DISCUSSION PAPER SERIES

No. 2011.1

July 2011

ETHNIC PATTERNS OF RETURNS TO EDUCATION IN BULGARIA: DO MINORITIES HAVE AN INCENTIVE TO INVEST IN EDUCATION?

■ **Claudia Trentini**

UNITED NATIONS

The UNECE series of Discussion Papers is intended to make available to a wider audience papers on matters of topical interest that have been prepared by the staff of the secretariat or commissioned by the secretariat from external experts. The purpose of the Discussion Papers is to contribute to the analysis and discussion of economic developments and policies in the UNECE region. The views expressed are those of the authors and do not necessarily reflect the official positions of the secretariat or of the member governments of the UNECE.

Ethnic patterns of returns to education in Bulgaria: Do minorities have an incentive to invest in education?

Claudia Trentini[1]

Abstract

It is widely accepted that disparities in education contribute to the poor labor market outcomes experienced by ethnic minority groups and consequently to their poverty. However, incentives to invest in education are significantly diminished if individuals are discriminated on the labor market and precluded from access to employment. In this paper we analyze differential educational benefits in Bulgaria and compare Roma returns to education with the majority population and the Turkish minority. We show that both ethnic minority groups have lower educational levels and employment rates than the majority population and that they also have lower returns to education. However, the gap in returns to education is much wider for the Roma with respect to both employment and labour-market earnings. The evidence suggests that this group is more vulnerable to discrimination, with a high percentage of the employment gap unexplained by differences in observable skills or characteristics.

---

[1] United Nations Economic Commission for Europe. The author has benefited from the comments and suggestions of Jaromir Cekota, Vitalija Gaucaite, Ralph Heinrich, Robert Shelburne and Lisa Warth. The usual disclaimer applies.

Introduction

The Roma are one of the largest and most disadvantaged ethnic minorities in the pan-European region. The Council of Europe average estimate of the Roma population in the European region is close to 11 million.[2] Of these about 6 million live in the territory of the European Union (EU), some 70 per cent of them in the 10 former communist countries that joined the EU in 2004 and 2007. The share of Roma in the population of post-communist EU countries averages 4 per cent, reaching approximately 10 per cent in Bulgaria.

The available evidence on the living conditions of this ethnic minority in Europe gives a dramatic picture of the situation: poverty and social exclusion are widespread, formal employment and education are extremely low, health is poor, housing is at best precarious (UNDP 2002, Ringold et al 2005, O'Higgins and Ivanov 2006). Moreover, not only education levels among Roma are worryingly low but also their vulnerability seems to be intensifying: in spite of the fact that younger Roma tend to have a somewhat higher educational achievement than older ones, the educational gap relative to majority groups of the same age has increased over time.

It is widely accepted that disparities in education contribute to the poor labor market outcomes experienced by ethnic minority groups and consequently to their poverty. At the same time occupational segregation or wage discrimination due to employer prejudice or stereotyping is considered a good reason for an individual to invest in relatively low education (Golbe 1985). Specifically, minority groups often enjoy lower benefits from education: not only on the labor market in terms of the probability of being employed and wages but also in terms of the quality of education obtained. This is the case when ethnic groups can only access ethnic segregated schools or poorly funded institutions in remote areas. Factoring in the higher costs of education related to inaccessibility of resources for lack of means or borrowing constraints that poor often face results in very low benefit-cost ratios for some minorities with respect to the majority population, leading to low educational investment. As a matter of fact, recent analyses of the Roma's poor labor market outcomes suggest that this group's incentives to invest in education could be very low.[3]

This paper analyzes differential educational benefits in Bulgaria and compare Roma returns to education with the majority population and the Turkish minority. The Turks are the largest minority in Bulgaria representing almost 10 per cent of the population according to the 2001 census (according to the same census data, Roma are only 6 per cent of the population). Like the Roma, the majority of Turks live in the countryside where they have much less access to infrastructure, work opportunities, and educational, cultural and health-care facilities than town-dwellers. In this paper it is shown that both ethnic groups have lower educational levels and employment rates than the majority population and that they also have lower returns to education. However, the gap in returns to education is much wider for the Roma with respect to

---

[2] This number includes 2.8 million Roma in Turkey and 1.2 million Roma in the former Soviet Union. Another 1 million Roma live in Western Balkans. For details, see the statistics link at http://www.coe.int/t/dc/files/themes/roma/default_en.asp

[3] About 70 per cent of Roma children drop out of education by middle school, largely because school holds little relevance to their lives (Gatenio Gabel 2009).

both employment and labour-market earnings. The evidence suggests that this group is more vulnerable to discrimination, with a high percentage of the employment gap unexplained by differences in observable skills or characteristics.

A quantitative analysis of the incentives to invest in education is important for developing sensible policies to integrate excluded minorities. This issue is particularly relevant in Bulgaria where the ethnic dimension of social polarization is quite dramatic: the share of the non-poor ethnic Bulgarians (56.13%) is 17 times bigger than that of the non–poor Turks (3.28%) and Roma (3.29%) (Kovacheva *et al*. 2005). Further, Bulgarian studies have pointed out how the ethnic pattern of poverty has been reproducing itself during the post-communist transition. The risk of poverty is highest among the long-term unemployed while low educational levels hamper employment even in the informal sector, perpetuating the negative cycle of poverty (Pamporov 2007).

Some studies focusing on Roma have claimed that their exclusion from mainstream society threatens Bulgaria's economic growth and political stability (World Bank 2010, UNDP 2002 and 2005).[4] This claim is based on long-term output, employment and demographic projections. Given the relatively high fertility rate of the Roma, their proportion in the general population has been increasing and is expected to keep rising over time (Ringold et al. (2005), Koytcheva and Philipov (2008)). A similar reasoning could be also applied to the Turkish minority in Bulgaria; however, the social exclusion of this minority appears to be less extreme than that of the Roma minority.

The paper is organized as follows. Section I provides information on the survey data used in our analysis as well as some descriptive information on the different population groups' educational achievement, living conditions, employment and earnings. Sections II, III, and IV report the results of estimations of the determinants of educational participation, employment and wages. Section V concludes.

I.   Data

We use data from the surveys collected in the framework of the Generations & Gender Programme (GGP). The programme is a system of national Generations and Gender Surveys (GGS) and contextual databases, which aims at improving the knowledge base for policy-making in participating European and Asian countries. The GGS is a panel survey of a nationally representative sample of the 18-79 year-old resident population in each participating country with at least three panel waves and an interval of three years between each wave. The contextual databases are designed to complement micro-level survey data with macro-level information on policies and aggregate indicators.

We use data from the first wave GGS for Bulgaria that were collected in 2004. While the main purpose of this survey is to investigate gender and family creation patterns, it can also be used to investigate labor market performance. Our sample comprises 9,192 ethnic Bulgarian (83.9%), 1,079 Turkish (9.8%) and 686 Roma

---

[4] See similar analyses for Hungary: Kertesi and Kézdi (2006) and (2010)

(6.3%) respondents aged 18 to 65 years and excludes other non-specified ethnic groups. The share of Roma in our sample is consistent with the 2001 national census data; however, it is below the above mentioned mid-range estimate of the Roma population by the Council of Europe (10.3 per cent).[5]

Survey participants self identify their own ethnic group. We identify individuals belonging to a minority as those who defined themselves as such or whose parents, one or both, belong to a minority group.[6]

The Annex table describes the sample characteristics. In many respects the Turkish community seems to be better integrated into the Bulgarian society than the Roma minority. Both Turkish and Roma populations are younger than the majority population – by 1.5 years and 6 years respectively. This reflects the different demographic patterns of the three population groups evident also from the number of children per adult and the number of siblings of the main respondent.[7] Gaps in education are quite wide: the average Bulgarian adult completed secondary schooling, while Turkish adults had a lower secondary level and the Roma only finished primary education. Both minorities' women have, on average, almost a year less of education than their male counterparts. A comparison of the respondents' level of education with that of their parents indicates a general increase in schooling years of about 2-3 years. Also, the observed similar level of partners' education is indicative of a very strong matching between couples.

