder of this paper is divided into four major sections. The next section describes data used in the empirical research. Section two focuses on the methodology that is adapted to identify drivers of Romani women disadvantage in education as well as changes that occurred with that regard over 2004-2011 time framework. Section three presents and carefully discusses main empirical results. Finally, section four gives concluding remarks.

[2] Detailed analysis of gender inequality among Roma in various aspects of human and economic development has been recently provided by Cukrowska and Kóczé (2013).

1. Data

The paper uses two data collections that were carried out primarily by the UNDP: 2011 UNDP/WB/EC regional survey on Roma communities and 2004 UNDP Regional Roma Survey. The data collection from 2004 is the first dataset on Roma that in addition to information on Romani population also includes the sample of non-Roma living in a close proximity to Roma. Similarly, the 2011 UNDP/WB/EC regional survey on Roma communities also accounts for both Roma and non-Roma living in a close proximity. The advantage of these datasets is that such definition of sampling design allows researchers not only to analyze the situation of Roma but also to compare Roma with non-Roma in terms of their social status, living conditions, education, employment migration patterns and poverty. The fact that the surveys comprise of Roma and non-Roma living in a close proximity to Roma may however have also some impact on the validity of the results. More specifically, non-Roma living in a close proximity to Roma may not be generally representative for all the non-Roma population. In consequence, the ethnic gaps based on the collected data may be lower than actually prevailing ones. On the other hand, the survey was conducted among the households, which were situated in the most disadvantaged and segregated Roma settlements or in the areas with compact Roma populations. As a consequence, those Roma who are living in integrated communities were not considered in the survey, so that the derived ethnic gaps may be overestimated. When interpreting the results it has to be thus clearly stated that the results cannot be generalized to the whole population and the estimated ethnic gaps represent the gaps between non-Roma and Roma living in a close proximity.

The 2004 data collection consists of data for nine countries of Southern Eastern Europe, in which the share of Roma population is sizeable: Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Kosovo, Macedonia, Montenegro, Romania and Serbia. In 2011 the survey was additionally extended to Central Eastern countries: Slovakia, Czech Republic, Hungary as well as the Republic of Moldova. From the countries of 2004 survey only Kosovo was not surveyed in 2011.[3] Within the 2004 survey a total of 8,273 households and 34,116 individuals were interviewed. In 2011 the scope of the survey was extended and it comprised information on 9,207 Roma households with a total of 41,334 individuals being interviewed and 4,274 non-Roma households with a total of 13,326 individuals.

To analyze changes over time, two datasets were combined to the extent that they remain consistent. Some of the definitions of the variables however differ in these two surveys and in consequence not all the variables could be included in the

[3] For more information on the sample design and survey implementation see Ivanov at el. (2012).

merged dataset. Moreover, as mentioned above not all the countries were covered by both the surveys and the coherent dataset is limited to the following countries: Albania, Bulgaria, Bosnia and Herzegovina, Croatia, Macedonia, Montenegro, Serbia, and Romania.

Simultaneous use of both surveys certainly allows for drawing conclusion on the changes in the relative status of Roma over time. However, while interpreting the results one needs to bear in mind that the survey's design is not longitudinal. This means that the individuals interviewed in each survey are not necessarily the same. As a result, while performing a comparative analysis, time-invariant fixed effects are not fully controlled for. Several factors are being listed for explaining the choice of such sampling design, including the factors of methodological and legal framework. In particular, Roma communities are still suspicious towards outsiders and survey takers and repeating the data collection in 2011 on the same clusters as in 2004 would be difficult because of the migration patterns (Ivanov et al., 2012).

2. Methodology

The scope of this paper is Romani women multiple disadvantage in the field of education. Three main explanatory variables regarding education are defined: 1) whether an individual succeeded in obtaining at least secondary education, 2) whether an individual has dropped out from school and 3) whether an individual can read and write.[4] All of these outcome variables are defined as dummy variables. The statistical inference regarding each of the educational outcomes is based on different sample. For the analysis of the educational level attainted a sample of individuals who are 25 to 64 years old is chosen. The upper bound is selected following the working age and the lower bound is chosen according to standard age of university/college graduate. For the analysis of school dropout rates the sample consists of individuals who are 9 to 17 years old. Finally, due to the small variation in the literacy rate among such young individuals, for the analysis of the literacy rates the sample is extended to individuals aged 9 to 25.

[4] The variables are derived based on the following questions: 1) "what is his/her highest attained education level? "; 2) "does s/he still attend school or training? "; 3) "can s/he read and write? ". The education levels are defined are follows: 1. None and incomplete basic; 2. Lower basic (1-4 years); 3. Incomplete upper basic; 4. Upper basic (5-8 years); 5. Incomplete secondary vocational/technical; 6. Secondary vocational/technical (1 or 2 years); 7. Secondary vocational/technical (3 or 4 years); 8. Incomplete secondary general; 9. Secondary general (4 years); 10. Incomplete college or university; 11. Associate (2 years) college; 12. Bachelor; 13. Masters; 14. PhD/Specialist. Obtaining at least secondary education corresponds to the highest attained levels denoted as 6, 7, 9, 10, 11, 12, 13, 14.

2.1. Determinants of Romani women's educational disadvantage

The analysis starts from the exploration of observable factors that may explain differences in the educational performance between men and women and Roma and non-Roma. Usually, empirical analyses of the ethnic and gender gaps are based on the estimation of the marginal effects on the dummy variables indicating gender and ethnic affiliations. When examining the gender and ethnic gaps simultaneously, one could account for both the indicators and investigate their marginal effects. Such approach would however assume that the gender gap is the same among the ethnic subpopulations (Roma and non-Roma) and the ethnic gap is the same among the subpopulations of men and women. Because of the high possibility that these gaps vary among the groups of individuals, the set of explanatory variables of the models should therefore include not only two dummy variables indicating whether an individual is a man or a woman and whether an individual belongs to the particular ethnic group but also their interaction.[5]

Since all three outcome variables are defined as dummy variables, the analysis of the factors that may contribute towards the formation of women's and Roma's disadvantage in education is based on the estimation of the probability models. In particular, present analysis is based on the linear probability model (LPM). The choice of the linear model instead of nonlinear probability models, such as probit or logit, is caused by to the fact that the set of explanatory variables that are of the main interest includes not only two dummy variables but also their interaction (as discussed above). Existing econometric literature has shown that in the non-linear models marginal effect on the interaction term is not equal to the marginal effect that is derived based on the common procedure done for single variables, which is usually enforced in the econometric software (for the discussion see Norton, Wang and Ai (2004) and Ai, Norton (2003)). In consequence, in the non-linear models the interpretation of the marginal effect of an interaction term is much more complex than in the linear case. To ease the understanding of the marginal effect of an interaction term, in the analysis presented in this paper, the linear probability model (LPM) is chosen and estimated with the OLS.[6] However, as

[5] Formally, consider the following model: $outcome_i = \alpha_1 + \alpha_2 Roma_i + \alpha_3 Female_i$ $\alpha_4 Interaction_i + \varepsilon_i$, where *Roma* is a dummy variable equal to 1 if an individual is Roma and 0 otherwise, *Female* is a dummy variable equal to 1 if an individual is female and 0 otherwise and *Interaction* is the interaction of these two. The marginal effects are the following: for non-Roma men α_1, for Roma men $\alpha_1 + \alpha_2$, for non-Roma women $\alpha_1 + \alpha_3$, for Roma women $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$.The resulting ethnic gap for men is α_2 and for women $\alpha_2 + \alpha_4$. Accordingly, the gender gap for non-Roma is given by α_3 and for Roma by $\alpha_3 + \alpha_4$.

