

Methodology annex

The survey questionnaire that was used to generate the data on which this report is based follows the philosophy of integrated household surveys, with separate components containing both household and individual modules. Within the individual module, each household member's profile was registered (demographic characteristics, economic status, education, health). The household module addresses issues related to the household in general (dwelling type, access to basic infrastructures, household items etc.). Questions related to incomes and expenditures were addressed in both modules, making it possible to crosscheck the results.

The primary universe under study consists of: (i) all the households in Roma settlements or areas of compact Roma population; (ii) displaced persons (IDPs/refugees); and (iii) non-Roma communities living in close proximity to Roma and the displaced. While Roma, refugees and IDPs are not Southeast Europe's only vulnerable groups, they are definitely among the most vulnerable.

The vulnerable group samples

The sampling of vulnerable groups in general and of Roma in particular is a major challenge in every survey targeting diversities and vulnerability. The first assumption of the survey was that major disparities in socio-economic status of the populations are most obvious (and can be explored best) at the level of municipality (or other relevant micro-territorial units). Since at this level vulnerability factors exist that affect both Roma and other communities, vulnerability profiles of the two groups (Roma and majority) in the same municipality were developed, in order to make possible the identification of those vulnerability factors that affect Roma.

The most difficult question in this regard is "Who is Roma?" and how to appropriately identify the survey respondents. The primary objective of the survey was to map the vulnerability of groups with common socio-

economic, cultural and linguistic patterns – irrespective of how potential survey respondents might identify themselves. Since Roma identity is often associated with underclass status and discrimination, the decision to avoid self-identification as 'Roma' is not infrequent. Simply asking potential survey respondents "Are you Roma?" is therefore unlikely to yield unbiased survey data. These issues are further complicated by the multiple ethnic identities that are commonly found in Southeast Europe (and not only among Roma). The question "Are you Roma?" implicitly suggests its antithesis ("You are not Bulgarian, Romanian, Macedonian etc.?"). This extensive possible confusion between 'ethnicity', 'nationality' and 'citizenship' further argues against relying solely on self-identification. In most countries, therefore, Roma are underreported in censuses, and officially registered sizes of Roma populations often differ dramatically from experts' estimates.

While accepting the belief that censuses understate the absolute numbers of Roma, the survey accepted that the census data provide reasonably adequate pictures of the structure and territorial distribution of those individuals who identify themselves as Roma. Since the absolute number of Roma populations is not known, random sampling was not possible, so a 'pyramid' sampling model was used instead. Within this model, various estimates of Roma population (including census data) constitute different tiers of the pyramid. The bottom of the pyramid constitutes the total ('real') number of Roma in a country. The top represents the hypothetical situation of total exclusion in which not a single person would self-identify as Roma. Census data constitute one of the pyramid's tiers, with the pyramid's strata reflecting the structure of the population. Under this model, if the 'propensity to underreport' (i.e., the share of Roma not willing to identify themselves as Roma) is distributed similarly in different regions within a country, the structure of the population reflected in the census tier would be identical to the structure of the total population. This should be

sufficiently unbiased to construct a representative stratified sample.

In practical terms, it was assumed that the propensity to underreport was identical for each region within an individual country. Based on this assumption, the Roma sample was taken as representative of the Roma population living in 'Roma settlements or areas of compact Roma population'. Those settlements and areas were defined as settlements where the share of Roma population equals or is higher than the national share of Roma population in the given country, as reflected in the census data. The share – not the absolute number – of Roma was used for identification of the sampling clusters. The knowledge that X per cent of Roma (as reported in the census) live in settlement Y was taken to mean that X per cent of the sample will be derived from settlement Y. In this way, the demographic structure of the sample reflects the demographic structure of the Roma population (as reflected in the census data in proportions).

At the first stage of the sample design the universe was defined as mentioned above, using 'average and above share of Roma in each settlement'. In the second phase, taking into consideration also Roma organizations' estimates of Roma populations, the distribution of the settlements and population sizes, sampling clusters were determined. Respondents were then identified using 'random route' selection processes (third stage).¹²⁸

Internal (self-identification) and external (outsider's identification) modes therefore prevail at different stages of the sampling process. Self-identification (reported during the census) was used in the first stage; external identification (assessment of local people, NGOs, experts) was employed in the second stage. In the third stage (respondents' selection), the results of the first two stages were confirmed or rejected by 'implicit endorsement of identification'. In practice this meant that having identified the sample clusters and the households to be interviewed, the introductory sentence at the beginning of the interview was "Good morning/day, we are conducting a survey among the Roma population. Would you like to be interviewed?" In case of explicit denial ("I am not Roma, why should you in-

terview me?") the interview was cancelled. Willingness to participate in the interview was interpreted as the household member's implicit endorsement of belonging to the universe under study.

