DECOMPOSING GAPS BETWEEN ROMA AND NON-ROMA IN ROMANIA

CHRISTOPHER RAUH
Université de Montréal

Abstract It is widely known that the Roma have been suffering persistent disadvantages. Yet, little empirical evidence exists. Using the censuses of 1977, 1992, 2002, and 2011, I provide a comprehensive overview of the past, present, and an outlook on the future of the Roma in Romania, home to a large and rapidly growing Roma community. Young Roma, in particular girls, are less likely to be attending school, indicating that lack of educational attainment is likely to persist. The Roma have worse housing conditions and face lower employment and higher unemployment levels. Amongst Roma, females are less likely to be employed than males. Oaxaca–Blinder decompositions of the ethnic and gender employment gaps reveal that the differences in employment cannot be fully explained by observables, such as age or education. Despite the seemingly dire picture, there are signs of improvement for more recent cohorts, as literacy rates have reached close to universal levels.

Keywords: Gypsies, Roma, inequality, Romania, Oaxaca–Blinder, gender gap, ethnic minorities

JEL classification: I24, J15, J21, J71

1. INTRODUCTION

The Roma form a large minority in the European Union with an estimated population of 4–9 million [Pan and Pfeil (2003), Liégeois (2007a, 2007b), Council of Europe (2012)]. Their low socio-economic status combined with violent incidents of discrimination has increased international alert and triggered the European Union to declare 2005–2015 "The Decade of Roma Inclusion" with the goal of improving their socio-economic status and social inclusion. Despite their poor outcomes being known in a popular sense, there is little systematic research documenting their situation. In this paper, I look at the development of a range of outcomes in Romania based on four population censuses conducted in 1977, 1992, 2002, and 2011. Outcomes and characteristics of family composition, housing conditions,

209

education, language, and the labor market are examined and segmented by age or year of birth, urban versus rural, and gender. According to the censuses, the growing Roma population of Romania has reached 614,010 (or 3.1%) in 2011, but estimates range up to 1.8–2.5 million [ERRC and Danova-Rusinova (2004)], as many Roma tend not to reveal their ethnicity out of fear of stigmatization.[1]

Looking at the Romanian censuses, I reveal many disadvantaged outcomes of Roma compared to the rest of the population, but also within Roma, among which females suffer from worse outcomes than men. The primary problem continues to be educational achievement, as school attendance and primary school completion remain low. The likelihood of young Roma attending school is low, sowing the seed for low educational achievement across future generations. The low levels of education are reflected in the labor market where unemployment among Roma is high and around 90% of those that work are clustered in blue collar occupations, many of which tend to belong to the informal sector. Controlling for observables, Roma have a 30% lower rate of employment and 76% higher rate of unemployment. Female Roma suffer from an even lower probability of employment than males. An Oaxaca–Blinder decomposition suggests that in 2011 still more than one quarter of the employment gap between Roma and non-Roma cannot be explained by observable characteristics. Comparing female and male Roma, the decomposition reveals that nearly 90% of the employment gap cannot be attributed to differences in observables. These two findings indicate that between the population groups, but also within Roma, discrimination or cultural practices might contribute to persistent inequalities beyond what we can observe in the data.

The level of segregation in the Romanian society is indicated by extremely low interethnic marriage rates that I document. In terms of housing, Roma suffer from worse conditions than non-Roma. In 2011, less than half of Roma dwellings compared to the rest of the population have a toilet, a sewage system or access to running water, and less than half the area and number of rooms per person. Also, access to electricity is considerably lower and there are more people per household. All these findings hold for rural as well as urban dwellings and potentially contribute to unhygienic circumstances and consequently health issues.[2]

The Roma odyssey can be dated back to as early as the 15th century, when they were traded as slaves in what today is Romania. Roma slavery was institutionalized and was not abolished until 1856 [Crowe (2007)]. Even after the end of slavery, misery continued as they were unskilled and a few received land. One of the darkest episodes of their history was the Nazi era in which between 1941 and 1942 an estimated 25,000–36,000 Roma were transported from Romania to concentration camps, and at least half fell victim to the Holocaust [Crowe (2007)]. During the socialist era, under the influence of the Soviet Union, Roma were forced into agricultural collectives and heavy industry, whereas their traditional occupations were declared illegal [Gilberg (1974), Beck (1984)]. In the 1980s, Romanian president Ceausescu resettled entire villages and neighborhoods in his attempt of forcing assimilation [Crowe (2007)]. Although they could surely not be considered an advantaged group in this historical period, they were largely

assured basic education and employment [O'Higgins (2010)], which seems to be one reason why their situation deteriorated between 1992 and 2002 with the rollback of the state. Discrimination and negative attitudes toward Roma remain high, as according to an Ethnobarometer survey taken in 2000, nearly 40% of non-Roma would prohibit Roma from settling in Romania [Ringold et al. (2005)].