[6] The author is aware of the methodological problems the analysis with the use of LPM involves. However given the interest in the marginal effect of being in the particular

the robustness check, respective nonlinear probit models are also estimated and presented in the Appendix (TABLE A. 1). In these models the coefficients on the interaction terms are corrected in the manner shown by Ai and Norton. The results obtained from probit and LPM are comparable which proves the robustness of the LPM estimates.

The model specification that is estimated with the use of LPM is the following:

$$Binary\ outcome_i = \alpha_1 + \alpha_2 Roma_i + \alpha_3 Female_i + \alpha_4 Interaction_i + \sum_{j=5}^{k} \alpha_j \boldsymbol{X}_i + \varepsilon_i; \tag{1}$$

where the vector X represents a set of control variables. Several sets of control variables that were found to have significant influence on Roma education (Brüggemann, 2012) are subsequently included in the equation. Such approach allows to compare how the initial values on Roma/gender coefficient change with subsequent controlling for additional factors; in other words it allows to identify factors that may explain Roma/gender differences in the educational outcomes.[7] Seven sets of control variables are considered: demographic characteristics, schooling conditions, health factor, living environment characteristics, family background variables, poverty and housing conditions. Demographic characteristics include age group dummies.[8] Schooling conditions consist of controls for the preschool and special school attendance. Health factor includes long-lasting illness indicator. Living environment controls are represented by two dummy variables: whether individual lives in an urban area and whether the primary school is located in walking distance. Following Brüggemann (2012) the family environment is accounted for by considering whether the household does posses books (more than 30). Moreover, following Kertesi and Kézdi (2011) the list of family background variables is additionally complemented with the indicator of internet access, the size of the household and a total number of unemployed adults living in the household. Moreover, in order to account for the gender dimension, education level of a partner of the household head, who in majority of the households is a woman, is also included. Poverty is additionally controlled for by the inclusion of the poverty indicator, i.e. whether a daily income per person is less than 4.30$

group (female/Roma) and the fact that the analysis involves the interaction term that is difficult to interpret in the non-linear case, the LPM is chosen keeping in mind that the predicted probabilities may be outside the interval of 0 and 1 and consequently the results may be biased and inconsistent (Horrace and Oaxaca, 2006). The problem of heteroskedasticity that is embedded in LPM is however addressed by the use of robust standard errors, which is now a common practice.

7 Similar approach was used for example by Kertesi and Kézdi (2011).

8 In the analysis of literacy rates demographic controls also include level of education attainted.

(PPP).[9] Finally, housing conditions include indicators of the number of rooms per person, squared meters per person, having a bathroom inside the dwelling and having an access to electricity supply. Moreover, the values and opinions on education and work of children may also impact the educational performance of Romani women. Two variables representing value of education are therefore additionally considered: 1) what is the perceived "appropriate" age to stop the education; 2) whether it is acceptable to work for children of primary school age.[10]

The econometric analysis is conducted based on all the available data and includes country specific fixed effects. Following Kertesi and Kézdi (2011) missing values of the explanatory variables are accounted for by the inclusion of dummy variables indicating missing status.

2.2. What about unobserved factors? Decomposing the gender and ethnic gaps in educational outcomes

Based on the regression models one may identify observable factors that to some extent explain Roma/gender lower educational performance. However, besides observable factors, individual unobservable characteristics, like ambition, ability, ingrained norms and values, etc. that are not fully accounted for by the models, may also influence Romani women's performance in terms of education. To analyze to what extent the difference in the educational outcomes may be explained by the distribution of the observable and unobservable factors, the ethnic and gender gaps are further decomposed using Ñopo nonparametric decomposition method (Ñopo (2004, 2008)).

Ñopo decomposition method has certain advantages over the regression models and decomposition methods that are based on them, including commonly applied Oaxaca-Blinder decomposition (1973)). In particular, in the regression models it is assumed that the impact of certain factors on the educational outcomes is fixed across the ethnic and gender subsamples, which is not necessarily true (e.g. 'better' family background may have more significant impact among the Roma

---

[9] The poverty line of 4.30$ (PPP) is chosen instead of for example 2.15$ or 1$ as the share of non-Roma individuals living below the poverty line defined on a lower level is relatively low (app. 3% in case of the poverty line at 2.15$).

[10] The consideration of such indicators is however problematic, because they are based on the answers of the survey question devoted to one randomly selected household member. The restriction of the sample to those who answered the question would significantly reduce its size. To avoid this problem it is assumed that the opinions of the random respondent are representative for the household he lives in and they are extrapolated for all the household members. Such procedure may however not be accurate. Therefore, these variables are not considered in the group of family background characteristics and only included as the last set of control variables.

community than among non-Roma). Moreover, these models assume that Roma and non-Roma as well as men and women are endowed with certain characteristics, which may be comparable. When measuring average outcomes between two groups of individuals, the problem of their limited comparability may however arise and cause serious problems for the validity of the inference. In particular, it may happen that the probability of observing an individual who shares comparable observable characteristics is close to zero, i.e. the possibility of finding comparable pairs of individuals is very low. For example it may happen that there is no Roma that shares the same family characteristics and living condition as non-Roma. In the econometric literature such situation is termed as a lack of so-called "common support". The method developed by Ñopo is based on nonparametric matching of 'otherwise comparable' individuals (in terms of observable characteristics), so it accounts for the distribution of the characteristics and the possibility of the lack of the common support. It also does not rely on any functional form such as the linearity of the regression models.

The method brings down to matching each female (Roma) with an 'otherwise comparable' male (non-Roma) and comparing their average outcomes. Once individuals are matched, the subsamples are divided in those for whom an adequate pair has been found (is in the common support) and those for whom there is no comparable individual, because no such individual who has comparable characteristics is observed in the sample (is out of the common support). Finally, the average outcomes are compared within those who are in the common support as well as those who could not be matched.[11] In consequence, the gap in the average outcome variables between two groups of individuals is decomposed into four components that consider the distribution of the characteristics:

$$\Delta = \Delta_X + \Delta_M + \Delta_F + \Delta_O, \tag{2}$$

where:
- Δ_X – the explained gap over the common support – the part of the gap that can be explained by the differences in the distribution of characteristics over the common support (for matched individuals);
- Δ_M – the explained part that can be explained by the differences in the distribution of characteristics of one group (non-Roma) that are in and out of the common support;
- Δ_F – the explained part that can be explained by the differences in the distribution of the characteristics between the individuals of the second group (Roma) that are in and out of the common support;

[11] The matching is performed with the use of 'Nopomatch' command available in Stata software. For a mathematical notation of the implemented procedure see Ñopo (2004, 2008).

– Δ_O – the unexplained part – a part that cannot be explained by the differences in the observed characteristics.

The ‘explained’ and ‘unexplained’ parts are interpreted in the similar manner as in the standard mean decomposition due to Oaxaca and Blinder (1973). The explained part is attributable to the observable differences in characteristics; the unobserved one in turn cannot be explained by them and is attributable to other complex structural factors that may cause lower achievement of one of the groups. Often, when gender inequality is considered the unexplained part is attributed to discrimination against women.

2.3. Is it getting better? Changes in education between 2004 and 2011

Lastly, to reveal whether the situation of Romani women in terms of education has became any better over the years 2004-2011 a difference-in-difference estimation proposed by Ashenfelter and Card (1985) is implemented. Although this estimation strategy is mainly used to track the changes of the policy introduction, it is also a straightforward tool that allows comparing the changes in the performance of two groups over the time.