In some cases (particularly in big cities and capitals), large Roma communities constitute relatively small shares of total populations. In such cases, the sampling methodology conformed to administrative subdivisions (usually the 'capital municipality' is divided into smaller municipalities and/or lower levels of self-government). These lower levels were then chosen as the sampling units. Such cases were also corrected typologically, introducing additional sampling points.

A similar approach was applied to refugees/displaced persons with the only difference that instead of census data for the first stage of the sampling design, official registries and data on refugees/displaced persons' distribution provided by relevant institutions dealing with displaced populations were used to outline the universe under study. In the second phase, based on these lists, the sampling clusters were determined through random sampling. At the third phase individual respondents were identified using 'random route' selection processes (third stage).

Control samples

In order to derive data for meaningful comparisons that would respond to the data needs of an area-based development approach, a control sample of populations that are not defined as 'vulnerable' in the context of this report (i.e. that are neither Roma, nor refugees/displaced) was constructed. Given the fact that the ethnic affiliation of those populations is diverse (in some cases they are a minority at the national level but constitute a local majority), their exact definition would be 'non-Roma and non-displaced persons living in close proximity to the two vulnerable samples'. The control groups' samples were constructed using similar procedures as for the two vulnerable groups. In the case of Roma, those are representative samples of non-Roma communities living in settlements with Roma communities of 'average and above' size. In the case of the displaced sample, the control

¹²⁸ To a certain extent the sampling process is similar to Leslie Kish's cluster sampling model.

group is non-displaced populations living in proximity. In the second stage of sampling (determining the size of the population and the sampling clusters), external identification was used to identify the 'proximity populations' (assessment of local people, local self-governments). In the third stage random route selection was also applied to select the individual households.

In cases of municipalities with a high share of Roma and the number of majority population not sufficient for creating a majority sample (for example, in cases of isolated Roma settlements or segregated neighbourhoods), a majority sample was based on a typologically similar settlement in the same district (administrative unit) with a Roma population equal to or higher than the national average. The criterion for choosing this settlement was that it be the 'closest village accessible by road connection'.

The desire to obtain comparable data for non-Roma and non-displaced populations living in close proximity to the two vulnerable groups surveyed reflected a major emphasis of the current analysis: its area-based development focus. The majority samples gave the survey the 'benchmark' needed for assessments of the depth of Roma and displaced persons' poverty and vulnerability vis-à-vis the control groups (non-Roma and non-displaced) living in similar socio-economic environments and sharing some of the challenges the two vulnerable groups are facing. Despite the sample design challenges it poses, this approach allows us to distinguish among various vulnerability factors, particularly those that are related to minority status (and hence can be attributed to various forms of discrimination), as opposed to manifestations of regional development disparities or depressed local economic circumstances. It also provides clues on how to tackle the issues of exclusion and marginalization. Although often determined by institutional factors and policies, exclusion occurs at the level of interaction. This is primarily the level of the community, where people have daily contact. Measuring the distance between Roma and non-Roma in areas they cohabit could be an important clue on how to tackle challenges of social distance.

It is important to bear in mind that this approach does not attempt to guarantee national representativeness for majority communities. Because they share similar so-

cio-economic circumstances, those populations may also be facing some vulnerability risks and thus also may be – but not necessarily are – vulnerable. If they are, the status of these populations would be worse than national averages – as is the case on some indicators compared to national averages. Whenever national indicators are available these are used as a benchmark to assess the vulnerability of the three groups covered in the survey.

Methodological costs and benefits of the Roma sample

The samples based on municipalities with average and above shares of Roma population are not fully representative for the entire Roma populations of the countries covered in this survey. They do, however, cover roughly 85 per cent of Roma in each country, and as such provide a good basis for developing quantitative socio-economic indicators of Roma welfare (quality of life, life expectancy, access to services, incomes etc.). The resulting samples are representative not just for residents of segregated Roma communities, but also for the majority of Roma.

The data generated by these samples are broadly consistent with census data, since this survey's data are based on relative numbers (structure and regional distribution) instead of absolute numbers of Roma registered in the censuses. This approach also gives some standardized criterion for majority sample selection. The major drawback of this sampling methodology is related to its application to municipalities where the share of Roma in the total population is below national averages. Because these municipalities effectively fall out of the scope of the sample, the conditions of Roma concentrated in 'mini-poverty pockets' or who are dispersed (presumably integrated with the majority) are not captured. Both groups are represented in the sample, however. In the first case, most of the 85 per cent of Roma who are captured by this survey methodology also live in similar poverty pockets, which benefit from representative sampling. In the second (integrated) case, this would be because a significant portion of the 85 per cent of Roma is functionally integrated (employed, maintaining contacts with majority communities and institutions) and thus typologically similar to dispersed (presumably integrated) Roma from the 15

per cent. Those of the 15 per cent who are 'dispersed and integrated' and self-identify themselves as Roma are typologically close to those who are integrated into the 85 per cent. Those who have been assimilated and do not self-identify as Roma fall out of the scope of the research, either because they don't meet the criterion of 'being Roma' (whatever that means) or because they don't meet the vulnerability criterion.