Concerning empirical studies on Roma, many studies are based on detailed non-representative surveys conducted by the UNDP in 2004 spanning nearly 30,000 observations across nine Eastern European countries.[3] The consensus is that Roma lag behind non-Roma in education, employment, and wages even when controlling for observables. Focusing on education using data collected by the UNDP in 2011, Brüggemann (2012) and O'Higgins and Brüggemann (2014) show that Roma are more likely to be sent to remedial schools, which are often equipped poorly [ERRC and Danova-Rusinova (2004)], leading to lower levels of achievement. Using the same data, Cukrowska and Kóczé (2013) show that Roma face poor housing conditions and among Roma, females achieve lower levels of education than men. For Roma actually still attending school in Hungary, Kertesi and Kézdi (2011b) find that a large share of the test-score gap between Roma and non-Roma students can be explained by social differences in income, wealth, and parental education. In Kertesi and Kézdi (2016), they can specifically link this gap to the home environment and parenting practices, finding that these factors can explain a substantial part. The representative nature of the data employed in my study allows us to conclude that not only test-score gaps are important, but also the enrollment gap, i.e., the extensive margin of education, plays an important role for children as young as seven. Moreover, this gap cannot be fully explained by parental characteristics.

An empirical study focusing on Roma inequality in Romania is based on a non-representative survey in 2000, finding that 70% of Roma were below the $4.30 Purchasing Power Parity (PPP) poverty line [Revenga et al. (2002)]. Examining total household expenditures during the previous month in the same dataset, Rat (2005) finds that 39% of Roma households belong to the lowest quintile. The poorest, in particular rural, Roma survive on day labor and informal activities, such as recycling scrap metal [Revenga et al. (2002)]. As a result of Roma underachievement in the labor market, the World Bank estimates a loss of 202–887 million Euro in Romania annually in productivity and fiscal contributions to the governments. The annual fiscal gains from bridging the employment gap are estimated at 2.4 times the total cost of investing in public education for all Roma children in Romania [De Laat (2010)].

The study on Roma that bears the closest relation in terms of methodology is by Kertesi and Kézdi (2011a). They investigate the magnitude and drivers of the wage gap and low formal employment rate of Roma in Hungary using two surveys spanning 15 years. Although the census data used in my study do not contain information on earnings, the censuses provide larger representative samples stretching over a longer time period and shed light on demographic trends and determinants of (un)employment, occupational inequalities, and education. The

presented study, to the best of my knowledge, is the first to provide a large scale descriptive quantitative overview with multiple data points of the Roma population in Romania documenting the evolution of disadvantages and allowing insights into trends and future challenges. I highlight the persistence of gaps in fertility rates and educational achievement, which already existed for birth cohorts as early as 1900. In Romania, the Roma form the largest group in terms of absolute numbers compared to the rest of Europe [Council of Europe (2012)].[4] Therefore, it is a particularly important place to understand the Roma disadvantages as the Roma are predominantly native, thereby forming a historically rooted minority. In many other countries, the Roma often have settled only recently or a few generations ago and hence other immigrants, rather than locals, could be considered the adequate control group. Their size and history make the Roma of Romania a particular compelling group to understand the evolution and trends of disadvantages faced. We know not only that gaps exist, but that they have persisted over generations and that they are likely to persist in the future. Moreover, the fact that the Roma population is growing at a high rate could exacerbate the impact of this lack of development and persistent inequality.

The paper is organized as follows: In Section 2, the data is described and basic demographics, household composition, and dwelling characteristics are compared. In Section 3, the human capital differences between Roma and the rest of the population are decomposed by looking at educational attainment, school attendance, literacy rates, and language. In Section 4, differences in labor market outcomes and occupation are explained, whereas Section 5 concludes.

2. DATA AND SAMPLE

The data are random representative 10% subsamples of each of the Romanian censuses from the National Institute of Statistics of Romania, which I obtained from IPUMS International [Minnesota Population Center (2017)].[5] The censuses include the entire Romanian population as all households are surveyed in face-to-face interviews irrespective of individual characteristics and legal situation. The 1977 census includes 619,904 households (4,309 Roma) and 1,937,021 individuals (19,716 Roma). In 1992, the sample covers 728,846 households (8,242 Roma) and 2,238,578 individuals (39,597 Roma). In 2002, there are observations of 732,016 households (11,378 Roma), and 2,137,967 individuals (52,619 Roma). In 2011, information on 746,908 households (14,730 Roma) and 1,991,924 individuals (61,401 Roma) is available. Information on the labor market is only available from the 1992, 2002, and 2011 censuses.[6]

2.1. Demographics

In Table 1, we see that Roma form a growing part of the Romanian population. While the non-Roma population has not increased between 1977 and 2011, the reported Roma population has tripled in the same time period, and now accounts

TABLE 1. Descriptive statistics of sample

	1977		1992		2002		2011	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ethnic Roma	0.010	0.1	0.018	0.13	0.025	0.15	0.031	0.17
Age	37.64	69.79	34.9	24.05	37.44	21.86	40.21	22.46
Number of children	0.73	1.12	0.73	1.13	0.67	1	0.54	0.9
Immigrant	0.01	0.11	0.01	0.09	0.01	0.08	0.01	0.08
Urban	0.45	0.50	0.54	0.50	0.52	0.5	0.54	0.50
Married	0.57	0.50	0.56	0.50	0.53	0.5	0.57	0.50
Labor market								
Employed			0.42	0.49	0.36	0.48	0.51	0.50
Unemployed			0.08	0.28	0.12	0.32	0.07	0.26
Not in labor force			0.54	0.50	0.59	0.49	0.45	0.50
Conditional on *employment*								
Hours worked					41.02	11.32	39.28	7.32
Wage/salary worker			0.80	0.40	0.70	0.46	0.70	0.46
Unpaid worker			0.02	0.15	0.15	0.36	0.12	0.32
Self-employed			0.18	0.39	0.14	0.34	0.18	0.39
Education								
Less than primary	0.50	0.50	0.30	0.46	0.25	0.43	0.17	0.38
Completed primary	0.25	0.44	0.31	0.46	0.28	0.45	0.27	0.44
Completed secondary	0.22	0.41	0.34	0.47	0.40	0.49	0.42	0.49
University graduate	0.03	0.16	0.05	0.22	0.07	0.25	0.14	0.35
Literate	0.95	0.21	0.97	0.17	0.97	0.16	0.99	0.11
Observations	1,937,021		12,238,578		12,137,967		11,991,924	