In the present setting the difference-in-difference estimation brings down to the comparison of the difference in the outcomes of two groups (e.g. Roma and non-Roma) in one point in time to the respective difference occurring in the subsequent point in time. Eventually, the resulting estimates show the difference in the differences, i.e. difference in the gaps (gender gap/ethnic gap) between 2004 and 2011. Formally, for the comparison of the ethnicity based gaps over 2004-2011 the model is formulated as:

$$\begin{aligned} Binary\ outcome_{it} = {} & \beta_1 + \beta_2 Roma_{it} + \beta_3 Female_{it} + \delta_1(year2011_t) + \\ & \delta_2(year2011 * Roma_i) + \textstyle\sum_{j=4}^{k} \beta_j \boldsymbol{X}_{it} + u_{it} \end{aligned} \tag{3}$$

and for the comparison of the gender based gaps over 2004-2011 the model is specified as:

$$\begin{aligned} Binary\ outcome_{it} = {} & \beta_1 + \beta_2 Roma_{it} + \beta_3 Female_{it} + \gamma_1(year2011_t) + \\ & \gamma_2(year2011 * female_i) + \textstyle\sum_{j=4}^{k} \beta_j \boldsymbol{X}_{it} + \omega_{it} \end{aligned} \tag{4}$$

where the coefficients of interest are δ_2 and γ_2 respectively, t is an additional subscript indicating time (t=2004, 2011), and year 2011 is a dummy variable equal to 1 if the observation is observed in year 2011 and 0 otherwise.

The difference-in-difference estimators are implemented both to the whole sample and subsamples based on ethnicity and gender. This means that changes in the gender gaps over the time are analyzed among all individuals as well as among Roma and non-Roma separately. Similarly, the changes in the ethnic gaps over 2004-2011 are analyzed among all individuals regardless of gender as well as among men and women separately. The analysis of the dynamics among the subsamples is thus in line with the approach used in section 2.1. that allows for the variation of the gender gaps among Roma and non-Roma as well as the ethnic gaps among men and women.

3. Empirical results

3.1. Summary statistics on educational outcomes

Before turning to the description of the empirical results obtained from econometric analysis, the summary measures of the outcome variables are discussed. The statistics on the level of education show that 24.6% of Roma aged 25 to 64 have no formal education at all compared to only 1.8% of non-Roma (Table 1).The respective percentages for women and men are 22.4% and 13.3%. The data thus suggest that Romani women to a higher extent have lower educational level than both Romani men and non-Romani women. Indeed, the disaggregated data show that 30.7% of Romani women have no formal education at all compared to 18.3% of Romani men and 2.4% of non-Romani women.

Such high percentage of Romani men and women who do not have formal education at all may be partly explained by the nature of the sample.[12] Taking into account the above results, it is not surprising that Romani women also experience higher dropout rates than Romani males (by approximately 2%) and non-Romani females (by 23%). On the other hand, summary measures on literacy rates show that the differences in the literacy rate are mainly due to the ethnicity and the gender dimension in the literacy gap is rather minor.

Table 1. Summary measures on the main educational outcome variables[13]

Variable	Roma	Non-Roma	Women	Men	Roma		non-Roma	
					Women	Men	Women	Men
Sample 9 to 17								
Dropout	0.321	0.093	0.292	0.273	0.331	0.310	0.099	0.087
No. Obs.	8304	1668	4939	5033	4118	4186	821	847
Sample 9 to 25								
Literacy	0.871	0.981	0.882	0.900	0.860	0.882	0.978	0.984
No. Obs.	14128	3232	8748	8612	7118	7010	1630	1602

[12] Because of the sampling design, which was limited to acquiring information in the areas of compact Roma settlements, those Roma who are living in integrated communities and have higher education were not considered in the survey.

[13] Detailed summary statistics on other variables used in the analysis are available from the author upon the request.

Sample 25 to 65								
No education	0.246	0.018	0.224	0.133	0.307	0.183	0.024	0.013
ISCED1	0.267	0.064	0.214	0.201	0.270	0.263	0.077	0.050
ISCED2	0.361	0.267	0.318	0.349	0.330	0.393	0.290	0.243
ISCED3	0.121	0.561	0.212	0.289	0.088	0.156	0.513	0.611
ISCED4 and 5	0.003	0.089	0.030	0.026	0.003	0.003	0.096	0.082
At least secondary education (ISCED 3)	0.124	0.650	0.242	0.315	0.091	0.159	0.609	0.694
No. Obs.	17402	7204	12554	12052	8887	8515	3667	3537

Source: 2011 UNDP/WB/EC regional survey on Roma communities

Notes: 1. Level of education is classified according to International Standard Classification of Education (ISCED) terminology. The lowest level is ISCED 1, which is primary education that usually starts at age of 6 and lasts between 4 to 6 years. ISCED 2 stands for lower secondary education that follows primary education and usually lasts between four to six years. ISCED 3 follows ISCED 2 and lasts between two to five years – students usually leave this level of education at age 17 to 20. Finally ISCED 4 refers to post-secondary but not tertiary education and ISCED 5 and higher for different levels of tertiary education (UNESCO, 2012).

3.2. Results: Drivers of the educational disadvantage

The results from the estimation of the linear probability models are presented in Table 2. The estimates show that among non-Roma, females are by 8.5% less likely than non-Roma males to obtain at least secondary education. The interaction term is insignificant, which indicates that Romani women in comparison to Romani men are not significantly more disadvantaged than non-Romani women compared to non-Romani men. The findings also show that Romani males if compared with non-Romani males are more than twice less likely to gain at least secondary education (marginal effect of -0.538). Similarly, the ethnic gap among women is not significantly different as the interaction term remains small and statistically insignificant.

Once controls variables are subsequently included in the estimated models, the findings show that family background characteristics explain substantial part of the ethnic gap. In particular, when the family background characteristics are controlled for, the ethnic gap reduces to -0.113 (see Model 6 – Family background controls in Table 2). At the same time, family background variables do not change the gender gap drastically (-0.073). Additional controlling for the housing conditions and values concerning education only slightly influences the initial findings.

The findings concerning the probability of dropping out of the school show that among non-Roma women are not subject to significantly higher dropout rates than men. Similarly to previous analysis, the insignificance of the interaction term means that Romani women are not found to be more disadvantaged due to gender than women from the majority of the population. However, significant gap in the dropout rates is present among Roma and non-Roma groups: Roma are by 11%

more likely to drop out from the school even once the ethnic differences in the demographic characteristics, family background, living environment and values concerning education are controlled for. The results indicate that similarly to the previous analysis, family background is the key observable determinant that contributes towards the formation of ethnic-based inequalities in the probability of dropping out of the school (see Model 6).

Table 2. Partial effects of linear probability models for three outcome variables: 'having obtained at least secondary education', 'having dropped out from the school', 'having an ability to read and write'

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se
Outcome variable: 'having obtained at least secondary education'									
Roma	-0.538***	-0.545***	-0.512***	-0.506***	-0.506***	-0.113***	-0.112***	-0.106***	-0.104***
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Female	-0.085***	-0.084***	-0.084***	-0.082***	-0.082***	-0.073***	-0.073***	-0.074***	-0.074***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
Interaction	0.018	0.017	0.018	0.018	0.018	0.003	0.003	0.003	0.003
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.008)	(0.008)	(0.008)	(0.008)
N	24 555	24 555	24 555	24 555	24 555	24 555	24 555	24 555	24 555
R2	0.326	0.330	0.346	0.349	0.350	0.677	0.677	0.678	0.678
Outcome variable: 'having dropped out from the school'									
Roma	0.243***	0.258***	0.214***	0.213***	0.212***	0.137***	0.132***	0.123***	0.110***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.015)	(0.015)	(0.015)	(0.015)
Female	0.013	0.018	0.016	0.017	0.018	0.016	0.015	0.015	0.015
	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)
Interaction	0.006	-0.001	0.003	0.002	0.001	0.001	0.003	0.004	0.006
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
N	9 972	9 972	9 972	9 972	9 972	9 972	9 972	9 972	9 972
R2	0.123	0.214	0.247	0.248	0.249	0.261	0.266	0.274	0.280
Outcome variable: 'having an ability to read and write'									
Roma	-0.112***	-0.062***	-0.044***	-0.043***	-0.043***	-0.027***	-0.025***	-0.019***	-0.015**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Female	-0.008	-0.011**	-0.010*	-0.010*	-0.010*	-0.009	-0.009	-0.010*	-0.010*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Interaction	-0.012	-0.003	-0.006	-0.006	-0.006	-0.006	-0.007	-0.006	-0.006
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
N	17 360	17 360	17 360	17 360	17 360	17 360	17 360	17 360	17 360
R2	0.089	0.148	0.170	0.173	0.174	0.180	0.183	0.190	0.192

Source: Estimated on the basis of the 2011 UNDP/WB/EC regional survey onRoma communities

Notes:

1. Model 1: Gross model; Model 2: + Age control; Model 3: + Schooling controls; Model 4: + Health control; Model 5: + Living environment controls; Model 6: + Family background controls; Model 7: + Poverty control; Model 8: + Housing conditions controls; Model 9: + Values controls.
2. Gross model and all successive models additionally include country fixed effects;
3. Robust standard errors reported in the parenthesis: *** p<0.01, ** p<0.05, * p<0.1;
4. Detailed estimation results are presented in the Appendix.