Looking at the self-identification done through the interview, asking each individual household member to state their ethnic affiliation, 16,198 Roma individuals declared Roma ethnicity out of 17,071 (95 per cent) individuals in the Roma sample. This proves that the sampling method chosen (indirect identification) corresponds very well to the self-identification method, without asking the question: "Are you Roma?".

Finally, it is important to realize that given the uncertainties concerning the size of Roma populations in these countries, the data and the analysis built on them have certain limitations. The survey does not provide precise answers to questions like "How many Roma live in poverty?" or "How many Roma have completed secondary education?". It instead gives answers to questions like "What share of Roma live in poverty?" and "What share of Roma have completed secondary education?". Based on these averages and using estimates of Roma population sizes, certain ranges for those variables can be provided. Such answers are useful for policy purposes because they outline the distance between

various groups, highlight the causes of these differences, and provide benchmarks against which future trends can be assessed. These benchmarks can easily be much more relevant than those based on census data. From a policy perspective, such benchmarks are crucial. The allocation of resources based on official census data (which underestimate the size of Roma communities) inevitably falls short of the scale of needs. Using benchmarks—even in range formats, as presented in this report—can be an important step towards more realistic and adequate policies.

Fieldwork and partnerships

Given the nature of the survey – addressing the needs of groups that are not easy to identify – fieldwork was another major challenge. A high level of trust was needed on the side of respondents – particularly in the case of Roma. As a specific and unique minority group, Roma in some countries show certain levels of distrust towards other ethnic groups and the ethnic majority as well. In order to overcome the possible distrust to enumerators, Roma interviewers were used for the fieldwork where possible (in countries where a sufficient number of trained Roma was available). In other cases Roma intermediaries were used (following the pattern of 'Roma assistant teachers'). These were either Roma 'assistant interviewers' (a Roma representative accompanying the experienced pollster) or local social workers or representatives of Roma NGOs. In

Achieved samples								
Country	Majority		Roma		Displaced		Total	
	Households	Household members	Households	Household members	Households	Household members	Households	Household members
Albania	450	1876	450	2479			900	4355
Bosnia and Herzegovina	404	1240	400	1941	398	1381	1202	4562
Bulgaria	500	1302	500	2176			1000	3478
Croatia	254	715	252	1252	197	656	703	2623
Kosovo	354	2275	354	2223			708	4498
Macedonia	377	1399	379	1836			756	3235
Montenegro	198	700	199	699	204	708	601	2107
Romania	601	1771	601	2905			1202	4676
Serbia	399	1270	399	1759	403	1553	1201	4582
Region	3537	12548	3534	17270	1202	4298	8273	34116

all cases the intermediaries were trained prior to the fieldwork (on the contents of the questionnaire, on general rules and procedures of an interview etc.). The general rule, however, was to approach the communities carefully, with respect and avoiding any suspicion about the purpose of the data collection. Since using 'assistant interviewers' or other intermediaries increased the costs of the survey substantially, this component was financed by the Council of Europe as a part of its "Roma under the Stability Pact" project.

The survey was executed by agencies-members of Gallup International, coordinated by the regional office of Gallup International in Sofia, which managed the data collection of the whole survey. Using Gallup International member agencies, this framework made possible applying similar standards and procedures in all countries covered by the

project, making cross-country comparisons possible and reliable. After the fieldwork was completed a thorough check was run on 10-15 per cent of the sample depending on the country. All errors were cleared and in one case (Montenegro) re-interviewing was necessary. Data control and validation was conducted centrally by the regional office of Gallup International.

From the outset all agencies involved were working in a coordinated manner under the methodological and conceptual guidance of the UNDP Bratislava Regional Centre. The methodology of the survey, sampling and fieldwork were broadly discussed with colleagues from the World Bank and members of the Data Experts Group. Three consultants (Gabor Kezdy, Valerie Evans and Dragana Radevic) were particularly instrumental in the final design of the methodology and sampling models.