Notes: For 1977 no labor market information is available. Hours worked are weekly hours worked conditional on employment. *Number of children* refers to own children in household.
Datasource: IPUMS International.

for 3.1% of the total population. Generally, the share of Roma is assumed to be underestimated in the censuses, as many are assumed not to reveal their ethnicity in surveys.[7]

Not only has the share of Roma of the total population increased, but also the population pyramid in Figure 1 indicates the important role Roma will play in the future. Despite life expectancy being about 10 years lower for Roma [McKee (1997)], the fact that only 11% of the rest of the population is younger than 10 years old, compared to 23% of Roma, suggests that a growing share of working age population will be Roma in the future.

A different way of showing why the Roma population has grown in the past and is likely to continue to grow in the future is Figure 2. The panels show the number of live births per female for Roma (black line) and non-Roma (gray line) and the corresponding 95% confidence interval as dashed lines. The left panel includes all females above age 40 from all samples by year of birth.[8] We can
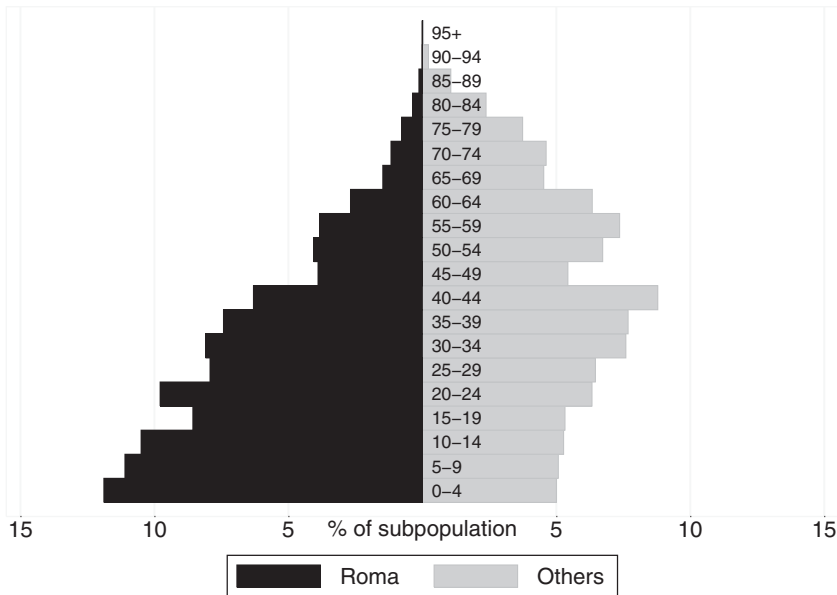
FIGURE 1. Population pyramid by ethnicity and age group in 2011. Datasource: IPUMS International.

see that the average number of children has been decreasing by cohort across the board. Although for non-Roma it has fallen below the replacement rate, it still is well above three children for Roma females. The right panel shows the number of children born for all women above age 14 in the 2011 sample. Here, it not only becomes clear that Roma have more children but also that many females already have children extremely early. At age 18, the average number of children born to a Roma female already is 0.57. In Online appendix Table B.3, I show that even controlling for education and county fixed effects, Roma females on average give birth to one additional child.

According to the censuses, the Roma are the third largest ethnic group in Romania (after Hungarians and Romanians) over the observed time period. In general, the population experienced urbanization between 1977 and 1992 during the transition from Communism, but afterwards the urban versus rural distribution has remained fairly stable. Although 54% of non-Roma live in urban neighborhoods in 2011, only 36% of Roma do so, as can be seen in Table B.1 of Online appendix. Interethnic marriage remains a rare event in Romania. Even in 2011, only 0.0013% of non-Roma household heads have a Roma spouse, whereas 93.1% of Roma household heads have a Roma spouse. On average, Roma are younger, have less education, and are less (more) likely to be (un)employed. In Table B.2 of Online appendix, we can see that within Roma, on average, females are less educated and less (more) likely to be (un)employed. In Sections 3 and 4, the education and employments gaps will receive systematic attention.
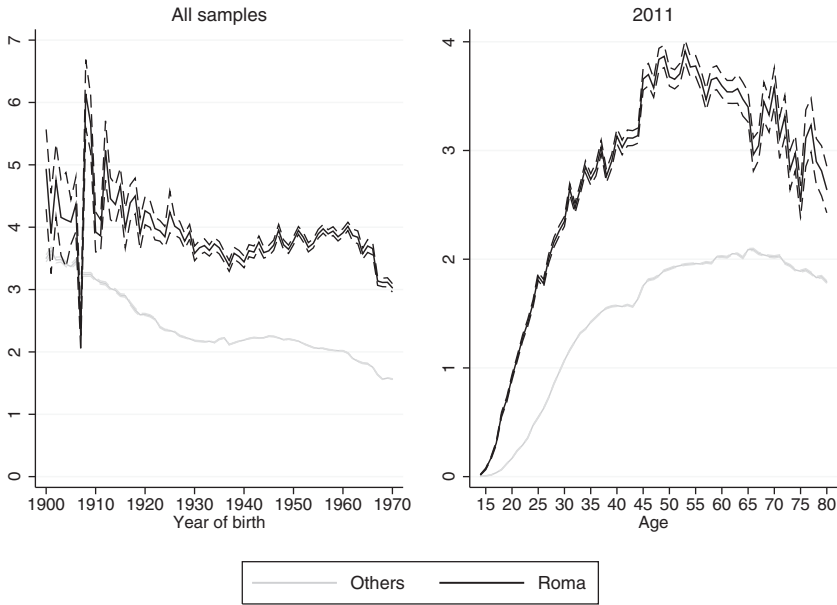
FIGURE 2. Number of children born per female by ethnicity. The left panel shows the average number of children by year of birth pooling all samples restricted to females above the age of 40. The right panel shows the average number of children by age using the 2011 sample only. The dashed lines represent the 95% confidence interval. Datasource: IPUMS International.