Although basically all the interviewed individuals were born in Bulgaria, among minorities only a few respondents declared having Bulgarian as the native language (less than 3 per cent of the Turks and 8 to 12 per cent of the Roma) and only a few more use it as a language at home. The data on employment status in the Annex table reveal large gaps between ethnic groups: employment rates of Roma are roughly half of the majority population, while Turks are somewhere in between. The next row in the table reports the percentage of workers in low skilled occupations defined as plant and machine operators, assemblers and more generally elementary occupations. The statistics reveal a highly segmented labor market with minorities, especially Roma, concentrated in low skilled positions. A comparison with the parents' occupation (when the respondent was 15 years old) indicates that the situation of ethnic minorities has worsened, considering the improvement in the occupational ladder of the majority population and the absolute worsening of the labor market status of ethnic minority women. The next 2 rows of the Annex table show the percentage of respondents who lived with their biological parents at age 15 and lived in a village at that time.

---

[5] The interviews are done face to face in the main language of the country. This most probably negatively influenced the inclusion of Roma communities in the GGP surveys. Moreover, samples can exclude up to 5 per cent of the target population (UNECE 2005). Unfortunately, exclusions are due to frame limitations or practical constraints – such as eliminating remote regions where survey collection would be prohibitively expensive. These two survey limitations can bring about a partial exclusion of Roma from the survey and - what is even more worrying - they imply an exclusion of the most disadvantaged among them, i.e. those living in the most remote areas and/or having the lowest exposure to the majority population and to education. Thus the actual living conditions of the Bulgarian Roma may well be worse, on average, than those described in this paper.

[6] This could be another reason as to why the share of Roma in the sample is below other populations' estimates.

[7] Roma have a life expectancy of up to 10 years lower than the majority population (UNDP 2002).

The GGS statistics describe the different family and fertility patterns, with Roma being less likely to marry legally their partner but having on average more children and Turks being somewhere in-between. The low percentage of married Roma couples reflects the fact that traditional Romani marriages are not always registered;[8] this is evident from the high percentages of cohabiting partners. The number of respondent's siblings would suggest that fertility has been decreasing for minorities as well; however, one has to keep in mind that many respondents might still be in their fertile period. The next 3 rows of the table describe the labor market participation of the different ethnic groups, showing the percentage of self employed, those working on a part time basis and the monthly wage in Bulgarian leva. Roma workers stand out for working often part time, being very rarely self employed and earning only half as much as the majority population. Turks on the contrary are more likely to be self employed and basically suffer no wage gap.

The last three indicators in the Annex table refer to the health and living conditions. The table reports the percentage of individuals suffering from chronic illness or disability or who are reporting "very bad" health for two large age groups: 17 to 45 and 46 to 65. While for the younger population there are not big differences across ethnic groups, for the more mature age group differences in health conditions become considerable, suggesting differences in standards of living. This is also confirmed by the following statistics on the percentage of households having access to a flushing toilet, or living in dwellings with a leaking roof, damp or rot walls, or/and floors. These figures show the dramatic living conditions of Roma: less than a fifth of them have access to sanitation, and more than half live in poorly insulated dwellings. Again, figures for the Turkish minority place them between the Roma and majority population.

II.   Determinants of education

We first examine the determinants of years of education. The survey only provides information on the highest degree obtained. Using a well defined classification of educational levels, we convert the degrees obtained into completed years of education.[9] Years of schooling are estimated separately for ethnic groups and gender by OLS as a function of age and the family characteristics when the respondent was undergoing schooling: an indicator for Bulgarian being the native language, a dummy variable indicating if the child was living with both parents, the number of brothers and sisters, if the family was living in a village, parental education and occupation. This last variable captures the income situation during the period of respondent's education. Dummies for missing values are added. The results are shown in table 1. The coefficient on age captures the fact that older people have higher years of schooling. A quadratic term in age is included to allow the possibility of schooling decreasing in the later years of life. The two age coefficients imply that keeping the

---

[8] In some Romani communities arranged marriage, child marriage and forced marriage are still prevalent as „traditional practices". These traditional marriages often take the form of "custom law" marriages. However, Roma "custom law" marriages are not to be confused with the Anglo-Saxon "common law" marriages and are not recognized by the state as legally binding. "Custom" means that the couple is viewed as married by the community, relatives and their own but not in the eyes of the administration (UNDP 2002, European Commission 2009).

[9] International Standard Classification of Education (ISCED) level codes are converted into years according to the Bulgarian classification of educational levels. Thus reported years of education refer to successful years at school and do not consider repeated years or interrupted studies.

other factors constant educational levels increase with age until individuals reach their late 40's for men and the early 40's for women.

Having Bulgarian as the native language is positively correlated with schooling, improving the educational outcome by close to 1 year for girls. For minority males this indicator is not significant given the small share of the group having Bulgarian as native language. As expected parental education (education mother M and education father F) is strongly correlated to educational achievement, and in particular maternal years of schooling seem to have a more powerful impact on educational outcomes of minorities than on the majority population. This is not surprising, considering the lower probability of the Turkish and Roma mothers to be active on the labor market and thus the higher amount of time they spend with their children.

Table 1
The OLS regression results
Dependent variable: years of schooling.

	males			females		
	Majority	Turks	Roma	Majority	Turks	Roma
age	0.284	0.464	0.42	0.346	0.561	0.452
	[0.018]***	[0.075]***	[0.105]***	[0.017]***	[0.067]***	[0.087]***
age2	-0.288	-0.525	-0.442	-0.378	-0.702	-0.536
	[0.021]***	[0.090]***	[0.138]***	[0.020]***	[0.084]***	[0.111]***
lang Bulg.	0.348	0.03	-0.071	0.709	1.318	1.008
	[0.522]	[0.724]	[0.733]	[0.389]*	[0.718]*	[0.502]**
education M	0.15	0.25	0.259	0.175	0.257	0.263
	[0.014]***	[0.054]***	[0.075]***	[0.014]***	[0.049]***	[0.069]***
education F	0.118	0.131	0.235	0.146	0.126	0.263
	[0.014]***	[0.051]**	[0.069]***	[0.013]***	[0.047]***	[0.064]***
low occ F	-0.186	-0.565	-0.044	-0.245	-0.178	0.131
	[0.081]**	[0.302]*	[0.460]	[0.076]***	[0.283]	[0.385]
low occ M	-0.214	-0.224	-0.441	-0.152	-0.801	0.063
	[0.084]**	[0.305]	[0.450]	[0.079]*	[0.293]***	[0.374]
brothers	-0.417	-0.347	-0.425	-0.488	-0.472	-0.372
	[0.048]***	[0.122]***	[0.141]***	[0.047]***	[0.131]***	[0.135]***
sisters	-0.259	-0.157	-0.285	-0.279	-0.556	-0.335
	[0.048]***	[0.121]	[0.143]**	[0.044]***	[0.107]***	[0.110]***
with parents	0.257	0.079	-1.574	0.382	-0.217	-2.962
	[0.284]	[0.997]	[1.278]	[0.266]	[0.804]	[1.634]*
village at 15	-0.343	0.84	0.742	-0.46	0.471	0.007
	[0.083]***	[0.312]***	[0.400]*	[0.078]***	[0.304]	[0.338]
Constant	3.994	-1.128	-1.024	2.638	-1.567	-0.565
	[0.686]***	[1.771]	[2.328]	[0.575]***	[1.572]	[2.211]
Obs	3881	488	283	4795	549	384
R-squared	0.26	0.25	0.3	0.33	0.39	0.35
Standard errors in brackets: * significant at 10%; ** significant at 5%; *** significant at 1%						

Source: UNECE Generations and Gender Surveys.