Data annex

Table A1

Average monthly household expenditure by category and country (euros)

		Transport	Food	Alcohol	Clothes	Housing	Medicine	Goods	Going out	Total
Albania	Roma	48.2	254.1	61.5	53.6	34.8	25.6	22.6	44.4	544.8
	Majority	51.4	391.2	67.4	133.6	97.5	35.0	35.5	63.4	875
Bosnia and Herzegovina	Roma	18.8	279.9	42.9	33.1	60.6	41.3	28.2	4.1	508.9
	Displaced	26.4	263.7	31.8	57.4	102.0	33.9	36.4	14.9	566.5
	Majority	142.8	286.1	34.0	66.4	106.3	29.3	38.8	25.7	729.4
Bulgaria	Roma	6.0	166.3	25.8	14.8	29.1	15.1	13.4	7.8	278.3
	Majority	18.6	133.3	17.8	24.5	53.2	27.9	12.0	8.0	295.3
Croatia	Roma	56.7	579.8	81.0	147.3	130.0	38.4	109.8	33.8	1176.8
	Displaced	64.1	303.7	56.5	54.0	143.0	20.3	54.7	18.4	714.7
	Majority	85.0	442.3	64.4	131.1	239.9	29.0	70.5	47.2	1109.4
Kosovo	Roma	69.7	347.1	95.8	173.5	158.2	121.5	59.1	48.1	1073
	Majority	128.1	580.2	120.7	261.0	209.1	123.3	73.7	91.4	1587.5
Macedonia	Roma	13.3	238.3	35.4	18.3	61.4	34.7	13.3	3.0	417.7
	Majority	36.7	279.3	28.5	52.7	118.2	30.7	35.2	13.6	594.9
Montenegro	Roma	28.2	323.3	63.6	98.9	63.2	35.2	37.6	31.0	681
	Displaced	47.5	314.0	63.2	94.6	125.0	60.0	37.2	57.0	798.5
	Majority	96.0	466.7	72.5	170.6	187.9	60.0	45.2	85.8	1184.7
Romania	Roma	42.7	193.9	50.9	68.6	52.0	42.1	21.0	42.7	513.9
	Majority	36.6	148.2	42.7	67.2	66.3	31.1	18.7	28.5	439.3
Serbia	Roma	43.0	206.3	35.2	46.5	58.2	31.9	38.5	13.0	472.6
	Displaced	41.9	222.1	36.7	61.9	93.0	31.7	31.8	17.0	536.1
	Majority	44.6	263.1	44.4	86.2	83.4	24.7	37.5	47.1	631

Table A2

Percentage of households owning each durable good

Household items	Majority	Roma	Displaced
Radio receiver	83	60	84
Refrigerator	94	59	89
Oven	83	53	86
TV set	96	80	86
Telephone	70	23	44
Car	45	13	30
CD player	30	8	17
Computer	22	2	13
Internet connection	12	1	7
Satellite dish	21	12	9
Mobile phone	62	29	56
Washing machine	71	31	69
Bed for each household member	90	47	77
Thirty and more books	61	9	28
Generator	7	3	2

Table A3

Household expenditures on bills in euros, with percentage of household total expenditures in parenthesis, by expenditure quintile

Sample	Quintile-group	Water	Electricity	Rent
Majority	1	96.9 (15%)	262.2 (22%)	79.8 (8%)
	2	195.0 (15%)	443.7 (18%)	99.3 (5%)
	3	87.0 (13%)	315.4 (20%)	102.2 (8%)
	4	95.1 (17%)	337.1 (20%)	115.7 (10%)
	5	55.0 (11%)	219.7 (15%)	105.2 (8%)
Roma	1	178.2 (28%)	343.8 (34%)	152.4 (6%)
	2	157.5 (32%)	322.8 (34%)	131.9 (9%)
	3	228.2 (35%)	398.3 (39%)	122.1 (12%)
	4	188.7 (34%)	325.6 (37%)	121.9 (11%)
	5	253.3 (29%)	303.3 (40%)	259.7 (10%)

Table A4

Inter-group inequality

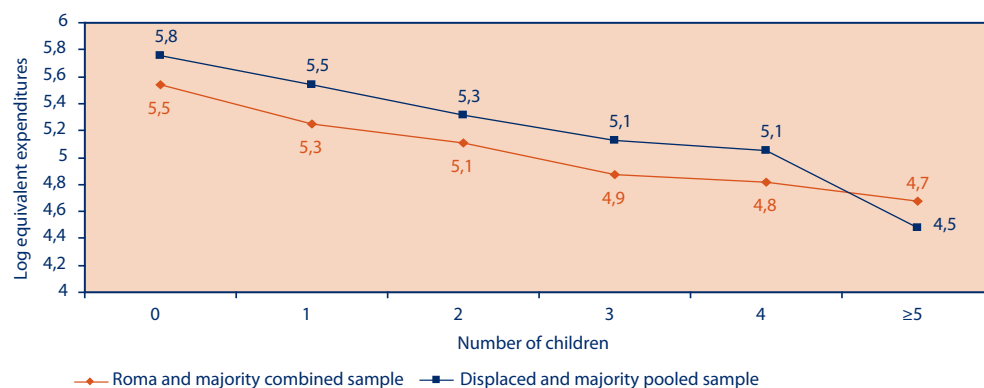
Gini indices, calculations based on total monthly expenditures per equalized capita in euros (number of observations).