Roma are said to have strong family ties, which is reflected in their household composition. In 2011, on average 4.5 people live in one Roma household, whereas non-Roma share a household amongst around 2.6 people on average.[9] Despite having more people in a household, the Roma are not equipped with larger housing units. On average, Roma dispose of 12 m² per capita, which is less than half of the area of others. These differences between the population groups have stayed large across the observed time period with no recognizable convergence. In Figure 3, we can see that Roma households, both rural and urban, are much less likely to be equipped with electricity, a sewage system, or toilets.

3. EDUCATION

The educational attainment of Roma and others by year of birth is shown in Figure 4. Most recent cohorts of Roma are again finally showing signs of increasing completion of primary schooling and actually a minuscule but nonzero share has completed tertiary education. In Figure B.1 of Online appendix, the equivalent is presented for male and female Roma. As can also be observed in Table B.2 of Online appendix, female Roma consistently achieve lower educational levels than
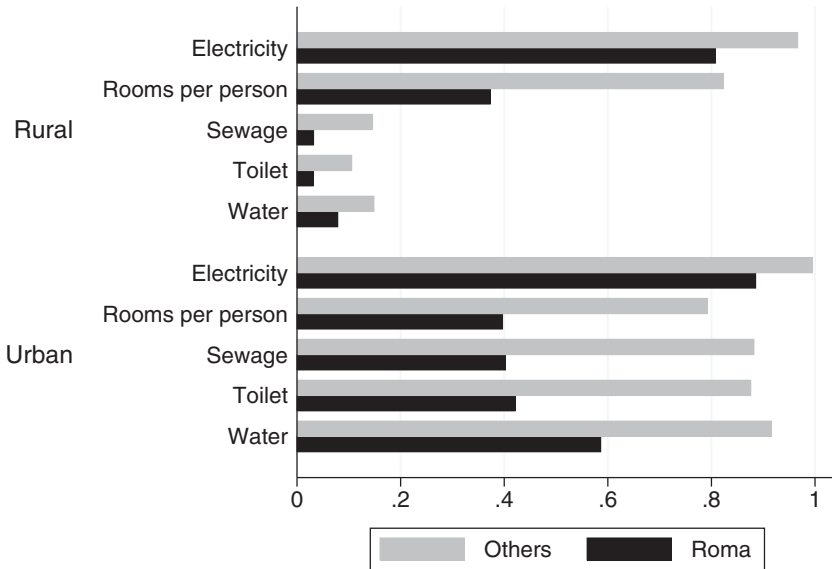
FIGURE 3. Dwelling characteristics of rural and urban Roma and others in 2011. "Room per person" is the average fraction of rooms per person shared in a household. All other bars represent the share of household dwellings equipped with electricity, a sewage system, and a toilet in rural (top) and urban (bottom) areas. Datasource: IPUMS International.

male Roma throughout the observed time period, with no systematic closure of the gap between male and female Roma noticeable.

3.1. School Attendance

The lack of universal primary school completion amongst future generations of Roma can be deducted from Figure B.2 of Online appendix, where the percentage attending school in 2011 and the 95% confidence interval is plotted on the y-axis against age on the x-axis. When comparing the share of Roma (black line) to non-Roma (gray line), the attendance curve of non-Roma strictly dominates.

In order to estimate the effects of different circumstances on educational attainment of Roma, I run a logistic regression estimating whether an individual aged 7–18 attends school in 2011, which provides insights into the educational attainment of the next generation of workers. The estimation controls for a vector of individual (X_i) and parental characteristics (Z_i). The baseline model of the probability of an individual i attending school (school$_i \in \{0, 1\}$) can be written as Prob(school$_i = 1|X_i, Z_i$, Roma$_i, \pi_s) = f(X_i, Z_i,$ Roma$_i, \pi_s)$.