The parental occupational level has a very different impact among ethnic groups and gender. Both maternal and paternal low skilled occupation indicators are strongly negatively correlated to the majority's boys and girls education, while they have no statistically significant impact on Roma. Most probably the indicators do not

provide valuable sources of variation for this group. Interestingly for Turks, paternal low skilled occupation is strongly negatively correlated to boys' outcomes while the same is true for maternal skills and girls' outcomes. This might reflect strong societal role models.

The presence of brothers seems to be more detrimental to schooling than the presence of sisters and this is particularly true for girls, who most probably need to help their mothers in taking care of the male siblings. Living in a village has a negative impact for the majority population while it is positively correlated to the educational achievement of minorities. This might depend on the different conditions minorities experience in rural versus urban settings. Particularly important might be the availability of "special" schools, where most pupils come from minorities. Residential segregation is common in Bulgaria, with ethnically homogeneous schools accommodating 70 per cent of Roma pupils in the country (ERRC 2004). As a matter of fact, Roma children are more likely to be retrieved from schools where they clearly represent only a minority. Special schools seem to be much more common in rural areas where the Roma and Turks concentrate. However, special schools tend to have substandard facilities and curricula and in general provide education of inferior quality so that the advantage of receiving more schooling in the countryside is not obvious.[10]

III.   Employment returns to education

In this section we analyze employment returns to education and their differential importance for the ethnic groups' labor market participation. We first show OLS regressions by groups and by gender of employment status on years of schooling, age, family characteristics, health status, the average (over the last 5 years) regional unemployment rates capturing the variation in provincial economic environment, and an indicator for those provinces where Roma presence is above the national average.[11] A linear probability model is preferred for the simpler interpretation of coefficients, especially for the analysis of the differential employment returns of education.[12] We estimate linear probability models of the following form separately for each group and gender:

$$E_i^* = \alpha_i + \beta S_i + \rho_E X_{Ei} + \varepsilon_i$$
with

---

[10] Teachers are not trained to teach students who may be of a different culture, which contributes to prejudice and ill-informed views as to the capacity and willingness of minorities children to learn, leading to low academic expectations (ERRC 2004).

[11] Given the objectives of the study, age rather than the more commonly included potential experience (age minus years of education) is used. An alternative is to use potential experience and instrument it with age. Using age directly provides more precise estimates albeit of a different parameter.

[12] Linear probability models cannot be correctly specified if there are unbounded right-hand-side variables. On the other hand, when right-hand-side variables are dummies that cover mutually exclusive categories, the model is saturated, and linear probability models are correctly specified and are, in fact, equivalent to probit and logit models. The model in fact includes almost only dummies. Moreover, the unconditional probabilities are in the middle range. Marginal effects of probit estimates are virtually identical to the coefficient shown in tables 2 and 3 and are available from the authors. In table 3 we take account of endogenous schooling and health conditions using a simple 2SLS. While a biprobit estimation is more efficient, 2SLS imposes less restrictive assumptions on the error structure while being consistent (Altonji, Elder and Taber (2005), Angrist (2001)).

$E = 1$ iff $E^* > 0$,

where E is the observed employment status of individual i, equal to 1 if employed and 0 otherwise, and E^* is the latent variable representing the 'tendency' to be employed, ρ and β denote the OLS coefficients, S the years of schooling, and X are the explanatory variables.

Table 2
OLS regression results
Dependent variable employment

	Males			**Females**		
	Majority	Turks	Roma	Majority	Turks	Roma
Yrs of schooling	0.028	0.028	0.016	0.03	0.03	0.017
	[0.003]***	[0.006]***	[0.007]**	[0.002]***	[0.005]***	[0.005]***
age	0.075	0.051	0.046	0.105	0.05	0.056
	[0.004]***	[0.012]***	[0.016]***	[0.003]***	[0.010]***	[0.010]***
age2	-0.097	-0.064	-0.062	-0.131	-0.059	-0.063
	[0.004]***	[0.014]***	[0.021]***	[0.004]***	[0.013]***	[0.013]***
Nr Children	-0.006	-0.046	-0.009	-0.015	-0.01	-0.034
	[0.009]	[0.019]**	[0.020]	[0.007]**	[0.019]	[0.014]**
Co-resident partner	0.116	0.131	0.038	-0.017	0.019	-0.01
	[0.019]***	[0.059]**	[0.075]	[0.014]	[0.050]	[0.052]
Poor health	-0.132	-0.247	0.006	-0.105	-0.084	-0.156
	[0.019]***	[0.052]***	[0.080]	[0.016]***	[0.052]	[0.056]***
Speak Bulgarian	0.21	0.067	0.057	-0.047	-0.004	0.008
	[0.110]*	[0.068]	[0.082]	[0.104]	[0.068]	[0.049]
rural	-0.063	-0.042	0.044	-0.052	-0.071	-0.042
	[0.017]***	[0.042]	[0.054]	[0.016]***	[0.041]*	[0.040]
Provincial Unempl	-0.007	-0.012	-0.004	-0.004	-0.01	-0.012
	[0.002]***	[0.005]**	[0.007]	[0.002]***	[0.005]**	[0.005]**
Roma concentration	0.008	0.042	-0.126	0.011	0.058	-0.054
	[0.015]	[0.053]	[0.068]*	[0.013]	[0.052]	[0.049]
Constant	-1.15	-0.551	-0.512	-1.539	-0.673	-0.679
	[0.131]***	[0.215]**	[0.265]*	[0.120]***	[0.176]***	[0.172]***
Observations	4118	511	295	5073	568	391
R-squared	0.24	0.18	0.1	0.29	0.14	0.17
Robust standard errors in brackets						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Source: UNECE Generations and Gender Surveys.

For Roma males the ability of the model to explain their employment probability is quite poor with a R-squared of only 10 per cent; moreover, almost no variable is statistically significant with the exceptions of age and education. For women on the contrary the number of children in the household and the health status are able to increase the fit of the model partly explaining the low labor market participation. This is quite in line with results in papers performing similar exercises like O'Higgings (2010) or Kertesi and Kezdi (2010). Geographic differences are likely to play an important role in explaining minorities' employment. This is because minorities tend to live in rural areas and remote villages, which are characterized by higher unemployment. Adding a full set of provincial dummy variables indeed raises the ability of the model to explain the variation of employment probability for Roma males as captured by the R-squared (going up to 19 per cent), however when one

accounts for the additional number of covariates used, the resulting R-squared (adjusted) is back to 9-10 per cent.

Nevertheless, the results of this regression are interesting per se, especially for what regards the schooling coefficient. While returns to education are relatively similar for the majority population and Turkish minority, they are about one third to over 40 percent lower for Roma. One additional year of schooling raises the probability of being employed by about 3 per cent for the first two groups while for the Roma this increases chances to be employed by less than 2 per cent. Differences in returns between genders are minimal. Usually the employment returns are higher for females; in Bulgaria this is true only for the majority population and Turkish minority.

The probability of employment rises with age although reaching a peak around 37 years for majority and Roma and around 40 years for Turkish men. The maximum talks longer to achieve for females ranging from 40 years for the majority to 44 years for Roma women. The number of children reduces the probability of employment more for women than men while the opposite is true for the presence of co-resident partners. Suffering from a disability or illness, as well as reporting poor health negatively influences the probability to be employed, except for Roma males. Speaking Bulgarian at home does not really have any significant effect on the employment chances especially for minorities. This can be due to the concentration of minorities in low skilled/manual occupations which do not require communication skills.[13] Given the insignificance of the language variable it will not be included in the next estimates to keep the model parsimonious. Provincial unemployment rates also negatively influence the probability to be employed for all groups examined with the exception of Roma males. We use averages of the 2000-9 period to avoid problems of selection, with workers moving to booming regions. However, looking at the dates when the households moved into the current accommodation, it is clear that the majority of the population is not moving in response of crises and that internal migration is not common.[14]

Assuming homogeneous returns to education across degrees, i.e. averaging over the academic career the impact of each year of school on the probability of being employed, leads to a quite unrealistic result that one year spent in primary school would have the same return as one year at university. Differences in returns to employment across degrees might well be very important, especially for minorities that have a relatively low educational participation and usually complete only primary or secondary education. To describe such differences, we show in the following table the same labor supply regression including instead of total years of education dummies for last degree obtained: primary (8 years), secondary (12 years), and tertiary (16 or more). The reference category is no education.