The results concerning self-perceived literacy show that the gender gap in literacy rate is rather negligible and it is mostly the ethnic gap that is significant. As in the previous estimations, the interaction term is insignificant indicating that the ethnic gap is essentially of the same size for men and women. The results once again confirm that family background constitutes the key factor that explains the ethnic gap in literacy of young persons (see Model 6).

3.3. Decomposition results

The decomposition results of the gender and ethnic gaps in obtaining at least secondary education are presented in Table 3. The findings show that only part of the gender gap in obtaining at least secondary education among both Roma and non-Roma may be attributable to the distribution of the observable characteristics. This means that the gender gap in educational level is to a high extent driven by unobservable factors that cannot be directly accounted for (e.g. parental attitudes towards education, educational aspirations, etc.). The explained part of the gender gap in education among Roma accounts for 37% and among non-Roma for 39%. Moreover, high Δ_M component among Roma indicates that significant part of the gap may be explained by the fact that for some Romani men there are no comparable Romani woman (i.e. some men are out of the common support). The fact that the component is positive means that men who remain unreached by Romani women tend to have "better" characteristics. In consequence, if Romani women had the same attributes as Romani men that remain unreached by them, then the gap in education attainment would decrease.

Table 3. Ñopo decomposition of the ethnic and gender gaps in obtaining at least secondary education

Outcome variable: 'having obtained at least secondary education'									
	Male-Female gap					Non-Roma-Roma gap			
	Roma		Non-Roma			Female		Male	
Total gap	**0.7501**		**0.1386**		**Total gap**	**5.7134**		**3.3676**	
Δ_O	**0.4762**	**63%**	**0.0849**	**61%**	Δ_O	**1.0338**	**18%**	**0.7531**	**22%**
Δ_M	0.5212	**69%**	-0.095	**-69%**	Δ_M	3.1753	**56%**	1.8454	**55%**
Δ_F	-0.2117	**-28%**	0.1124	**81%**	Δ_F	0.1854	**3%**	-0.127	**-4%**
Δ_X	-0.0356	**-5%**	0.0363	**26%**	Δ_X	1.3189	**23%**	0.8961	**27%**
Explained total	**0.2739**	**37%**	**0.0537**	**39%**	**Explained total**	**4.6796**	**82%**	**2.6145**	**78%**

Source: Estimated on the basis of the 2011 UNDP/WB/EC regional survey on Roma communities

Note: The is performed based on the following variables: age dummies, special school indicator, long-standing illness indicator, urban area indicator, books indicator, Internet indicator, log of HHs size, total number of unemployed adults in the HH, head of the HH has at least secondary education, partner of the HH's head has at least secondary education. Country fixed effects are additionally controlled for.

Once the ethnic gaps are considered a meaningful part is explained by the differences in the distribution of the characteristics of Roma and non-Roma. Similarly, detailed decomposition results show that there are some characteristics of non-Roma individuals which are not shared by Roma that contribute towards the persistence of the gap in education. If Roma achieved characteristics of non-Roma – especially as regards family background - then the gap would significantly decrease - by more than 50%. This is true both for men and women. The results indicate that the unobservable factors also contribute towards explaining the ethnic gap in education, which is shown by the unexplained component. Nevertheless, mostly the gap remains due to the differences in the observable factors and particularly family 'endowments' of Roma and non-Roma.

The decomposition results of the ethnic gap in the probability of dropping out from the school (outcome variable: 'having dropped out from the school') are presented in Table 4. The results of the decomposition of the gender gaps in the dropout rates and the results from the decomposition of the gender and ethnic gaps in the literacy rates (outcome variable: 'having an ability to read and write') are not presented as they are rather small and insignificant.

Table 4. Ñopo decomposition of the ethnic gap in dropout rates

Outcome variable: 'having dropped out from the school'					
Non-Roma-Roma gap					
	Female		Male		
Total gap	**-0.7151**		**Total gap**	**-0.723**	
Δ_O	**-0.8821**	**123%**	**Δ_O**	**-0.502**	**69%**
Δ_M	-0.059	**8%**	Δ_M	-0.2256	**31%**
Δ_F	0.4122	**-58%**	Δ_F	0.2562	**-35%**
Δ_X	-0.1862	**26%**	Δ_X	-0.2514	**35%**
Explained total	**0.167**	**-23%**	**Explained total**	**-0.2208**	**31%**

Source: Estimated on the basis of the 2011 UNDP/WB/EC regional survey on Roma communities

Note: The decomposition of the gaps in the dropout rate is performed based on the following variables: age dummies, special school indicator, long-standing illness indicator, urban area indicator, walking distance to school indicator, books indicator, Internet indicator, log of HHs size, total number of unemployed adults in the HH, head of the HH has at least secondary education, partner of the HH's head has at least secondary education. Country fixed effects are additionally controlled for.

The findings show that among females the ethnic gap in dropout rates is left unexplained, suggesting that the differences in the distribution of characteristics among Romani women and non-Romani women are not solely enough to explain Romani women higher dropout rates. Among men, the differences in the distribution of characteristics to some extent may explain Romani men higher dropout rates. However, still large part (app. 70%) is due to some other complex unobserved factors that cause Romani boys higher dropout rates. According to Pantea

(2009) these factors may include: 1) customs and values, 2) conditions of Romafamilies and communities, 3) structural constrains related to school systems, 4) (perceived) returns of education, 5) policy ineffectiveness. Furthermore, these factors are likely to intertwine and consequently may have a long lasting impact on the lives of Romani women.

3.4. Difference-in-difference estimation results

The last section presents the dynamics in gender and ethnic inequality in education over the years 2004-2011.