	All	Majority	Roma	Displaced
	.4335 (7706)	.3939 (3372)	.4383 (3315)	.3475 (1019)
Albania	.4783 (891)	.4082 (442)	.3860 (449)	N/A
Bosnia and Herzegovina	.3648 (1171)	.3154 (398)	.3983 (382)	.3051 (391)
Bulgaria	.3405 (935)	.2948 (468)	.3094 (467)	N/A
Croatia	.3805 (612)	.2863 (241)	.3773 (240)	.3767 (131)
Kosovo	.3796 (706)	.3427 (354)	.3866 (352)	N/A
Macedonia	.3446 (732)	.2859 (359)	.3140 (373)	N/A
Montenegro	.3236 (446)	.2551 (144)	.3596 (155)	.2829 (147)
Romania	.3401 (1157)	.3002 (587)	.3188 (570)	N/A
Serbia	.4051 (1056)	.3097 (379)	.4746 (327)	.3551 (350)

FIGURE A1

Linear relationship between log equivalent expenditures and number of children

For Roma and majority combined samples as well as displaced and majority pooled samples



Description of variables (Used in group-related determinants of poverty regression)	
Variable	Mean (Standard deviation in parenthesis)
Household expenditures equivalized	627.5 (622.5)
Log expenditures equivalized	5.3 (0.8)
Number of children	0.9 (1.3)
Dummy Variable	Percentage of Sample
Skill level of household head	
- Skilled	63.3
- Unskilled	36.7
Education level of household head	
- Well educated	62.5
- Poorly educated	37.5
Country of residence	
- Albania	14.6
- Bosnia and Herzegovina	10.7
- Bulgaria	11.7
- Kosovo	15.1
- Macedonia	10.8
- Montenegro	4.7
- Romania	15.7
- Serbia	15.1

Group-related determinants of poverty	
Coefficients for linear regression analysis of log equivalized expenditures with membership of the Roma sample, country of residence, capital or rural locality, number of children in household, and the skill and education level of household head	
Roma	-0.42***
Rural	-0.23***
Capital	0.19***
Number of children	-0.10**
Skill level	0.32***
Education level	0.39***
Albania	-1.08***
Bosnia and Herzegovina	-0.01
Bulgaria	-0.62***
Kosovo	-0.78***
Macedonia	-0.53***
Montenegro	-0.50***
Serbia	-0.84***
Romania	-0.82***
A	5.87***
R-squared	0.53

***p<0.01, **p<0.05

Table A7

Description of variables (Used in regression analysis of determinants of poverty within each group)		
	Mean (Standard deviation in parenthesis)	
Variable	Roma	Majority
Household expenditures equivalized	73.5 (89.1)	140.7 (132.1)
Log expenditures equivalized	4.98 (0.79)	5.7 (0.70)
Number of children	1.24 (1.50)	0.53 (0.97)
	Percentage of Sample (with each level of variable)	
Dummy Variable	Roma	Majority
Skill level of household head		
- Skilled	33.1	84.2
- Unskilled	66.9	15.8
Education level of household head		
- Well educated	38.8	85.9
- Poorly educated	61.2	14.1
Country of residence		
- Albania	14.4	15.0
- Bosnia and Herzegovina	11.2	9.9
- Bulgaria	12.6	10.4
- Kosovo	12.9	18.1
- Macedonia	10.6	11.1
- Montenegro	4.0	5.6
- Romania	16.8	14.1
- Serbia	10.2	10.1

Table A8

Determinants of poverty within each group		
Coefficients for linear regression analysis of log equivalized expenditures for Roma and majority households with the country of residence, locality, and number of children in a household and the skill level and level of education of the household head		
	Roma	Majority
Rural	-0.75***	-0.29***
Capital	0.00	0.24***
Number of children	-0.10*	-0.00
Skill level	0.32***	0.35***
Education level	0.39***	0.42***
Albania	-1.14***	-1.00***
Bulgaria	-0.51***	-0.70***
Bosnia and Herzegovina	0.00	-0.00
Macedonia	-0.45***	-0.55***
Serbia	-0.97***	-0.75***
Montenegro	-0.38**	-0.55***
Romania	-0.72***	-0.91***
Kosovo	-0.61***	-0.86***
A	5.43***	5.85***
R-squared	0.35	0.43