---

[13] ERRC (2006) reports that the jobs most Roma do are at the bottom end of the labor market; low-level menial jobs. However, only a very small number of Roma work in restaurant/hotel type work or in shops, which is surprising given that such occupations usually offer some unqualified opportunities for people at the lower end of the labor market. ERRC evidence suggests that Roma are excluded from employment involving contact with the public or with food.

[14] Using 2009 provincial unemployment rates is not affecting the results significantly.

Table 3
OLS regression results
Dependent variable employment

	Males			Females		
	Majority	Turks	Roma	Majority	Turks	Roma
Primary	-0.105	0.172	0.119	0.218	0.072	0.061
	[0.123]	[0.084]**	[0.061]*	[0.088]**	[0.065]	[0.042]
Secondary	0.078	0.261	0.203	0.365	0.259	0.212
	[0.123]	[0.087]***	[0.106]*	[0.087]***	[0.076]***	[0.083]**
Tertiary	0.15	0.7	-	0.478	0.538	-
	[0.123]	[0.092]***		[0.087]***	[0.097]***	
age	0.074	0.055	0.047	0.104	0.055	0.059
	[0.004]***	[0.011]***	[0.016]***	[0.003]***	[0.010]***	[0.010]***
age2	-0.095	-0.069	-0.061	-0.13	-0.067	-0.067
	[0.004]***	[0.013]***	[0.022]***	[0.004]***	[0.012]***	[0.013]***
Nr Children	-0.007	-0.049	-0.012	-0.015	-0.014	-0.037
	[0.009]	[0.019]**	[0.020]	[0.008]**	[0.019]	[0.014]***
Co-resid. partner	0.116	0.127	0.035	-0.017	0.023	-0.005
	[0.019]***	[0.059]**	[0.075]	[0.014]	[0.050]	[0.053]
Poor health	-0.131	-0.257	-0.016	-0.105	-0.085	-0.16
	[0.019]***	[0.052]***	[0.081]	[0.016]***	[0.051]*	[0.056]***
rural	-0.063	-0.042	0.057	-0.05	-0.048	-0.04
	[0.016]***	[0.043]	[0.054]	[0.016]***	[0.042]	[0.040]
Prov. Unempl	-0.007	-0.012	-0.005	-0.004	-0.009	-0.013
	[0.002]***	[0.005]**	[0.007]	[0.002]***	[0.005]*	[0.005]***
Roma concentr.	0.006	0.038	-0.13	0.011	0.048	-0.051
	[0.015]	[0.053]	[0.068]*	[0.013]	[0.053]	[0.048]
Constant	-0.623	-0.527	-0.503	-1.551	-0.641	-0.676
	[0.141]***	[0.218]**	[0.270]*	[0.110]***	[0.174]***	[0.173]***
Observations	4116	511	293	5068	568	390
R-squared	0.25	0.18	0.09	0.29	0.15	0.16
Robust standard errors in brackets						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Source: UNECE Generations and Gender Surveys.

From table 3 it is clear that for minorities, secondary education has an important positive impact on employment, raising the probability of being employed, in comparison to someone without education, by about 26 per cent for Turkish males and 20 per cent for Roma males. The marginal effect of primary education on the employment of the Roma may well reflect the substandard quality of their primary schooling. As a matter of fact, Roma children are often segregated into either Roma-only schools and classes or sent to special schools for children with developmental problems; both types of segregation have been shown to result in poor quality education and low performance on national tests (UNICEF 2010). The reasons which lead to this educational segregation are various and also include language barriers. However, Romani children often do not even have the chance to start at a regular/mainstream school. In Bulgaria the so-called Romani ghetto schools were established especially for Romani children and are located in or close to the segregated ghetto-like Romani neighborhoods. In addition to these, a number of schools, primarily schools located in villages, have become predominantly Romani or all-Romani due to demographic shifts in the past decade. Although the all-Romani ghetto schools follow standard curriculum, and are formally categorized as regular

10

schools, the material conditions and the quality of education in them are markedly inferior as compared to other mainstream schools attended primarily by non-Romani students (ERRC 2004). The inferior quality of education is likely to deter prospective employers from hiring Roma and thus could partly explain the insignificant return to primary education for minorities that are unable to access mainstream schooling.

For males from the majority population none of the three education degrees seems to have a positive and significant impact on employment; this is due to the negligible share of population with no education chosen as a base for comparison and to their high labor market participation at all educational levels. For females from the majority population on the contrary, employment returns to education seem to be very high and rising with the degree achieved, reaching almost 48 per cent for the college educated women.

We now look at the differential impact of education on employment for the different groups, interacting years of schooling with ethnic group. In doing this we take account of endogeneity of years of schooling and poor health conditions. Statistical endogeneity results from unobserved determinants of education that also influence employment. This typically occurs when education (and indeed all forms of human capital) is in part the result of constrained optimizing decisions made within the household. As such, the decision to attend school depends on market conditions and public services outside the household, as well as preferences and resource constraints within the household. To the extent that there exist unobservables that influence both education and the probability to be employed or the wage rate, the error terms will be correlated and OLS estimates of the coefficient on education will be biased by this correlation. Therefore, biased estimates of returns are unlikely to capture the effect of education on employment or earnings, an issue of significant concern to policymakers. The following analysis aims to correct for the estimation bias.

Household level characteristics observed when the respondent was 15 years old are used as instruments for completed education. These include the variables analyzed in the determinants of education: an indicator of household location: if it was in a village, and some variables reflecting the household size and composition: the number of siblings and if the child was still living with the parents at 15. Both mother's and father's completed educational levels are used. These can influence child schooling in a variety of ways. First, they serve as a proxy for permanent income (especially paternal education). Second, they may reflect parental preferences. Third, they may affect the education production process. Parental occupations, which are highly correlated with these two, have also been considered and do not change the results substantively. They are excluded in order to keep the analysis parsimonious. We also use partner education, which we found highly correlated among partners. In fact, recent work (Weiss, 1999 and Chiappori et al, 2009) suggest that marriage is subject to assortative mating even for the majority population while Stewart (2009) reviews recent evidence supporting the idea that assortative matching is a common practice for ethnic minorities. The validity of the instruments proposed depend on whether spousal or parental education does affect employment and wage rates. Spousal education or parental background might influence husband's/son's education through nepotism or networks effects. However, parental and spouse background

information has been adopted as instruments elsewhere, and, more importantly, its validity is ultimately an empirical question.

Investment in schooling is not the only form of human capital investment that individuals undertake. Health is arguably important in the decision to participate in the labor market and probably influences wages as well. Thus the omission of health may lead to biased estimates. To measure health we use an indicator which is equal to one in the presence of a disability or a long standing/chronic illness or if self reported health status is "very bad". As this indicator is only capturing severe health conditions it is arguably only influencing labor market participation decisions. As a consequence, we will control for it in the employment equation and we use it to control for selection in the wage equations.

As with schooling decisions, investments in health and their subsequent outcomes will also be endogenous. This is particularly likely for short term measures like the self-reported current health status. While it is possible that current health influences labor market participation or wages, there may also be a feedback effect whereby current earnings affect consumption patterns, nutritional intakes, and health of workers. Instruments for the health measure include indicators of sanitation, and housing quality, the number of disabled household's members, and provincial level indicators of the number of physicians per population, number of hospitals and provincial number of deaths for cardiovascular reasons.[15] Quality of housing and sanitation determine the hygienic conditions in which household members live influencing their health conditions.[16] They are supposed to influence employment and wages only indirectly through health. We then add the number of disabled household members, excluding the respondent, because genetic predispositions could run in the family. In fact, in the case of ethnic minorities in Bulgaria, endogamous marriage seems to be the norm (Pamparov 2007). This could account for a relatively high genetic predisposition to disease (Luba et al, 2005).