Table 5. Difference-in-difference estimates of the educational outcomes

Outcome variable: 'having obtained at least secondary education'					
Gender effect 2004	-0.113***	(0.005)	Roma effect 2004	-0.092***	(0.007)
Gender effect 2011	-0.092***	(0.004)	Roma effect 2011	-0.078***	(0.006)
Diff-in-diff estimator	0.021***	(0.006)	Diff-in-diff estimator	0.014**	(0.007)
Among Roma			**Among males**		
Gender effect 2004	-0.103***	(0.007)	Roma effect 2004	-0.081***	(0.008)
Gender effect 2011	-0.090***	(0.005)	Roma effect 2011	-0.076***	(0.007)
Diff-in-diff estimator	0.013	(0.008)	Diff-in-diff estimator	0.004	(0.009)
Among non-Roma			**Among females**		
Gender effect 2004	-0.123***	(0.007)	Roma effect 2004	-0.098***	(0.008)
Gender effect 2011	-0.094***	(0.008)	Roma effect 2011	-0.080***	(0.006)
Diff-in-diff estimator	0.029***	(0.011)	Diff-in-diff estimator	0.018*	(0.009)
Outcome variable: 'having dropped out from the school'					
Gender effect 2004	0.037***	(0.013)	Roma effect 2004	0.157***	(0.020)
Gender effect 2011	0.019**	(0.009)	Roma effect 2011	0.097***	(0.015)
Diff-in-diff estimator	-0.018	(0.016)	Diff-in-diff estimator	-0.061***	(0.023)
Among Roma			**Among males**		
Gender effect 2004	0.052***	(0.019)	Roma effect 2004	0.155***	(0.026)
Gender effect 2011	0.019*	(0.011)	Roma effect 2011	0.085***	(0.021)
Diff-in-diff estimator	-0.032	(0.021)	Diff-in-diff estimator	-0.07**	(0.031)
Among non-Roma			**Among females**		
Gender effect 2004	0.014	(0.012)	Roma effect 2004	0.156***	(0.029)
Gender effect 2011	0.001	(0.014)	Roma effect 2011	0.106***	(0.022)
Diff-in-diff estimator	-0.013	(0.018)	Diff-in-diff estimator	-0.049	(0.034)
Outcome variable: 'having an ability to read and write'					
Gender effect 2004	-0.009	(0.007)	Roma effect 2004	-0.024**	(0.010)
Gender effect 2011	-0.021***	(0.005)	Roma effect 2011	0.006	(0.008)
Diff-in-diff estimator	-0.012	(0.008)	Diff-in-diff estimator	0.030***	(0.011)

Among Roma			Among males		
Gender effect 2004	-0.021**	(0.010)	Roma effect 2004	-0.017	(0.013)
Gender effect 2011	-0.021***	(0.006)	Roma effect 2011	0.007	(0.011)
Diff-in-diff estimator	0	(0.012)	Diff-in-diff estimator	0.023	(0.015)

Among non-Roma			Among females		
Gender effect 2004	0.006	(0.005)	Roma effect 2004	-0.022	(0.015)
Gender effect 2011	-0.010*	(0.006)	Roma effect 2011	0.006	(0.012)
Diff-in-diff estimator	-0.016**	(0.007)	Diff-in-diff estimator	0.028	(0.017)

Source: Estimated based on the 2011 UNDP/WB/EC regional survey on Roma communities and 2004 UNDP Roma Rerional Survey.

Notes: DiD estimator based on the regression with the following right-hand side variables: dropout rate: age, illness indicator, urban area, books indicator, Internet indicator, log of household size, total number of unemployed adults in the household, household's head and partner's education, poverty indicator, squared meters per capita, rooms per capita, bathroom indicator, electricity indicator and country fixed effects; literacy: lower basic or no education, age, urban area, books indicator, Internet indicator, log of household size, household's head and partner's education, poverty indicator, bathroom indicator, electricity indicator and country fixed effects; education level: age, illness indicator, urban area, books indicator, Internet indicator, log of household size, total number of unemployed adults in the household, household's head and partner's education, poverty indicator, squared meters per capita, rooms per capita, bathroom indicator, electricity indicator and country fixed effects.

Standard errors reported in the parenthesis: *** p<0.01, ** p<0.05, * p<0.1.

Detailed estimation results are available from the author upon the request.

The findings presented in Table 5 show that, over the years 2004-2011 there has been a positive change in the female's probability of acquiring at least secondary education. This means that in 2011 women were less unlikely to obtain at least secondary education than in 2004. Similar patterns have been revealed for Roma and non-Roma groups. The difference in difference estimates moreover indicates that negative premium from being Roma has declined. More detailed consideration of the changes in the ethnic gap within females and males shows that similar patterns are present for both groups, i.e. the ethnic gap in the educational attainment has decreased both for males and females.

As far as dropout rates are considered, both the ethnic and gender gap in the average dropout rates were higher in 2004 and they have significantly declined till 2011. As regards the ethnic gap (i.e. Roma effect), the most significant decline occurred among males – the probability of dropping out the school for this group of individuals has declined nearly by half (if compared with non-Roma males). Among females, changes in the ethnic gap in the dropout rates were less significant, but still present.

The estimates for literacy rate show that over 2004-2011 the overall gender gap has increased but still remains relatively low. The closer examination of the gender gap reveals that increase of the negative premium from being a female is found for non-Roma individuals; for Roma the gender gap in literacy rate is stable in this two time periods. Roma effects (ethnic gaps) regarding literacy rates are found to be insignificant and rather small in magnitude.

Overall, among both Roma and non-Roma females are slightly less unlikely to obtain secondary and higher education and less likely to drop out from school. Still however the gender gap in these dimensions is present. Alike, changes in the ethnic gap in educational outcomes between 2004 and 2011 were positive but rather minor. They were also more pronounced among men than women.

Conclusion

This paper seeks to outline intersectional inequalities based on gender and ethnicity that persist in educational performance. The findings from 2011 UNDP/WB/EC regional survey on Roma communities show that Romani women suffer from lower educational attainment – both in comparison to non-Romani women (higher disparities) and Romani men. They have also higher dropout rates and lower literacy rates than non-Romani women. Roma-non-Roma gap in education is however more significant than the gender gap, which means that Romani women are much more disadvantaged compared to non-Romani women than to Romani men. Still, they face double disadvantage, especially in chances of acquiring higher education, as they are exposed both to the ethnic and gender gap. It has to be however clearly noted that the data comprise information on Romani and non-Romani individuals who are living in a close proximity and the results cannot be generalized to the whole population. The analysis of the dynamics of the gender and ethnic gaps in education resolves that although some positive changes have been observed, the pace of the progress is the same among the subpopulation disaggregated by gender and ethnicity, leaving Romani women still behind the majority of the population. The analysis thus reveals that Romani women are still exposed to a higher risk of failure of obtaining education and thus greater economic vulnerability and dependence from men.

The regression analysis moreover reveals some interesting patterns with respect to the possible factors that influence the persistence of the gender and ethnic gaps in educational outcomes. It indicates that family background and perceived value of education are the key observable driving forces of the existing ethnic gaps in education. The results from the decompositions moreover show that differences in endowments in these attributes of men and women as well as Roma and non-Roma do not entirely explain their unequal performance and there are other unobservable factors that contribute towards the persistence of the gender and ethnic differences. In order to confront these factors there is therefore a high need for gender sensitive policy interventions that will raise the awareness of the problem and facilitate Romani women greater education.

Ensuring better education prospects for Romani women goes beyond personal gains of the certain generation and makes an intergenerational impact on their children as well as on wider communities (e.g. Summers, 1994; World Bank, 2001). Measures and attempts to facilitate Romani women quality education are therefore highly needed as they provide means for alleviating the poverty of Roma and their greater integration with majority population. Common actions aimed at better education of Romani women may also further lead to their economical and political empowerment and thus greater civic engagement that is critical for Roma minority social integration. Consequently, investing in Romani women will make the efforts in regards Roma inclusion more effective for the benefit of the present as well as future generations.

References

Ai C., Norton E. (2003). Interaction terms in logit and probit models, Economics Letters, 80, 123–129.

Ashenfelter O., Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs, Review of Economics and Statistics, 67, 648–660.

Blinder A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources, 8, 436-455.

Brummet, Q. (2008).The effect of gender inequality on growth: a cross country empirical study. The Park Place Economist, Vol. 16, 13-23.

Brüggemann, C. (2012). Roma Education in Comparative Perspective. Findings from the UNDP/World Bank/EC Regional Roma Survey. Roma Inclusion Working Papers. Bratislava: UNDP.

Cukrowska E., Kóczé A. (2013). Interplay between gender and ethnicity: exposing structural disparities of Romani women. Analysis of the UNDP/World Bank/EC regional Roma survey data. Roma Inclusion Working Papers. Bratislava: UNDP.

Davalos, M. E., Sattar, S., Simler, K., Bidani, B., Reva, A., Tsirunyan, S., Orlando, M. B. (2012). Kosovo - Gender gaps in education, health, and economic opportunities. The World Bank Group Europe and Central Asia region poverty reduction and economic management. Washington DC. World Bank.