***p<0.01, **p<0.05, *p<0.1

Table A9 Description of variables used in the Roma determinants of education logistic regression	
Mean and standard deviation (in brackets) for continuous variables	
Age (years)	21.4
Age-squared	723.4
Percentage of each sample with each level of dependent variables used in logistic regression analyses	
Elementary to primary education	
- Elementary	56.9
- Primary	43.1
Primary to secondary	
- Primary	79.6
- Secondary	20.4
Percentage of each sample with positive responses for the explanatory dummy variable	
Role-model (positive)	34.9
Male	42.4
Poor	51.6
With chronic illness	16.1

Table A10 Odds ratios for explanatory variables in the Roma determinants of education logistic regression (standard errors shown in parenthesis)¹²⁹		
	Elementary to primary	Primary to secondary
Male	1.53*** (0.07)	1.48*** (0.13)
Age	0.89*** (0.01)	0.90*** (0.03)
Age-squared	1.00*** (0.00)	1.00*** (0.00)
Poor	0.66*** (0.06)	0.51*** (0.14)
Household head's education	1.69*** (0.07)	1.97*** (0.14)
Illness	0.82** (0.09)	1.48 (0.00)
R-squared ¹³⁰	0.10	0.08

***p<0.01, **p<0.05, *p<0.1

¹²⁹ Due to the small number of Roma with secondary or tertiary education, logistical regression analyses were not performed for individuals with secondary or tertiary education.

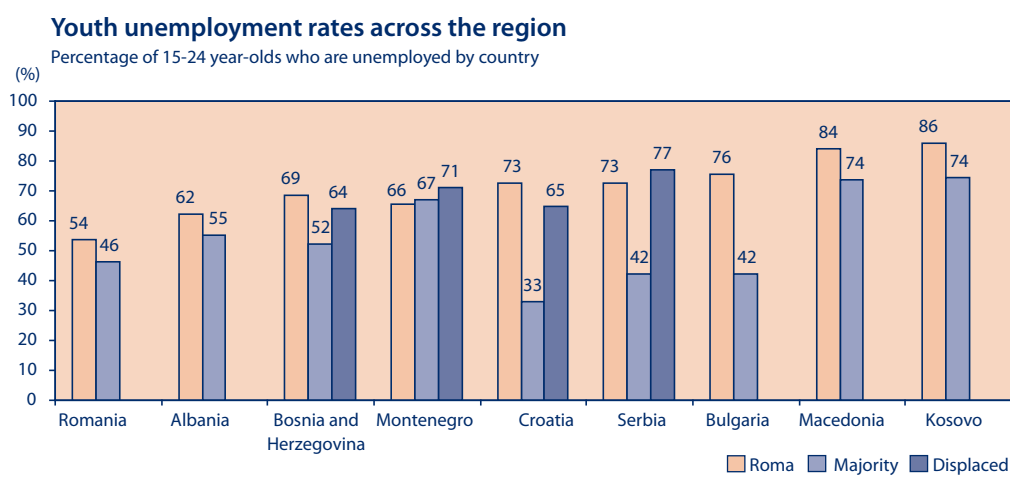
¹³⁰ Using Nagelkerke R-squared value.

Table A11 Description of variables used in group-related determinants of education logistic regression analyses	
Mean and standard deviation (in brackets) for continuous variables	
Age	24.5 (18.0)
Age-squared	924.1 (1258.6)
Percentage of total sample with positive responses for dependent dummy variable	
Elementary to primary dummy	
- Elementary	52.8
- Primary	47.2
Primary to secondary dummy	
- Primary	52.7
- Secondary	47.3
Secondary to tertiary dummy	
- Secondary	83.9
- Tertiary	16.1
Percentage of total sample with positive responses for the explanatory dummy variable	
Roma	60.4
Well-educated head	49.4
Male	40.8
Poor	37.2
With chronic illness	15.5

Table A12 Increase in odds of increasing education from one level to the next associated with each explanatory variable in the group-related determinants of education logistic regression analyses (standard errors shown in parenthesis)			
	Elementary to primary	Primary to secondary	Secondary to tertiary
Roma	0.36*** (0.08)	0.20*** (0.08)	0.22*** (0.70)
Male	1.53*** (0.06)	1.58*** (0.08)	0.82 (0.13)
Age	0.90*** (0.00)	0.85*** (0.01)	0.88*** (0.03)
Age-squared	1.00*** (0.00)	1.00*** (0.00)	1.00*** (0.00)
Poor	0.68*** (0.06)	0.37*** (0.09)	0.18*** (0.42)
Household head's education	1.53*** (0.06)	2.53*** (0.09)	3.33*** (0.34)
Illness	0.78*** (0.08)	0.82* (0.11)	0.66** (0.18)
R-squared ¹³¹	0.12	0.33	0.14

***p<0.01, **p<0.05, *p<0.1

FIGURE A2



¹³¹ Using Nagelkerke R-squared value.