The model we are going to estimate is the following:

$$E_i^* = \alpha_i + \beta S_i + \rho_E X_{Ei} + \varepsilon_i$$

$$S_i = \delta_i + \rho_S X_{Si} + \upsilon_i$$

with

$$E = 1 \text{ iff } E^* > 0,$$

where the first line is the labor equation analyzed earlier. Common to both equations the set of covariates X_E includes age and age squared, number of children, an indicator showing whether a co-resident partner is available, the average unemployment rate in the district, an indicator individuating districts with a high proportion of Roma population, and ethnic dummies. X_S includes X_E and the instruments described above, measured when the individual was 15: an indicator

---

[15] Cardiovascular diseases are the leading cause of death for working age population in Bulgaria (data from Bulgarian National Statistical Institute).

[16] A significant factor contributing to bad health and a high early mortality rate is poor living conditions. According to the Bulgarian National Statistical Institute, almost half of the Roma population lacked potable water in their homes in 2001, and one fifth of Roma people live in homes where they have less than 4 sq. m. per capita. Lack of proper hygiene and overpopulation are the cause of infectious disease and epidemics (Fundación Secretariado Gitano (2009)).

showing whether the household was in a village, and some variables reflecting the household size and composition: the number of siblings and if the child was still living with the parents, and parental and partner's educational level.[17]

Table 4
OLS and 2SLS regression results
Dependent variable employment

	OLS				2SLS			
	males	females	males	females	males	females	males	females
Schooling	0.031	0.032	0.03	0.031	0.04	0.03	0.044	0.027
	[0.003]***	[0.002]***	[0.003]***	[0.002]***	[0.006]***	[0.005]***	[0.006]***	[0.005]***
T*school	-0.001	-0.01	-0.002	-0.01	0.013	-0.009	0.006	-0.011
	[0.006]	[0.005]*	[0.006]	[0.005]*	[0.012]	[0.009]	[0.012]	[0.009]
R*school	-0.016	-0.017	-0.016	-0.017	-0.028	-0.015	-0.033	-0.017
	[0.007]**	[0.006]***	[0.007]**	[0.006]***	[0.013]**	[0.010]	[0.013]**	[0.010]
Health			-0.133	-0.105			-0.18	-0.203
			[0.016]***	[0.014]***			[0.096]*	[0.092]**
Turk	-0.06	0.041	-0.051	0.041	-0.176	0.029	-0.088	0.04
	[0.066]	[0.055]	[0.065]	[0.055]	[0.130]	[0.094]	[0.131]	[0.093]
Roma	-0.07	0.004	-0.066	0	0.064	-0.02	0.135	-0.023
	[0.062]	[0.049]	[0.062]	[0.049]	[0.114]	[0.085]	[0.113]	[0.085]
age	0.073	0.099	0.072	0.099	0.069	0.1	0.067	0.099
	[0.003]***	[0.003]***	[0.003]***	[0.003]***	[0.004]***	[0.004]***	[0.004]***	[0.004]***
age2	-0.095	-0.125	-0.092	-0.122	-0.091	-0.126	-0.086	-0.122
	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.004]***	[0.005]***	[0.004]***
children	-0.017	-0.021	-0.018	-0.021	-0.013	-0.022	-0.014	-0.022
	[0.008]**	[0.006]***	[0.008]**	[0.006]***	[0.008]*	[0.007]***	[0.008]*	[0.007]***
partner	0.124	-0.007	0.12	-0.011	0.119	-0.006	0.113	-0.015
	[0.017]***	[0.014]	[0.017]***	[0.014]	[0.017]***	[0.014]	[0.017]***	[0.014]
rural	-0.049	-0.055	-0.048	-0.054	-0.037	-0.059	-0.03	-0.056
	[0.014]***	[0.014]***	[0.014]***	[0.013]***	[0.016]**	[0.015]***	[0.016]*	[0.015]***
Unempl	-0.007	-0.006	-0.007	-0.006	-0.007	-0.006	-0.007	-0.006
	[0.002]***	[0.001]***	[0.002]***	[0.001]***	[0.002]***	[0.001]***	[0.002]***	[0.001]***
H Roma	0.002	0.012	0.001	0.011	0.004	0.012	0.004	0.01
	[0.014]	[0.013]	[0.014]	[0.013]	[0.014]	[0.013]	[0.014]	[0.013]
Constant	-0.923	-1.497	-0.887	-1.475	-0.992	-1.479	-0.989	-1.44
	[0.068]***	[0.060]***	[0.067]***	[0.059]***	[0.083]***	[0.069]***	[0.084]***	[0.073]***
Obs	4924	6032	4924	6032	4924	6032	4922	6028
R-squared	0.24	0.29	0.25	0.3	0.23	0.29	0.044	0.027
First stage statistics					*Partial R2*	*Partial R2*	*Partial R2*	*Partial R2*
Schooling					0.2237	0.2586	0.2391	0.2735
T*school					0.2518	0.3595	0.2566	0.3625
R*school					0.3213	0.3532	0.3219	0.3563
Health							0.0316	0.033
					F(12,4898)	*F(12,6008)*	*F(12,4889)*	*F(12,5997)*
Schooling					117.58	174.62	80.85	118.81
T*school					137.35	281.07	88.83	179.47
R*school					193.21	273.38	122.12	174.72
Health							8.39	10.76

Source: UNECE Generations and Gender Surveys.

---

[17] Estimates of E which include the health status involve the estimation of a third equation for health. Instruments for the health status are explained in the text.

In table 4 it is evident that the difference in returns to years of schooling we found in the previous regressions is statistically significant for Roma. There are no statistically significant differences in educational returns for the Turkish minority. In OLS when the health indicator is added, coefficients on schooling decrease, suggesting that they are correlated with health conditions. Once we instrument education point estimates of schooling returns increase for the male majority population while the difference with the Roma minority widens. This is in line with the argument of Card (1999) and others that OLS estimates are biased downwards because individuals with high discount rates choose low levels of schooling, which have a higher marginal rate of return.

On the contrary for females 2SLS education coefficients decrease making the difference among ethnic groups less statistically significant. This would be more consistent with the omitted variable bias, namely ability. The coefficient for Roma females is relatively high and goes in the right direction but it is imprecisely estimated, with standard errors quite big. The coefficient on health increases in the 2SLS estimates. This can be due to weak instruments, and it is confirmed by low first stage statistics (both the partial R-squared and the F-test) for the health indicator.

Employment gap decomposition

We take a look at the relative importance of education and individual characteristics to explain employment gaps. We decompose employment gaps using the Blinder and Oaxaca methodology. The next table shows the results of this decomposition using the model estimated above and controlling for schooling endogeneity. In this case we run 2SLS regressions separately for each group and gender to decompose the gap between Turkish and Roma minorities and the majority population of the same gender.

The first row of the table reports the raw employment gap of the minorities' groups with respect to the majority population. The first set of variables and coefficients listed under the heading "Explained" measures the difference that is due to the different characteristics of the minority group with respect to the majority population (i.e. less education, younger, etc.). It is also called "endowment term." That means that if the regression coefficients in the minorities' samples were the same as the majority coefficients, the employment rate differentials would be equal to the figure shown in the row "total explained".

The unexplained part reflects differences in coefficients and is usually attributed to discrimination, and other unobserved differences for example in education quality, other skills and/or preferences. The most important element in the endowment term is due to differences to education. It accounts for more than 75 and 65 per cent of the employment gap for Turkish and Roma males respectively, and for almost 50 per cent for females. This is surprising considering that no indicator for education quality is used; presumably having that information would allow explaining even a bigger proportion of the employment gap. Age differences work slightly in favor of minorities, as they are younger. The presence of children negatively influences more females than males. Working in a rural environment or in an

economically depressed province also contributes to the gap.[18] It is evident that the unexplained gap is wider for Roma than for Turkish males, pointing to the possibility that Roma are more discriminated or more excluded. In the case of females, levels of discrimination are similar across ethnic groups.