Individual controls include gender, the mother tongue of the child, whether they live in a rural or urban area, and county fixed effects π_s. Age is controlled for by using a polynomial of degree 3 in order to account for the non-linear effect of
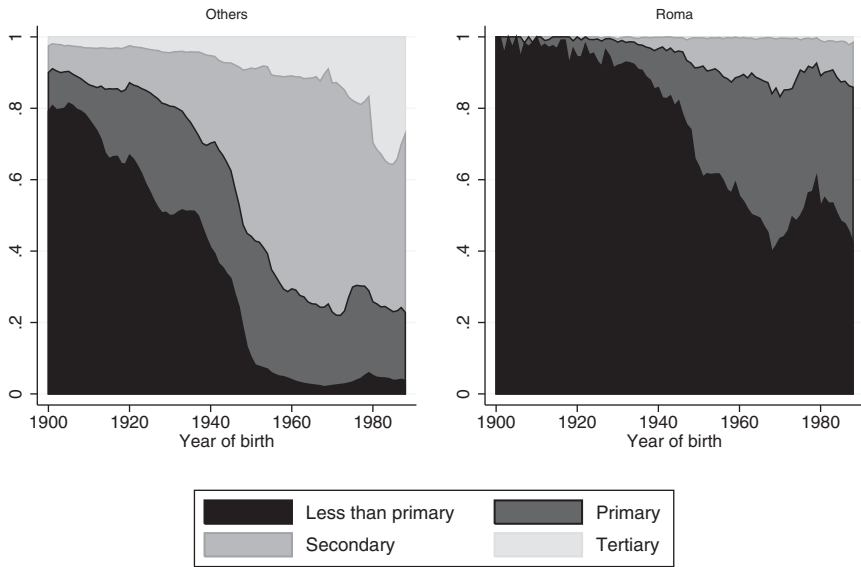
FIGURE 4. Highest level of educational attainment of non-Roma and Roma by year of birth. This figure shows the fraction of the population and their highest level of education obtained by year of birth. The sample is restricted to individuals above the age of 22. Datasource: IPUMS International.

age on school attendance.[10] Parental characteristics include maternal educational attainment and whether a father is in the household. Educational attainment is split into dummy variables for primary schooling completed and secondary or more schooling completed. Less than primary schooling forms the baseline category. In specification (1) of Table 2 one can see that despite all these controls, Roma still have a lower likelihood of attending school, as the Roma dummy has a negative coefficient, which is significant at the 1% level. Given a baseline attendance of 88%, Roma are, ceteris paribus, 7% less likely to be attending school.

Given the data, we can only speculate about reasons such as parental demand for child labor, or lack of commitment or valuation of education.[11] Looking at the signs and significance of other coefficients, one sees that more factors contribute to the lower likelihood of Roma attending school. The probability of attending school is positively associated with maternal educational attainment and having a father present. Also, living in an urban settlement increases the probability of school attendance. Since Roma are less educated and more likely to live in rural areas, one can imagine that the education gap is far from closing. Furthermore, speaking a Roma language as a mother tongue decreases the likelihood of school attendance, which nearly half of Roma children do. This suggests that the misalignment of language spoken at home and language of instruction could play a role in explaining part of Roma underachievement in education.

TABLE 2. Logistic regression estimating whether child attends school

Dependent variable: School attendance

		Roma	
	All	Female	Male
	(1)	(2)	(3)
Roma	− 0.064***		
	(0.003)		
Female	0.011***		
	(0.001)		
Immigrant	0.009	0.054	0.037
	(0.007)	(0.133)	(0.115)
Roma language	− 0.019***	− 0.088***	− 0.051***
	(0.002)	(0.013)	(0.011)
Urban	0.010***	0.021*	− 0.019*
	(0.001)	(0.012)	(0.011)
Age	0.571***	0.958***	0.829***
	(0.016)	(0.080)	(0.068)
Age squared	− 0.041***	− 0.073***	− 0.060***
	(0.001)	(0.007)	(0.005)
Age cubed	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)
Single mother	− 0.007***	− 0.015	− 0.014
	(0.002)	(0.017)	(0.016)
Maternal education			
Primary school	0.054***	0.096***	0.093***
	(0.003)	(0.013)	(0.012)
At least secondary school	0.092***	0.135***	0.149***
	(0.003)	(0.023)	(0.023)
Sample mean	0.879	0.685	0.718
County FE	Yes	Yes	Yes
Observations	218,697	6,305	7,073
R^2	0.390	0.185	0.199

Notes: Robust standard errors in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Reported coefficients are marginal effects. The sample is restricted to individuals aged 6–18. In column (1), both Roma and non-Roma are included, whereas in columns (2) and (3) the sample is restricted to female and male Roma, respectively. County fixed effects and a constant term are included in all specifications. Baseline categories for dummy variables are the following: less than primary schooling for maternal education, male (*Female*), rural (*Urban*), non-Roma (*Roma*).