Horrace W. C., Oaxaca R. L. (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. Economics Letters, 90, 321-32.

Ilisei I. (2013). Education of Roma Women between Feminism and Multiculturalism Case study: Roma Women in Romania. Journal of Social Science Education, 12(1), 67-74.

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

Ivanov A., Kagin, J. Kling, J. (2012). Integrated household surveys among Roma populations: one possible approach to sampling used in the UNDP-World Bank-EC Regional Roma Survey 2011. Roma Inclusion Working Papers. Bratislava: UNDP.

Kertesi G., Kézdi G. (2011). The Roma/Non-Roma Test Score Gap in Hungary, American Economic Review, 101(3), 519-25.

Ñopo H. (2004). The Gender Wage Gap in Peru 1986-2000: Evidence from a Matching Comparisons Approach. Económica. Departamento de Economía. Facultad de Ciencias Económicas. Universidad Nacional de La Plata 0(1-2), 9-37.

Ñopo H. (2008). Matching as a Tool to Decompose Wage Gaps. The Review for Economics and Statistics, 90(2), 290-299.

Norton, E. C., H. Wang and C. Ai. (2004). Computing interaction effects and standard errors in logit and probit models. The Stata Journal, 4(2), 154-167.

Oaxaca R. L. (1973). Male-Female Wage Differentials in Urban Labor Markets. International Economic Review, 14, 693-709.

Pantea, M. C. (2009). Gender Mainstreaming in the National Action Plans for the Decade of Roma Inclusion. UNIFEM CEE office and UNDP Bratislava Regional Centre, http://www.unifem.sk/uploads/doc/GM_in_Roma_Decade_NAPs.pdf

Schultz, T.P. (1993) Investments in the Schooling and Health of Women and Men: Quantities and Returns. The Journal of Human Resources, 28 (4), 694-734.

Summers, L.H. (1994). Investing in all the People: Educating Women in Developing Countries. Washington, D.C.: The World Bank

UNDP (2013). Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World

UNESCO (2010) GLOBAL EDUCATION DIGEST 2010 Comparing Education Statistics Across the World. Special focus on gender.

UNESCO (2012) International Standard Classification of Education ISCED 2011

World Bank (2001). Engendering development: through gender equality in rights, resources, and voice. World Bank policy research report.

APPENDIX

TABLE A. 1. Marginal effects of probit models for three outcome variables: 'having obtained at least secondary education', 'having dropped out from the school', 'having an ability to read and write'.

Variables	Model 1	Model 2	Model 3	Model 4
	coef/se	coef/se	coef/se	coef/se
Outcome variable: 'having obtained at least secondary education'				
Roma	-0.536***	-0.548***	-0.516***	-0.507***
	(0.009)	(0.010)	(0.010)	(0.010)
Female	-0.073***	-0.0725***	-0.072***	-0.0692***
	(0.009)	(0.009)	(0.009)	(0.009)
Interaction	0.017	0.015	0.015	0.012
	(0.013)	(0.013)	(0.013)	(0.013)
Outcome variable: 'having dropped out from the school'				
Roma	0.235***	0.241***	0.214***	0.214***
	(0.012)	(0.011)	(0.012)	(0.012)
Female	0.022	0.027	0.026	0.028
	(0.030)	(0.031)	(0.031)	(0.031)
Interaction	0.007	0.005	0.006	0.005
	(0.017)	(0.017)	(0.018)	(0.018)
Outcome variable: 'having an ability to read and write'				
Roma	-0.088***	-0.021***	-0.015***	-0.015***
	(0.005)	(0.004)	(0.003)	(0.003)
Female	-0.019	-0.009	-0.008	-0.008
	(0.015)	(0.006)	(0.005)	(0.005)
Interaction	-0.014***	0.004	0.005	0.005
	(0.007)	(0.010)	(0.012)	(0.011)

Notes:

1. The probit estimates correspond to estimates from the linear probability model presented in Table 2.
2. Model 1: Gross model; Model 2: + Age control; Model 3: + Schooling controls; Model 4: + Health control; Model 5: + Living environment controls; Model 6: + Family background controls; Model 7: + Poverty control; Model 8: + Housing conditions controls; Model 9: + Values controls.
3. Gross model and all successive models additionally include country fixed effects;
4. Robust standard errors reported in the parenthesis: ***p<0.01, **p<0.05, *p<0.1;
5. Marginal effects of the interaction term are computed according to the formula: $\Delta 2\Phi(x\alpha)/\Delta x1\Delta x2 = \Phi(\alpha 1 + \alpha 2 + \alpha 12 + x\alpha) - \Phi(\alpha 1 + x\alpha) - \Phi(\alpha 2 + x\alpha) + \Phi(x\alpha)$. For the discussion on the technical problems with the derivation of marginal effects and standard errors of interaction terms for probit and logit models see: Ai and Norton (2003), Norton et al. (2004)

Model 5	Model 6	Model 7	Model 8	Model 9
coef/se	coef/se	coef/se	coef/se	coef/se
-0.508***	-0.162***	-0.156***	-0.143***	-0.138***
(0.010)	(0.014)	(0.014)	(0.013)	(0.013)
-0.069***	-0.119***	-0.117***	-0.116***	-0.117***
(0.009)	(0.013)	(0.013)	(0.013)	(0.012)
0.012	0.004	0.003	0.001	0.000
(0.013)	(0.011)	(0.010)	(0.010)	(0.010)
0.214***	0.153***	0.150***	0.145***	0.136***
(0.012)	(0.017)	(0.017)	(0.017)	(0.018)
0.029	0.025	0.025	0.026	0.029
(0.031)	(0.033)	(0.032)	(0.032)	(0.032)
0.005	0.000	0.002	0.002	0.002
(0.018)	(0.022)	(0.023)	(0.023)	(0.023)
-0.015***	-0.009***	-0.009***	-0.008***	-0.007***
(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
-0.008	-0.007	-0.007	-0.007	-0.007
(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
0.005	0.013	0.013	0.015	0.015
(0.011)	(0.015)	(0.015)	(0.016)	(0.016)

TABLE A. 2 Partial effects of linear probability model estimates; dependent variable 'having obtained at least secondary education', sample of individuals 25 to 65 years old

Variables coef/se		Model 1 coef/se	Model 2 coef/se	Model 3 coef/se
	Roma	-0.538***	-0.545***	-0.512***
		(0.008)	(0.008)	(0.009)
	Female	-0.085***	-0.084***	-0.084***
		(0.011)	(0.011)	(0.011)
	Interaction	0.018	0.017	0.018
		(0.012)	(0.012)	(0.012)
Demographic characteristics	Age (25 to 29)		0.099***	0.066***
			(0.010)	(0.010)
	Age (30 to 34)		0.079***	0.048***
			(0.010)	(0.010)
	Age (35 to 39)		0.097***	0.069***
			(0.010)	(0.010)
	Age (40 to 44)		0.104***	0.081***
			(0.011)	(0.011)
	Age (45 to 49)		0.102***	0.088***
			(0.011)	(0.011)
	Age (50 to 54)		0.067***	0.058***
			(0.011)	(0.011)
	Age (55 to 59)		0.042***	0.038***
			(0.011)	(0.011)
Schooling conditions	Special school			-0.129***
				(0.015)
	Special school - missing			-0.061***
				(0.010)
	Preschool			0.139***
				(0.007)
Health	Long lasting illness			
	Long lasting illness - missing			
Living environment	Urban			

Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
coef/se	coef/se	coef/se	coef/se	coef/se	
-0.506***	-0.506***	-0.113***	-0.112***	-0.106***	-0.104***
(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
-0.082***	-0.082***	-0.073***	-0.073***	-0.074***	-0.074***
(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
0.018	0.018	0.003	0.003	0.003	0.003
(0.012)	(0.012)	(0.008)	(0.008)	(0.008)	(0.008)
0.042***	0.041***	0.050***	0.051***	0.058***	0.058***
(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
0.027***	0.026**	0.034***	0.035***	0.042***	0.041***
(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
0.049***	0.048***	0.044***	0.045***	0.050***	0.050***
(0.010)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
0.066***	0.065***	0.042***	0.042***	0.047***	0.046***
(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
0.077***	0.075***	0.040***	0.040***	0.045***	0.045***
(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
0.050***	0.049***	0.026***	0.026***	0.030***	0.029***
(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
0.035***	0.034***	0.020***	0.020***	0.023***	0.022***
(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
-0.120***	-0.122***	-0.026**	-0.027**	-0.028***	-0.028***
(0.015)	(0.015)	(0.011)	(0.011)	(0.011)	(0.011)
-0.056***	-0.057***	-0.020**	-0.018**	-0.017**	-0.016**
(0.010)	(0.010)	(0.008)	(0.008)	(0.008)	(0.008)
0.139***	0.139***	0.043***	0.043***	0.043***	0.043***
(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
-0.051***	-0.052***	-0.023***	-0.022***	-0.021***	-0.021***
(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
-0.140***	-0.140***	-0.070***	-0.069***	-0.067***	-0.066***
(0.022)	(0.022)	(0.017)	(0.017)	(0.017)	(0.017)
	0.034***	0.005	0.005	0.004	0.004
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)

Family background variables	Books indicator			
	Books indicator - missing			
	Internet indicator			
	Internet indicator - missing			
	log of HHs size			
	Total number of unemployed in the HH			
	HH's head has at least secondary education			
	HH's head has at least secondary education - missing			
	Partner of HH's head has at least secondary education			
	Partner of HH's head has at least secondary education - missing			
Poverty	Poverty indicator			
	Poverty indicator - missing			
Housing conditions	Squared meters per capita			
	Rooms per capita			
	Squared meters per capita - missing			
	Rooms per capita - missing			
	Bathroom inside			
	Bathroom inside - missing			
	Electricity			
	Electricity - missing			

		0.033***	0.032***	0.026***	0.026***
		(0.006)	(0.006)	(0.006)	(0.006)
		0.046**	0.046**	0.046**	0.046**
		(0.019)	(0.019)	(0.019)	(0.019)
		0.027***	0.025***	0.019***	0.018***
		(0.005)	(0.005)	(0.005)	(0.005)
		-0.024	-0.024	-0.023	-0.023
		(0.024)	(0.024)	(0.024)	(0.024)
		-0.009**	-0.007**	0.007*	0.008*
		(0.003)	(0.003)	(0.004)	(0.004)
		-0.000	0.001	0.000	0.000
		(0.002)	(0.002)	(0.002)	(0.002)
		0.474***	0.473***	0.470***	0.469***
		(0.008)	(0.008)	(0.008)	(0.008)
		0.196*	0.198*	0.192*	0.193*
		(0.103)	(0.103)	(0.102)	(0.102)
		0.371***	0.372***	0.371***	0.370***
		(0.009)	(0.009)	(0.008)	(0.008)
		0.129***	0.129***	0.128***	0.128***
		(0.005)	(0.005)	(0.005)	(0.005)
			-0.019***	-0.013***	-0.012***
			(0.004)	(0.004)	(0.004)
			-0.002	-0.002	-0.002
			(0.005)	(0.005)	(0.005)
				0.000	0.000
				(0.000)	(0.000)
				0.004	0.005
				(0.012)	(0.012)
				0.020***	0.020***
				(0.005)	(0.005)
				-0.013	-0.014
				(0.025)	(0.025)
				0.003	0.003
				(0.005)	(0.005)
				-0.033	-0.034
				(0.021)	(0.021)
				0.023***	0.022***
				(0.004)	(0.004)
				0.038	0.039*
				(0.024)	(0.024)

Values	Appropriate age to stop education			
	Appropriate age to stop education - missing			
	Acceptable to work than go to school for children at primary school age			
	Acceptable to work than go to school for children at primary school age - missing			
	Number of observations	24 555	24 555	24 555
	R2	0.326	0.330	0.346

Notes:

1. Robust standard errors reported in the parenthesis; ***p<0.01, **p<0.05, *p<0.1
2. All regressions control for country fixed effects.

TABLE A. 3 Partial effects of linear probability model estimates; Dependent variable: 'having dropped out from the school', sample of individuals 9 to 17 years old

Variables coef/se		**Model 1**	**Model 2**	**Model 3**
		coef/se	**coef/se**	**coef/se**
	Roma	0.243***	0.258***	0.214***
		(0.013)	(0.013)	(0.013)
	Female	0.013	0.018	0.016
		(0.015)	(0.016)	(0.016)
	Interaction	0.006	-0.001	0.003
		(0.018)	(0.018)	(0.018)
Demographic characteristics	Age (9 to 14)		-0.281***	-0.273***
			(0.009)	(0.009)
Schooling conditions	Preschool			-0.063***
				(0.016)
	Preschool - missing			0.217***
				(0.030)
	Special school			-0.173***
				(0.009)
	Special school - missing			0.002
				(0.029)
Health	Long lasting illness			
	Long lasting illness - missing			

					0.000
					(0.000)
					0.006
					(0.006)
					-0.013***
					(0.004)
					-0.013*
					(0.008)
24 555	24 555	24 555	24 555	24 555	24 555
0.349	0.350	0.677	0.677	0.678	0.678

Model 4	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
coef/se	**coef/se**	**coef/se**	**coef/se**	**coef/se**	
0.213***	0.212***	0.137***	0.132***	0.123***	0.110***
(0.013)	(0.013)	(0.015)	(0.015)	(0.015)	(0.015)
0.017	0.018	0.016	0.015	0.015	0.015
(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)
0.002	0.001	0.001	0.003	0.004	0.006
(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
-0.272***	-0.272***	-0.277***	-0.276***	-0.277***	-0.276***
(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
-0.073***	-0.070***	-0.082***	-0.082***	-0.083***	-0.083***
(0.016)	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)
0.214***	0.213***	0.202***	0.196***	0.187***	0.180***
(0.030)	(0.030)	(0.029)	(0.029)	(0.029)	(0.029)
-0.173***	-0.173***	-0.158***	-0.154***	-0.150***	-0.145***
(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
0.005	0.003	-0.004	0.001	-0.013	
(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	
0.080***	0.080***	0.084***	0.083***	0.080***	0.080***
(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
-0.055	-0.054	-0.050	-0.054	-0.046	-0.056
(0.039)	(0.039)	(0.039)	(0.040)	(0.039)	(0.040)

Living environment	Urban			
	School in walking distance			
Family background variables	Books indicator			
	Books indicator - missing			
	Internet indicator			
	Internet indicator - missing			
	log of HHs size			
	Total number of unemployed in the HH			
	HH's head has at least secondary education			
	HH's head has at least secondary education - missing			
	Partner of HH's head has at least secondary education			
	Partner of HH's head has at least secondary education - missing			
Poverty	Poverty indicator			
	Poverty indicator - missing			