Table 5
The Blinder-Oaxaca decomposition of employment gaps

	males		females	
	Turks	Roma	Turks	Roma
employment gap	**16.7**	**36.0**	**21.6**	**39.0**
Explained				
schooling	12.5	23.7	10.4	18.8
health conditions	0.4	0.2	0.2	-0.5
age	-2.2	-2.5	-1.2	1.5
Nr of children	0.2	0.4	0.9	1.5
Co-resid. partner	-1.4	-1.6	0.1	0.2
rural	1.8	1.5	2.2	2.0
Unempl rate	1.3	1.4	1.0	0.8
High Roma	-0.0	-0.3	-0.0	-0. 3
Total explained	**12.6**	**22.9**	**13.5**	**24.0**
percentage				
explained	**75.3**	**63.5**	**62.6**	**61.4**
Total unexplained	**4.1**	**13.1**	**8.1**	**15.1**

Source: UNECE Generations and Gender Surveys.

IV.   Earnings returns to education

In this sub-section we examine the earnings returns to education. Table 6 shows OLS regression estimates of log earnings on a series of variables influencing income: years of schooling, age, indicators for being self employed, working part-time, and in a public company, rural and high Roma province indicators, and the average provincial unemployment rate.

For males from both minorities returns to education are about 30 to more than 50 per cent lower, while minorities' females do not seem to have any significant return to education at all. Being self employed implies higher earnings for most groups, while working part time clearly reduces income. Majority population workers earn a lower salary in the countryside; on the contrary, minority workers achieve higher earnings in the countryside. This could suggest that in general earnings are lower in rural areas irrespective from the ethnic group.

The economic cycle negatively influences earnings of almost all groups with the exception of male Roma. Provinces with a high concentration of Roma also seem to be characterized by lower wages, at least for majority workers.

---

[18] Inserting a whole set of provincial dummies significantly increases the percentage of gap explained for Roma males to 67.2 per cent, slightly reduces it for Turkish males (73.3) and females (56.5) and leaves it roughly unchanged for Roma females. This indicates that geographic differences are an important factor to understand Roma males' employment gap.

Table 6
OLS regression results
Dependent variable log earnings

	Males			Females		
	Majority	Turks	Roma	Majority	Turks	Roma
Yrs school	0.067	0.046	0.032	0.066	0.012	0.017
	[0.006]***	[0.015]***	[0.015]**	[0.004]***	[0.013]	[0.014]
age	0.038	0.01	0.032	0.027	0.033	0.013
	[0.007]***	[0.025]	[0.026]	[0.006]***	[0.047]	[0.026]
age2	-0.049	-0.011	-0.062	-0.031	-0.045	-0.028
	[0.009]***	[0.031]	[0.036]*	[0.008]***	[0.069]	[0.035]
Self-empl.	0.247	0.494	0.49	0.164	0.974	0
	[0.043]***	[0.185]***	[0.420]	[0.038]***	[0.255]***	[0.150]
Part-time	-0.224	-0.226	-0.451	-0.331	-0.201	-0.243
	[0.049]***	[0.189]	[0.177]**	[0.053]***	[0.174]	[0.086]***
public	0.099	-0.131	0.071	-0.015	0.083	-0.088
	[0.027]***	[0.106]	[0.117]	[0.020]	[0.083]	[0.089]
rural	-0.179	0.194	0.228	-0.099	0.005	-0.029
	[0.028]***	[0.095]**	[0.115]*	[0.026]***	[0.085]	[0.081]
Unempl	-0.023	-0.04	-0.017	-0.019	-0.036	-0.027
	[0.004]***	[0.013]***	[0.011]	[0.003]***	[0.010]***	[0.010]***
High Roma	-0.067	0.028	-0.046	-0.081	0.161	0.156
	[0.025]***	[0.148]	[0.130]	[0.019]***	[0.125]	[0.104]
Constant	3.502	4.129	3.775	3.399	3.914	4.203
	[0.144]***	[0.486]***	[0.463]***	[0.129]***	[0.731]***	[0.452]***
Obs	2336	271	104	2795	216	93
R-squared	0.18	0.15	0.22	0.2	0.24	0.24
Robust standard errors in brackets						
* significant at 10%; ** significant at 5%; *** significant at 1%						

Source: UNECE Generations and Gender Surveys.

Again we analyze differentials in returns to education. Like in the previous analysis schooling is considered endogenous. Moreover, as usual in earnings equations it is necessary to correct for selection bias and this is particularly important in this context because of minorities' low employment rates. The selection bias arises from the fact that for unemployed individuals there is no information on earnings, and thus the corresponding observations cannot be used when estimating the earnings equation. As a consequence the analysis is limited to a non-random sample constituted only by those who report positive earnings. As it is likely that those factors that influence the employment probability also influence the earnings equation, we estimate the probability of being employed on the whole sample and correct the earnings equation contemporaneously. We control for selection bias using as exclusion restrictions the number of children, married status and health status. While the first two variables might influence labor market participation decisions especially of women, the last one is a good instrument for both genders.

In table 7 we report in the following order separately for males and females: OLS results, then results controlling only for the selection bias, 2SLS controlling for the endogeneity of schooling and then results controlling both for the schooling endogeneity and selection bias.

Table 7
OLS, 2SLS and LIML regression results

	OLS		2SLS selection		2SLS schooling		LIML	
	males	fem.	males	fem.	males	fem.	males	fem.
Schooling	0.072	0.067	0.068	0.065	0.123	0.109	0.128	0.107
	[0.006]***	[0.004]***	[0.006]***	[0.004]***	[0.013]***	[0.010]***	[0.013]***	[0.009]***
T*school	-0.042	-0.058	-0.052	-0.066	0.006	-0.089	-0.046	-0.097
	[0.019]**	[0.015]***	[0.014]***	[0.011]***	[0.030]	[0.021]***	[0.026]*	[0.017]***
R*school	-0.049	-0.048	-0.052	-0.054	-0.068	-0.076	-0.088	-0.084
	[0.014]***	[0.012]***	[0.018]***	[0.015]***	[0.032]**	[0.026]***	[0.033]***	[0.023]***
Turk	0.476	0.702	0.575	0.818	0.067	1.141	0.665	1.258
	[0.208]**	[0.190]***	[0.154]***	[0.126]***	[0.331]	[0.241]***	[0.283]**	[0.188]***
Roma	0.307	0.365	0.449	0.508	0.707	0.836	1.01	0.99
	[0.127]**	[0.122]***	[0.161]***	[0.131]***	[0.282]**	[0.229]***	[0.286]***	[0.188]***
age	0.031	0.027	-0.002	-0.003	0.023	0.022	-0.015	-0.015
	[0.007]***	[0.006]***	[0.008]	[0.008]	[0.007]***	[0.006]***	[0.008]*	[0.008]*
age2	-0.041	-0.032	0.002	0.006	-0.031	-0.024	0.017	0.022
	[0.008]***	[0.008]***	[0.010]	[0.010]	[0.009]***	[0.008]***	[0.010]*	[0.010]**
Self- empl	0.28	0.233	0.276	0.215	0.263	0.208	0.244	0.186
	[0.045]***	[0.041]***	[0.040]***	[0.031]***	[0.039]***	[0.031]***	[0.041]***	[0.032]***
Part time	-0.223	-0.284	-0.342	-0.365	-0.187	-0.261	-0.315	-0.361
	[0.050]***	[0.050]***	[0.049]***	[0.040]***	[0.046]***	[0.037]***	[0.050]***	[0.040]***
public	0.078	-0.013	-0.073	-0.117	0.032	-0.068	-0.13	-0.186
	[0.026]***	[0.019]	[0.034]**	[0.028]***	[0.031]	[0.024]***	[0.036]***	[0.031]***
rural	-0.105	-0.075	-0.099	-0.048	-0.032	-0.015	-0.024	0.014
	[0.027]***	[0.024]***	[0.029]***	[0.024]**	[0.031]	[0.027]	[0.033]	[0.027]
unempl	-0.025	-0.021	-0.022	-0.018	-0.026	-0.02	-0.022	-0.017
	[0.003]***	[0.003]***	[0.003]***	[0.002]***	[0.003]***	[0.002]***	[0.003]***	[0.002]***
H Roma	-0.068	-0.063	-0.087	-0.077	-0.054	-0.063	-0.069	-0.075
	[0.025]***	[0.019]***	[0.028]***	[0.021]***	[0.027]**	[0.021]***	[0.028]**	[0.021]***
Constant	3.56	3.384	4.478	4.115	3.025	2.878	3.938	3.763
	[0.138]***	[0.127]***	[0.174]***	[0.178]***	[0.199]***	[0.168]***	[0.209]***	[0.179]***
Obs	2709	3103	4924	6032	2711	3104	4924	6032
R-squared	0.19	0.21			0.14	0.18		
First stage statistics					Partial R2	Partial R2		
Schooling					0.2161	0.205		
T*school					0.1741	0.2795		
R*school					0.3431	0.3322		
					F(12,2686)	F(12,3080)		
Schooling					61.71	66.19		
T*school					47.17	99.56		
R*school					116.93	127.67		