In columns (2) and (3), I restrict the sample to Roma only and split the analysis by gender to see whether different characteristics have a differential impact on boys and girls. For girls, speaking a Roma language seems to be more detrimental. Living in an urban area increases attendance for girls while reducing it for boys. Although the other covariates exhibit similar effects across boys and girls, the overall attendance rate is lower for girls than for boys (68.5% vs. 71.8%). This
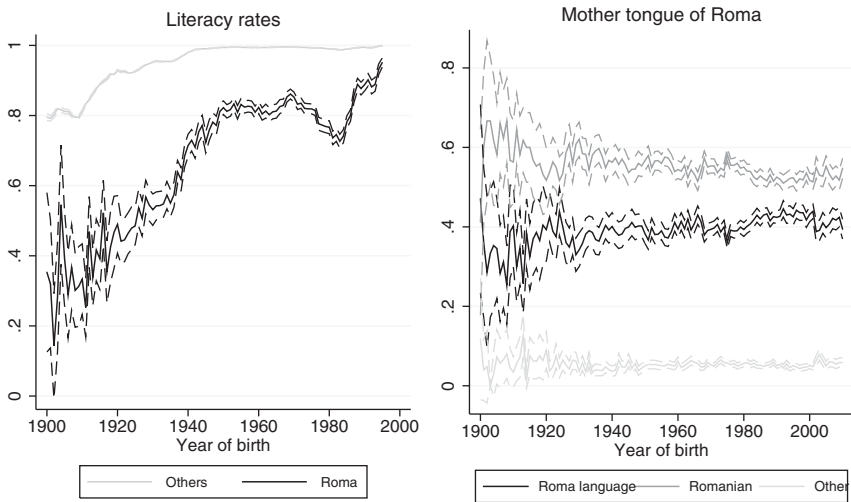
FIGURE 5. Literacy rates and mother tongue of Roma by year of birth. The dashed lines represent the 95% confidence interval. The left panel shows literacy rates by ethnicity and year of birth. The sample is restricted to individuals above the age of 15. The right panel shows the shares of the Roma population speaking different mother tongues by year of birth. Datasource: IPUMS International.

could stem from the traditional role of the male breadwinner and females' roles as housewives, therefore causing underinvestment in the education of female Roma.

3.2. Literacy and Language

The low educational attainment of Roma is also reflected in literacy rates. In the left panel of Figure 5, literacy rates and corresponding 95% confidence intervals by year of birth of individuals aged 16 and over of both population groups are plotted for all four censuses together. For Roma, literacy rates, indicated by the black line, stagnate around the 80% level for cohorts born between 1950 and 1980. Then, for those born around 1980, the literacy rate actually drops steeply to almost 70%. The last cohorts of the sample enjoy an upward trend again, and the graph indicates that for most recent cohorts the gap is nearly closed.

Non-Roma females (gray line in left panel of Figure B.3 of Online appendix) catch up to males at the beginning of the 1940s achieving nearly full literacy. However, for Roma females, indicated by the gray line in the right panel, the gap persists across all birth cohorts in the sample until close to the very end, as it seems that those born after 1985 have finally caught up to males.

We previously saw that having a Roma language as mother tongue reduces the probability of school attendance by 8.8 and 5.1 percentage points for female and male Roma, respectively. In the right panel of Figure 5, we can see that around 50%

of Roma of the most recent cohort still indicate that their mother tongue is a Roma language (black line), suggesting strong cultural transmission across generations. One reason contributing to dropout rates of Roma might be a misalignment of language of instruction and mother tongue of a large proportion of Roma. This is illustrated in Figure B.4 of Online appendix, where the school attendance rates by age in 2011 of Roma with Romanian as their mother tongue (gray line) strictly dominates the attendance rates of Roma who have a Roma language as mother tongue (black line). This has been recognized by the government, and efforts have been made to provide a number of special educational initiatives providing Roma children with a Roma language as language of instruction [ERRC and Danova-Rusinova (2004)].

4. LABOR MARKET

The low educational attainment of the Roma population of Romania is associated with worse outcomes in the labor market. Although there is no data on labor market outcomes for 1977, a similar gap as in education and literacy is observed for Roma in the labor market between 1992 and 2011. Although the entire population seems to have suffered between 1992 and 2002 due to the economic crisis, potentially attributable to mismanagement of privatization during the 1990s, both groups show improvements from 2002 to 2011.

In Figures B.5 and B.6 of Online appendix, I display employment (only salary/wage workers) and unemployment rates, respectively, for males (top) and females (bottom) for 1992 (left), 2002 (middle), and 2011 (right) over the lifecycle.[12] The gray line is for non-Roma, whereas the black line represents Roma. The dashed lines represent the 95% confidence interval. One can tell that for both males and females, there is a substantial (un)employment gap over most of the lifecycle. For males, the unemployment gap over the lifecycle has narrowed from 1992 to 2011, whereas for females it has widened substantially. In 2011, young Roma females face unemployment rates close to 80% compared to about 50% for young Roma males.

In order to estimate the disadvantage Roma experience in the labor market, I run logit regressions with employment excluding unpaid work [columns (1) and (2)], including only salary/work employment [columns (3) and (4)], and unemployment [columns (5) and (6)[as the dependent variables while restricting the sample to individuals aged 16–64 not enrolled in education.[13] If Roma disadvantages stem from reasons beyond their education and other observable characteristics, then policies would have to consider that addressing the educational gap might not be sufficient. I control for a vector of individual characteristics X_i including educational attainment, age, age squared, gender, literacy, and whether an individual lives in an urban or a rural area. Education is split into dummy variables for completion of primary, secondary, or tertiary schooling (with less than primary education forming the baseline category). Additionally, I control for county s and year t fixed effects, which are denoted as π_s and ϕ_t, respectively. The estimation

of the probability of an individual i being (un)employed ((un)emp$_i \in \{0, 1\}$) takes the form $Prob((\text{un})\text{emp}_i | X_i, \text{Roma}_i, \pi_s, \phi_t) = h(X_i, \text{Roma}_i, \pi_s, \phi_t)$.