FIGURE A3

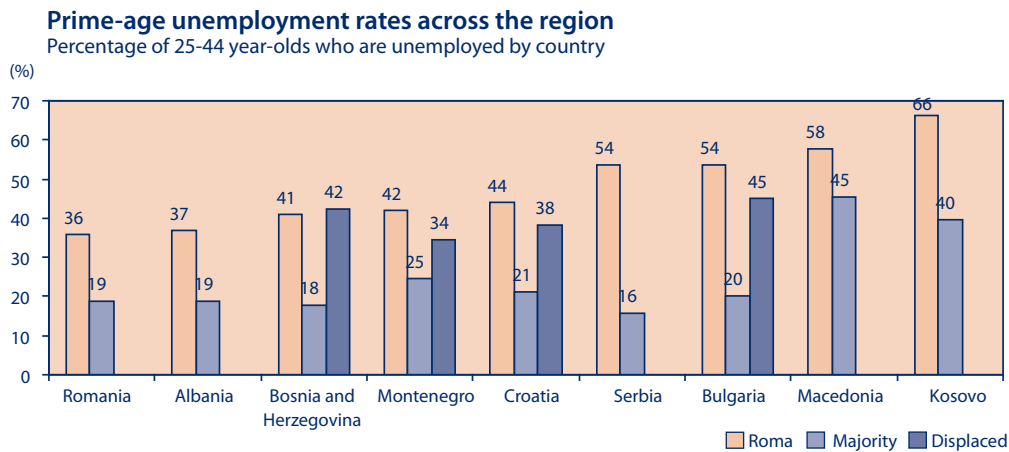


FIGURE A4

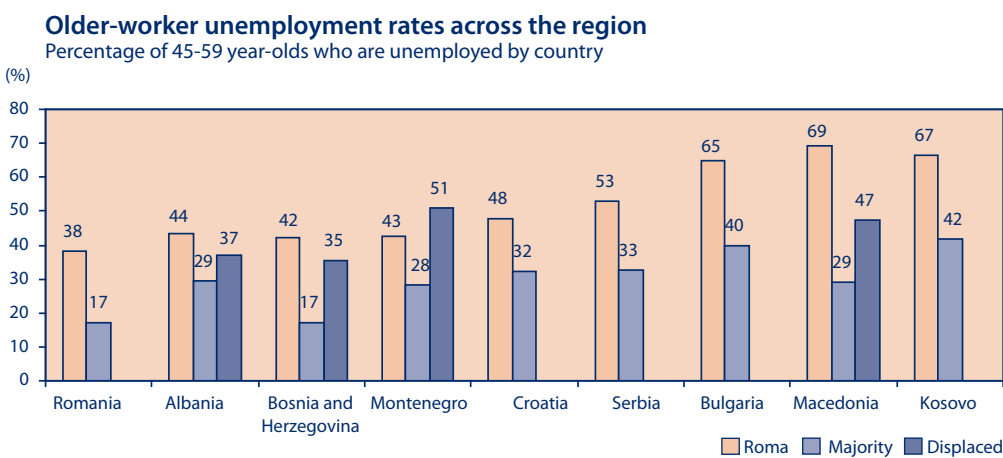


Table A13

The effect of education on the probability of employment ¹³²						
	Majority		Roma		Displaced	
Coefficients on schooling (compared to the base category: no schooling, elementary or incomplete primary education)	Effect	Std. error	Effect	Std. error	Effect	Std. error
Men						
Completed primary	.23*	.13	.06	.04	-.30*	.18
Completed secondary	.45***	.12	.37***	.07	-.20	.16
Completed tertiary	.90***	.04	.69***	.25	.07	.18
N	3446		4865		1195	
Pseudo-R ²	.15		.10		.12	
Women						
Completed primary	.28**	.13	.28***	.05	-.09	.17
Completed secondary	.72***	.12	.49***	.09	.31**	.15
Completed tertiary	1.39***	.13	.53*	.29	.89***	.18
N	2928		3727		908	
Pseudo-R ²	.19		.10		.07	
Baseline employment probability	.61		.28		.41	

*p<0.1, **p<0.05, ***p<0.01

¹³² The table reports the effects, in terms of the percentage point improvement in employment probability, of raising one's education level from no- or elementary-education level to a primary, secondary or tertiary level, respectively. The probit model was estimated separately for adult men and women (16 years and above). Apart from the education variables, the model also included country-specific intercepts and age and age-squared variables.