Source: UNECE Generations and Gender Surveys.

To control both for the endogeneity of schooling and for the selection bias, we basically estimate a five equations recursive model with maximum likelihood (LIML): a normal wage equation, three schooling equations (one for each group) and one selection equation.

$$\ln w_i = \alpha_i + \beta S_i + \rho_w X_{wi} + \nu_i$$
$$S_i = \delta_i + \rho_S X_{Si} + \upsilon_i$$
$$E_i^* = \alpha_i + \beta S_i + \rho_E X_{Ei} + \varepsilon_i$$

with

$$E = 1 \text{ iff } E^* > 0,$$

where the second and third equations have been analyzed in the previous section. The second equation is estimated for each ethnic group. The first equation is the wage equation where ln w is the natural logarithm of earnings of individual *i, S* is the years of schooling, and the covariates X_w include age and age squared, dummies for workers self-employed, working in the public sector and part-time, the usual regional indicators.

Controlling for the endogeneity of schooling raises returns to education for all groups but more so for the majority, increasing the differential with minority groups, while controlling for selection reduces returns to education in some cases dramatically as for Turkish males. Results in last two columns show that returns to education are significantly lower for Roma (both males and females) and for Turkish women. Returns for Turkish men are lower but the difference with the majority population is only marginally significant.

This is in contrast to O'Higgins (2010) results that wage returns to education for Roma and non-Roma are relatively similar. On the contrary, our results corroborate the idea that lower educational attainment by minorities is at least partly motivated by lower perceived returns both in terms of earnings and in terms of employment.

V.   Conclusions

This paper analyzes the factors determining employment and wages for ethnic minorities in Bulgaria. It shows that in spite of the fact that individuals belonging to ethnic minorities and who could achieve secondary schooling have a probability of 20-25 per cent higher than illiterates to find gainful employment, minorities have very low incentives to invest in education given the lower returns - with respect to the majority population - in terms of prospective employment and wages in the labor market.

The relatively low returns to education for disadvantaged minorities in general and the Roma minority in particular, should be considered by the policy makers designing policies that aim to break the vicious cycle: poor education, poor labor market outcomes, poverty, welfare dependency. Given existing fertility patterns, it is likely that minorities will make up a growing share of the population over time. Integrating them into the workforce could boost growth through heightened productivity, new skills, increased consumption and investment. Moreover, lower requirements for welfare spending would enable the government to invest in infrastructure and raise productive capacity (UNDP 2005, World Bank 2010).

The analysis of the situation of disadvantaged ethnic minorities on the labor market, based on the initial generations and gender survey of Bulgaria, implies that forward-looking policies are needed for their social and economic integration in a number of areas, including education, health, housing, labor market and social security. Among these the issue of gainful employment constitutes an important precondition for successful social integration.

The high unemployment rate of minorities is most frequently considered to be a consequence of supply-side factors such as the low or absent educational qualifications; lack of relevant work skills, and their concentration in economically depressed regions. Results in this paper support this view suggesting that improving education levels of minorities could help reducing a big part of the employment gap. However, for Roma apart of education differences, and family structure for females, few other observable characteristics can explain the employment differential. The high percentage of unexplained employment gap could be attributed not only to unobserved differences in education quality, other skills and/or preferences but also - as claimed in many other studies (ERRC 2007, UNDP 2002 and 2005, Ringold et al. 2006) - to discrimination.

Against this background, labor market policies will thus need to improve employability and labor market competence of vulnerable groups through active policies and also tackle discrimination with well designed affirmative actions.[19] Given the important share of the working age population excluded by the labor market it is particularly important to involve private businesses in this exercise without relying exclusively on public work programmes. In addition, the desegregation of public education and decisive improvement of its quality in low-income districts are needed in order to reduce racial prejudice and upgrade the education and skills of Roma and Turkish children and youth with a view to improving their employability in the formal sector of the economy. This would in the longer term also improve the sustainability of the Bulgarian social-security system as a whole.

---

[19]Recent evidence shows that to tackle discrimination a mixture of approaches is needed including anti-discrimination legislation**,** proactive equality policies and positive actions (ERRC 2007).

References

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. (2005) "An Evaluation Of Instrumental Variables Strategies For Estimating The Effects Of Catholic Schooling," *Journal of Human Resources*, 2005, v40(4,Fall), 791-821

Angrist, Joshua D, (2001). "Estimations of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business & Economic Statistics*, American Statistical Association, vol. 19(1), pages 2-16, January.

Chiappori, Pierre-André, Iyigun, Murat and Weiss, Yoram (2009) "Investment in Schooling and the Marriage Market"*, The American Economic Review*, Volume 99, Number 5, pp. 1689-1713

European Commission (2009), *Ethnic minority and Roma women in Europe: A case for gender equality?* Brussels, EGGSI — Expert group on gender equality, social inclusion, health and long-term care.

ERRC (2004), *Stigmata: Segregated schooling of Roma in Central and Eastern Europe*, European Roma Rights Centre, Budapest.

ERRC (2007), *The Glass Box: Exclusion of Roma from Employment,* European Roma Rights Centre, Budapest.

Fundación Secretariado Gitano (2009), *Health and the Roma Community, analysis of the situation in Europe: Bulgaria, Czech Republic, Greece, Portugal, Romania, Slovakia, Spain*, Madrid

Gatenio Gabel, Shirley (2009), "The growing divide: the marginalisation of young Roma children in Bulgaria", *International Journal of Social Welfare*, Volume 18, pages 65–75

Golbe, D. (1985), "Imperfect signalling, affirmative action, and black-white wage differentials", *Southern Economic Journal* 51(4): 842-848.

Kertesi, Gábor and Kézdi, Gábor (2006), "Expected long-term budgetary benefits to Roma education in Hungary," *Budapest Working Papers on the Labour Market* No. 5.

Kertesi, Gábor and Kézdi, Gábor (2010), "Roma Employment in Hungary After the Post-Communist Transition". *The Economics of Transition,* forthcoming.

Kovacheva Siyka, Radka Peeva and Tsvetan Andreev (2006), *National Report Bulgaria Socio-economic trends and welfare policies*, Universiteit Utrecht

Koytcheva, Elena and Dimiter Philipov (2008), **"**Bulgaria: Ethnic differentials in rapidly declining fertility", *Demographic Research,* Volume 19 (13), pp 361-402

Luba Kalaydjieva, Bharti Morar, Raphaelle Chaix, and Hua Tang (2005), "A newly discovered founder population: the Roma/Gypsies"*, BioEssays* vol 27, pp 1084–1094

Mintchev, Vesselin, Venelin Boshnakov and Alexander Naydenov (2011), "Sources of Income Inequality: Empirical Evidence from Bulgaria", The wiiw Balkan Observatory Working Paper 89

O'Higgins, Niall and Ivanov, Andrey (2006), "Education and Employment Opportunities for the Roma", *Comparative Economic Studies*, vol. 48 Pages 6-19.