The coefficient of the Roma dummy is (positive) negative and significant at the 1% level for (un)employment in all model specifications exhibited in Table 3. The exhibited marginal effects indicate that Roma have a disadvantage in the labor market that goes beyond their lack of education and differing geographic location. For instance, in columns (1) and (3), we find that Roma suffer a 16 and 13 percentage-point employment deficit, respectively, whereas in column (5) we see that the unemployment surplus amounts to 7.2 percentage points. Relative to baseline rates of 57% or 48% for employment and 9.5% for unemployment, this amounts to a 29% or 27% lower rate of employment and 76% higher rate of unemployment.

In column (2) and (4), I interact the Roma dummy with year dummies in order to see the trend of the Roma penalty in employment. I find that in 2002, the Roma disadvantage in employment decreased relative to 1992, but this improvement is muted by 2011 (and has even worsened for salary/wage work). I also interact the Roma dummy with the female dummy, finding that a female Roma, conditional on all other controls, has an even lower probability (-9.9 or -6.2 percentage points) of being employed. Whether this is due to discrimination within Roma and/or the gender role in Roma families, where the male is assumed to be the breadwinner, or due to discrimination from others, cannot be distinguished with the available data. By interacting the Roma dummy with literacy, I find that the return to literacy in terms of employment is only about half of what it is for non-Roma (6.7 vs. 11.4 percentage points) for self-employment or salary/wage employment. Remarkably, when only looking at salary/wage employment in column (4), illiterate male Roma are actually more likely to be employed ($+3.4$ percentage points) than their non-Roma equivalent. However, for literate Roma, which form the great majority, the disadvantage remains large and significant.

4.1. Decomposition of Employment Gap and Differences in Occupations

In order to identify how much of the gap in terms of salary/wage employment can be explained by observable characteristics, I decompose it following Oaxaca (1973) and Blinder (1973). Following the adaptation for a logit model presented in Yun (2004) and Borooah and Iyer (2005), the two-fold Oaxaca–Blinder decomposition splits the difference in employment rates into the part explained by differences in endowments X of the individuals and the gap arising due to discrepancies between the coefficients γ of the regressors of the two population groups. The labor market differential can be expressed as

$$\underbrace{\bar{X}^{\text{Roma}}\hat{\gamma}^{\text{Roma}} - \bar{X}^{\text{other}}\hat{\gamma}^{\text{other}}}_{\text{Overall difference}} = \underbrace{(\bar{X}^{\text{Roma}} - \bar{X}^{\text{other}})\hat{\gamma}^{\text{other}}}_{\text{Explained difference}} + \underbrace{\bar{X}^{\text{Roma}}(\hat{\gamma}^{\text{Roma}} - \hat{\gamma}^{\text{other}})}_{\text{Unexplained difference}}$$

(1)

TABLE 3. Logistic regression estimating (un)employment

Dependent variable: Employed/unemployed

	Employment		Salary/wage emp.		Unemployment	
	(1)	(2)	(3)	(4)	(5)	(6)
Roma	−0.163***	−0.096***	−0.131***	0.034***	0.072***	0.082***
	(0.002)	(0.006)	(0.002)	(0.007)	(0.001)	(0.004)
Female	−0.146***	−0.144***	−0.108***	−0.107***	−0.018***	−0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female × Roma		−0.099***		−0.062***		0.012***
		(0.003)		(0.004)		(0.002)
Literate	0.104***	0.114***	0.141***	0.183***	−0.012***	−0.017***
	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)
Literate × Roma		−0.047***		−0.151***		0.011***
		(0.005)		(0.006)		(0.004)
Immigrant	−0.082***	−0.082***	−0.098***	−0.098***	0.000	−0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Urban	0.048***	0.048***	0.149***	0.149***	0.035***	0.035***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Age	0.062***	0.062***	0.059***	0.059***	−0.019***	−0.019***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age squared	−0.001***	−0.001***	−0.001***	−0.001***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Roma 2002		0.044***		0.013***		−0.033***
		(0.004)		(0.005)		(0.003)
Roma 2011		−0.001		−0.013***		−0.033***
		(0.004)		(0.004)		(0.003)
Education completed						
Primary	0.001	0.000	0.098***	0.097***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Secondary	0.133***	0.133***	0.257***	0.257***	−0.002***	−0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tertiary	0.328***	0.328***	0.411***	0.410***	−0.059***	−0.059***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Sample means	0.570	0.570	0.479	0.479	0.095	0.095
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,850,069	3,850,069	3,850,069	3,850,069	2,666,019	2,666,019
R^2	0.215	0.216	0.282	0.282	0.103	0.104

Notes: Robust standard errors in parentheses. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Reported coefficients are marginal effects. The dependent variables are employment (either salary/wage workers or self-employed) in columns (1) and (2), salary/wage employment in columns (3) and (4), and unemployment in columns (5) and (6). The sample includes all individuals aged 16–64 not enrolled in education. County and year fixed effects and a constant term are included in all specifications. Baseline categories for dummy variables are the following: less than primary schooling (education), male (*Female*), non-literate (*Literate*), rural (*Urban*), non-Roma (*Roma*).

where $\bar{X}^{\text{Roma}}$ and $\bar{X}^{\text{other}}$ are the average attributes of the Roma and non-Roma, whereas $\hat{\gamma}^{\text{Roma}}$ and $\hat{\gamma}^{\text{other}}$ are the coefficients estimated from separate regressions for Roma and non-Roma, respectively. The left-hand side of equation (1) is the overall difference, which is displayed in the third row of Table 4. The first term on the right-hand side estimates the employment gap that can be attributed to differences in observable characteristics between the two groups and is decomposed in the middle block, whereas the second term is the residual that captures the effect that could be due to discrimination or cultural differences in the bottom block of Table 4. The rows within the middle and bottom block sum up to the explained and unexplained share displayed in rows four and five, respectively.