	-0.002	-0.000	0.000	0.006	0.009
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
	-0.037***	-0.034***	-0.032***	-0.033***	-0.032***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
		-0.063***	-0.056***	-0.048***	-0.044***
		(0.010)	(0.010)	(0.010)	(0.010)
		-0.071**	-0.076**	-0.079**	-0.080**
		(0.035)	(0.035)	(0.034)	(0.034)
		-0.031***	-0.024**	-0.006	-0.002
		(0.010)	(0.010)	(0.011)	(0.011)
		-0.072	-0.080*	-0.077*	-0.088**
		(0.045)	(0.045)	(0.045)	(0.043)
		0.045***	0.042***	0.040***	0.034***
		(0.012)	(0.012)	(0.013)	(0.013)
		-0.010**	-0.015***	-0.013***	-0.012***
		(0.004)	(0.004)	(0.004)	(0.004)
		-0.065***	-0.061***	-0.053***	-0.051***
		(0.011)	(0.011)	(0.011)	(0.011)
		0.034	0.035	0.040	0.023
		(0.061)	(0.062)	(0.060)	(0.061)
		-0.015	-0.015	-0.013	-0.010
		(0.013)	(0.012)	(0.012)	(0.012)
		0.034***	0.033***	0.031***	0.028**
		(0.011)	(0.011)	(0.011)	(0.011)
			0.081***	0.069***	0.065***
			(0.011)	(0.011)	(0.011)
			0.062***	0.058***	0.048***
			(0.015)	(0.015)	(0.015)

Housing conditions	Squared meters per capita			
	Rooms per capita			
	Squared meters per capita - missing			
	Rooms per capita - missing			
	Bathroom inside			
	Bathroom inside - missing			
	Electricity			
	Electricity - missing			
Values	Appropriate age to stop education			
	Appropriate age to stop education - missing			
	Acceptable to work than go to school for children at primary school age			
	Acceptable to work than go to school for children at primary school age - missing			
	Number of observations	9 972	9 972	9 972
	R2	0.123	0.214	0.247

Notes: 1. Robust standard errors reported in the parenthesis; *** p<0.01, ** p<0.05, * p<0.1

2. All regressions control for country fixed effects.

				0.001	0.001
				(0.001)	(0.001)
				0.053*	0.045
				(0.032)	(0.032)
				-0.012	-0.016
				(0.021)	(0.021)
				-0.065	-0.080
				(0.050)	(0.051)
				-0.079***	-0.080***
				(0.016)	(0.016)
				0.006	0.013
				(0.074)	(0.075)
				-0.067***	-0.063***
				(0.010)	(0.010)
				-0.059	-0.060
				(0.077)	(0.077)
					-0.000**
					(0.000)
					0.017
					(0.016)
					0.077***
					(0.009)
					0.070***
					(0.021)
9 972	9 972	9 972	9 972	9 972	9 972
0.248	0.249	0.261	0.266	0.274	0.280

TABLE A. 4 Partial effects of linear probability model estimates; dependent variable: 'having an ability to read and write', sample of individuals 9 to 25 years old

Variables **coef/se**		**Model 1**	**Model 2**	**Model 3**
		coef/se	coef/se	coef/se
	Roma	-0.112***	-0.062***	-0.044***
		(0.005)	(0.005)	(0.005)
	Female	-0.008	-0.011**	-0.010*
		(0.005)	(0.005)	(0.005)
	Interaction	-0.012	-0.003	-0.006
		(0.008)	(0.007)	(0.007)
Demographic characteristics	Age (9 to 14)		0.114***	0.097***
			(0.008)	(0.007)
	Age (15 to 19)		0.042***	0.036***
			(0.005)	(0.005)
	Basic or no education		-0.196***	-0.182***
			(0.006)	(0.005)
	Basic or no education - missing		-0.194***	-0.160***
			(0.022)	(0.022)
Schooling conditions	Preschool			-0.056***
				(0.012)
	Special school			-0.227***
				(0.022)
	Special school - missing			0.066***
				(0.004)
Health	Long lasting illness			
	Long lasting illness - missing			
Living environment	Urban			

Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
coef/se	coef/se	coef/se	coef/se	coef/se	
-0.043***	-0.043***	-0.027***	-0.025***	-0.019***	-0.015**
(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
-0.010*	-0.010*	-0.009	-0.009	-0.010*	-0.010*
(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
-0.006	-0.006	-0.006	-0.007	-0.006	-0.006
(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
0.095***	0.095***	0.102***	0.101***	0.097***	0.096***
(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
0.035***	0.035***	0.041***	0.042***	0.040***	0.040***
(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
-0.180***	-0.180***	-0.174***	-0.170***	-0.164***	-0.162***
(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
-0.159***	-0.160***	-0.162***	-0.158***	-0.152***	-0.148***
(0.022)	(0.022)	(0.024)	(0.024)	(0.025)	(0.025)
-0.047***	-0.046***	-0.041***	-0.041***	-0.041***	-0.040***
(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
-0.225***	-0.225***	-0.222***	-0.218***	-0.212***	-0.211***
(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)
0.066***	0.066***	0.061***	0.059***	0.057***	0.056***
(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
-0.063***	-0.062***	-0.064***	-0.064***	-0.062***	-0.062***
(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
0.020	0.021	0.020	0.021	0.018	0.020
(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
	-0.025***	-0.026***	-0.026***	-0.030***	-0.032***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)

Family background variables	Books indicator				
	Books indicator - missing				
	Internet indicator				
	Internet indicator- missing				
	log of HHs size				
	Total number of unemployed in the HH				
	HH’s head has at least secondary education				
	HH’s head has at least secondary education - missing				
	Partner of HH’s head has at least secondary education				
	Partner of HH’s head has at least secondary education - missing				
Poverty	Poverty indicator				
	Poverty indicator - missing				

		0.017***	0.013***	0.007	0.005
		(0.004)	(0.004)	(0.004)	(0.004)
		0.051***	0.054***	0.059***	0.060***
		(0.013)	(0.013)	(0.014)	(0.014)
		0.007*	0.003	-0.010**	-0.011**
		(0.004)	(0.004)	(0.005)	(0.005)
		-0.006	-0.005	-0.005	-0.002
		(0.021)	(0.021)	(0.021)	(0.021)
		-0.045***	-0.042***	-0.040***	-0.039***
		(0.006)	(0.006)	(0.007)	(0.007)
		0.012***	0.014***	0.013***	0.013***
		(0.002)	(0.002)	(0.002)	(0.002)
		0.009**	0.007	0.002	0.001
		(0.005)	(0.005)	(0.005)	(0.005)
		0.100***	0.102***	0.095***	0.100***
		(0.032)	(0.032)	(0.032)	(0.032)
		-0.008*	-0.007	-0.011**	-0.012***
		(0.005)	(0.005)	(0.005)	(0.005)
		-0.022***	-0.021***	-0.020***	-0.019***
		(0.006)	(0.006)	(0.006)	(0.006)
			-0.052***	-0.045***	-0.044***
			(0.007)	(0.007)	(0.007)
			-0.025***	-0.025***	-0.022***
			(0.007)	(0.007)	(0.007)

Housing conditions	Squared meters per capita			
	Rooms per capita			
	Squared meters per capita - missing			
	Rooms per capita - missing			
	Bathroom inside			
	Bathroom inside - missing			
	Electricity			
	Electricity - missing			
Values	Appropriate age to stop education			
	Appropriate age to stop education - missing			
	Acceptable to work than go to school for children at primary school age			
	Acceptable to work than go to school for children at primary school age - missing			
	Number of observations	17 360	17 360	17 360
	R2	0.089	0.148	0.170

Notes:

1. Robust standard errors reported in the parenthesis; ***p<0.01, **p<0.05, *p<0.1

2. All regressions control for country fixed effects.

					0.001*	0.001*
					(0.000)	(0.000)
					0.006	0.009
					(0.015)	(0.015)
					-0.007	-0.007
					(0.008)	(0.008)
					0.017	0.019
					(0.022)	(0.022)
					0.028***	0.028***
					(0.010)	(0.010)
					0.040	0.036
					(0.040)	(0.040)
					0.050***	0.049***
					(0.005)	(0.005)
					-0.056	-0.053
					(0.044)	(0.044)
						0.000***
						(0.000)
						-0.004
						(0.009)
						-0.026***
						(0.005)
						-0.023**
						(0.012)
	17 360	17 360	17 360	17 360	17 360	17 360
	0.173	0.174	0.180	0.183	0.190	0.192