O'Higgins, Niall (2010), "'It's not that I'm a racist, it's that they are Roma': Roma Discrimination and Returns to Education in South Eastern Europe," *International Journal of Manpower*, Emerald Group Publishing, vol. 31(2), pages 163-187, May.

Pamporov, Alexey (2007), "Social exclusion of Roma in Bulgaria", paper presented at The Fourth International Conference of the EAPS Working Group "Second Demographic Transition" Social Exclusion And The Changing Demographic Portrait Of Europe, Budapest.

Ringold, Dena, Orenstein, A. Mitchell and Wilkens Erika (2005), *Roma in an Expanding Europe. Breaking the Poverty circle*, World Bank, Washington

Stewart, Frances (2009), "Horizontal Inequality: Two Types of Trap," *Journal of Human Development and Capabilities*, 10: 3, pp. 315 — 340

UNDP (2002), *The Roma in Central and Eastern Europe: Avoiding the Dependency Trap,* Regional Human Development Report. UNDP: Bratislava.

UNDP and Ernst & Young (2005), *Employing the Roma Insights from Business*. UNDP: Bratislava.

UNECE (2005), *Generations & Gender Programme: Survey Instruments*. New York and Geneva.

UNICEF (2010), *Towards Roma Inclusion A Review of Roma Education Initiatives in Central and South-Eastern Europe*. Geneva.

World Bank (2010), *Economic Costs of Roma Exclusion*, Europe and Central Asia Human Development Department.

Annex 1: GGP data

We use data from the surveys collected in the framework of the *Generations & Gender Programme* (GGP). The GGP is made of two major components, totally independent from each other at the data gathering level, but that could be interactive at the statistical analysis level: the Generations & Gender Survey (GGS) and the Contextual Database (CDB). The GGS consists of a panel survey of three waves (three years apart) in which 10 000 individuals aged from 18 to 80 are followed. The CDB on the other hand, relates to more than 200 variables, of national and/or regional level, sometimes qualitative but more frequently quantitative (time series from 1970 up to present in most cases), related to a wide range of topics: health, economy, employment, culture, education, demography, pensions, etc.

While for some topics, as for example welfare state provisions and educational systems, we referred to the CDB, our main source of data is the GGS. The surveys are nationally representative surveys which ensure international comparability of data at least for the core questionnaire. Some of the modules are optional, e.g. housing, ethnicity and nationality, previous partners, intention to break-up, preventing some cross country comparisons. The first wave was conducted in 2004 for Bulgaria. The GGS questionnaire covers a wide range of topics related to the household and the relations among genders and generations. Main respondents can be either men or women aged between 18 and 80.

Annex Table 1

	Majority		Turkish minority		Roma	
	m	f	m	f	m	f
Obs	**4119**	**5073**	**511**	**568**	**295**	**391**
percentage	**83.9%**		**9.85%**		**6.26%**	
age	**39.8**	**38.6**	**37.6**	**37.0**	**33.5**	**33.8**
	13.5	12.8	12.5	12.0	11.3	11.4
education	**12.9**	**13.6**	**9.8**	**9.3**	**6.9**	**5.9**
	2.58	2.87	3.47	3.86	3.78	3.87
education father	**10.6**	**11.0**	**6.3**	**6.3**	**4.1**	**4. 1**
	4.08	4.10	4.48	4.44	3.81	4.09
education mother	**10.5**	**11.0**	**5.1**	**5.7**	**3.5**	**3.4**
	4.27	4.15	4.24	4.34	3.47	3.74
education partner	**13.3**	**13.3**	**9.1**	**10.1**	**6.4**	**7.0**
	2.96	2.69	3.66	3.44	3.74	3.90
Bulgarian native language	**0.991**	**0.990**	**0.041**	**0.035**	**0.082**	**0.123**
	0.093	0.098	0.199	0.184	0.274	0.329
Bulgarian used at home	**0.993**	**0.995**	**0.092**	**0.099**	**0.163**	**0.254**
	0.080	0.067	0.289	0.298	0.370	0.436
employed	**0.670**	**0.612**	**0.505**	**0.398**	**0.313**	**0.224**
	0.470	0.488	0.500	0.489	0.460	0.415
Low skilled occupation	**0.298**	**0.177**	**0.435**	**0.353**	**0.620**	**0.595**
	0.458	0.382	0.497	0.479	0.488	0.493
Low skilled occ father	**0.380**	**0.368**	**0.515**	**0.501**	**0.578**	**0.557**
	0.485	0.482	0.500	0.500	0.495	0.497
Missing info (occ father)	**0.056**	**0.061**	**0.041**	**0.057**	**0.058**	**0.077**

	0.229	0.240	0.199	0.232	0.234	0.267
Low skilled occ mother	**0.316**	**0.306**	**0.395**	**0.395**	**0.486**	**0.472**
	0.465	0.461	0.489	0.489	0.501	0.500
Missing info (occ mother)	**0.015**	**0.014**	**0.012**	**0.022**	**0.020**	**0.021**
	0.122	0.116	0.111	0.147	0.142	0.142
with parents at 15	**0.929**	**0.923**	**0.940**	**0.915**	**0.918**	**0.913**
	0.258	0.267	0.238	0.278	0.274	0.283
village at 15	**0.370**	**0.335**	**0.698**	**0.704**	**0.548**	**0.550**
	0.483	0.472	0.459	0.457	0.499	0.498
rural area	**0.239**	**0.198**	**0.630**	**0.599**	**0.554**	**0.540**
	0.427	0.400	0.483	0.490	0.498	0.499
Number of children	**0.754**	**1.01**	**1.21**	**1.53**	**1.66**	**1.93**
	0.912	0.924	1.198	1.196	1.621	1.590
Married	**0.569**	**0.621**	**0.587**	**0.614**	**0.405**	**0.442**
	0.495	0.485	0.493	0.488	0.493	0.496
Co-resident partner	**0.623**	**0.678**	**0.749**	**0.748**	**0.766**	**0.775**
	0.485	0.467	0.434	0.434	0.424	0.418
house owner	**0.385**	**0.395**	**0.308**	**0.303**	**0.177**	**0.177**
	0.487	0.489	0.462	0.460	0.382	0.382
Number of siblings	**1.130**	**1.151**	**2.291**	**2.368**	**3.112**	**3.347**
	1.078	1.089	1.865	1.783	2.057	2.072
Self-employed	**0.155**	**0.130**	**0.211**	**0.200**	**0.109**	**0.092**
	0.349	0.326	0.383	0.383	0.290	0.288
part time	**0.079**	**0.064**	**0.130**	**0.140**	**0.250**	**0.230**
	0.260	0.236	0.311	0.329	0.403	0.404
Earnings	**151.8**	**119.0**	**144.5**	**128.8**	**82.5**	**67.6**
	267.3	110.8	253.6	239.9	78.6	56.2
Poor Health (age 17 -45)	**0.104**	**0.128**	**0.128**	**0.139**	**0.126**	**0.124**
	0.305	0.335	0.334	0.346	0.333	0.330
Poor Health (age 46 -65)	**0.331**	**0.459**	**0.415**	**0.538**	**0.543**	**0.586**
	0.474	0.499	0.495	0.498	0.504	0.487
Access to toilet	**0.848**	**0.878**	**0.421**	**0.394**	**0.184**	**0.219**
	0.359	0.328	0.494	0.489	0.388	0.414
Living in humid dwelling	**0.188**	**0.202**	**0.302**	**0.267**	**0.510**	**0.545**
	0.391	0.402	0.459	0.443	0.501	0.499
Nr disabled HH members	**0.094**	**0.099**	**0.194**	**0.125**	**0.197**	**0.146**
	0.343	0.393	0.467	0.420	0.460	0.400