The returns to education by ethnic group						
	Majority		Roma		Displaced	
Percentage change in wage arising from increasing one's educational level from	% change	Std. error	% change	Std. error	% change	Std. error
Males						
Elementary to primary	.18	.13	.29***	.04	.39***	.15
Primary to secondary	.25**	.12	.18***	.06	.23*	.13
Secondary to tertiary	.23*	.13	.62***	.19	.48***	.14
N	2173		2244		598	
R ²	.92		.93		.81	
Females						
Elementary to primary	.17	.15	.24***	.07	.37	.25
Primary to secondary	.38***	.14	.41***	.11	.51***	.22
Secondary to tertiary	.37***	.14	.21	.28	.67***	.24
N	1420		880		301	
R ²	.93		.93		.77	

*p<0.1, **p<0.05, ***p<0.01

Group-related wage gap								
The estimated wage of majority and Roma men and women relative to the average wage of an uneducated member of the majority								
	Men				Women			
	No education	Primary	Secondary	Tertiary	No education	Primary	Secondary	Tertiary
Majority – region	0	21	52	86	0	17	60	120
Roma – Bulgaria	-41	-24	-11	4	-34	-18	9	23
Roma – Kosovo	-46	-31	-18	-5	-37	-22	4	17
Roma – Serbia	-43	-27	-14	1	-45	-32	-9	2
Roma- Croatia	-14	11	31	53	-47	-34	-13	-1
Roma – Macedonia	-51	-37	-25	-13	-48	-35	-14	-3
Roma – Romania	-43	-27	-14	0	-53	-41	-22	-12
Roma – Montenegro	-46	-31	-19	-5	-63	-54	-38	-30
Roma – Albania	-61	-50	-41	-32	-64	-55	-40	-32
Roma – Bosnia and Herzegovina	-55	-42	-31	-20	-71	-64	-52	-45

Description of variables (Used in the group-related determinants of poverty regression)	
Variable	Mean (Standard deviation in parenthesis)
Household expenditures equivalized	136.6 (122.5)
Log expenditures equivalized	5.7 (0.7)
Number of children	0.4 (1.0)
Dummy Variable	Percentage of Sample
Displaced	25.3
Capital	17.5
Rural	29.8
Well-educated head	85.6
Skilled head	82.7
Country of residence	
- Bosnia and Herzegovina	31.9
- Croatia	16.7
- Montenegro	17.1
- Serbia	34.3

Group-related determinants of poverty Coefficients for linear regression analysis of log equivalized expenditures with membership of the displaced sample, country of residence, capital or rural locality, number of children in household, and the skill level and level of education of the household head		
	Original model	Reduced form model
Displaced	-0.30***	-0.30***
Rural	-0.12***	-0.12***
Capital	0.12**	$10^{-2} \times -9.8^{**}$
Capital*Displaced	$10^{-2} \times -6.0$	
Number of children	$10^{-2} \times -9.9^{***}$	$10^{-2} \times -9.9^{***}$
Skill level	0.38***	0.38***
Education level	0.65***	0.65***
Croatia	$10^{-2} \times -5.4$	
Montenegro	-0.46***	-0.45***
Serbia	-0.83***	-0.79***
A	5.66***	5.65***
R-squared	0.47	0.47

***p<0.01, **p<0.05

Table A18

Description of variables		
(Used in the regression analysis of determinants of poverty within each group)		
Variable	Mean (Standard deviation in parenthesis)	
	Displaced	Majority
Household expenditures equivalized	708.8 (547.5)	1020.9 (730.4)
Log expenditures	5.56 (0.87)	6.10 (0.65)
Number of children	0.82 (1.13)	0.51 (0.82)
Dummy variable	Percentage of sample (with each level of variable)	
	Displaced	Majority
Locality		
- Rural	32.0	16.7
- Capital	16.1	23.7
- Urban	51.9	59.6
Skill level of household head		
- Skilled	71.8	90.3
- Unskilled	28.2	9.7
Education level of household head		
- Well educated	84.6	91.8
- Poorly educated	15.4	8.2
Country of residence		
- Bosnia and Herzegovina	33.0	32.0
- Croatia	16.4	20.4
- Montenegro	16.9	15.7
- Serbia	33.6	32.0

Table A19

Determinants of poverty within each group		
Coefficients for linear regression analysis of log equivalized expenditures for displaced and majority households with the country of residence, locality, and number of children in a household and the skill level and level of education of the household head		
	Displaced	Majority
Rural	-0.15**	-0.10*
Capital	$10^{-2} \times 4.5$	0.12**
Number of children	$10^{-2} \times -5.1^*$	-0.15***
Skill level	0.36***	0.42***
Education level	0.73***	0.52***
Serbia	-0.91***	-0.71***
Montenegro	-0.43***	-0.46***
A	5.29	5.74
R-squared	0.45	0.40

***p<0.01, **p<0.05