The first three columns are dedicated to each year separately of a comparison between non-Roma (group 1) and Roma (group 2), whereas the last three columns compare across gender within Roma, i.e., Roma males (group 1) to Roma females (group 2). The difference in employment between non-Roma and Roma has remained fairly constant around 30 percentage points. However, the share of this gap explained by observables has increased from 58% in 1992 to 72% in 2011. Although this could be interpreted in a good sense as a potential decrease in discrimination in the labor market, given that now observables can account for nearly three times as much as in 1992, the fact that the actual gap is nearly constant is worrying. Education, which I define as highest educational attainment as well as a binary variable for literacy, is responsible for most of the explained gap. For instance, in 2011 education accounts for 21.5 of the 22.2 explained percentage-point gap (97%).

Restricting the sample to only Roma, the explained part of the salary/wage employment gap between males and females has decreased from nearly 20% in 1992 to 13% in 2011, as can be seen in the last three columns of Table 4. Females' disadvantages in the labor market are larger than observables suggest. The gap does not seem to be attributable to motherly obligations as including the number of children in the household hardly affects the results. The good news is that the gender gap has decreased from 23 percentage points in 1977 to 12 percentage points in 2011. However, this narrowing seems to have more to do with reduced employment rates of males rather than an increase in female employment.

As can be seen in Table 5, Roma are clustered in blue collar occupations (nearly 90%).[14] In contrast, the rest of the population increased its share of white collar occupations from 30% in 1992 to 41% in 2011. The largest share of Roma work in elementary occupations, traditionally low skilled with low remuneration. The differential distribution across occupational categories is reflected by the Duncan dissimilarity index, where 0 indicates perfect similarity and 1 indicates perfect dissimilarity in occupations. Noticeably, this index increased from 0.37 in 1992 to 0.53 in 2002, before declining to 0.32 in 2011, meaning that in 2011 32% of workers would have to change occupation in order to equalize distributions.

TABLE 4. Oaxaca–Blinder decomposition of employment

	Across ethnicity Non-Roma vs. Roma			Within Roma Male vs. female		
Dependent variable: Salary/wage employment						
	(1992)	(2002)	(2011)	(1992)	(2002)	(2011)
Overall						
Group 1	0.553***	0.422***	0.467***	0.378***	0.189***	0.216***
	(0.000)	(0.000)	(0.000)	(0.005)	(0.003)	(0.003)
Group 2	0.261***	0.138***	0.157***	0.145***	0.087***	0.097***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Difference	0.292***	0.284***	0.310***	0.234***	0.102***	0.119***
	(0.003)	(0.002)	(0.002)	(0.006)	(0.004)	(0.004)
Explained	0.170***	0.198***	0.222***	0.046***	0.020***	0.016***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)
Unexplained	0.122***	0.086***	0.088***	0.187***	0.082***	0.104***
	(0.003)	(0.002)	(0.002)	(0.006)	(0.004)	(0.004)
Explained						
Education	0.163***	0.210***	0.215***	0.042***	0.016***	0.013***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Age	− 0.024***	− 0.032***	− 0.019***	0.006***	0.002***	0.003***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Female	− 0.001	− 0.000	− 0.001**			
	(0.001)	(0.000)	(0.000)			
Roma language	0.011***	− 0.008***	− 0.010***	− 0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Urban	0.021***	0.028***	0.036***	− 0.001	− 0.000	− 0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
# of own children in HH				− 0.000	0.001***	0.001***
				(0.000)	(0.000)	(0.000)
Unexplained						
Education	0.197***	0.321***	0.171***	0.081*	− 0.021	− 0.146***
	(0.040)	(0.032)	(0.024)	(0.046)	(0.025)	(0.029)
Age	0.485***	0.564***	0.243***	− 0.273***	− 0.042	− 0.191***
	(0.037)	(0.041)	(0.025)	(0.048)	(0.036)	(0.035)
Female	0.015***	0.044***	0.030***			
	(0.004)	(0.004)	(0.002)			
Roma language	0.026***	0.016***	0.027***	0.006	− 0.002	0.010***
	(0.003)	(0.003)	(0.002)	(0.005)	(0.003)	(0.003)
Urban	0.028***	0.073***	− 0.016***	− 0.038***	− 0.011***	0.036***
	(0.003)	(0.004)	(0.002)	(0.004)	(0.003)	(0.003)
# of own children in HH				0.048***	0.016***	0.021***
				(0.006)	(0.004)	(0.004)

TABLE 4. Continued

	Dependent variable: Salary/wage employment					
	Across ethnicity Non-Roma vs. Roma			Within Roma Male vs. female		
	(1992)	(2002)	(2011)	(1992)	(2002)	(2011)
Constant	−0.630***	−0.932***	−0.366***	0.363***	0.142***	0.375***
	(0.059)	(0.059)	(0.038)	(0.066)	(0.043)	(0.044)
Share of gap explained	0.582	0.697	0.716	0.197	0.196	0.134
Observations	1,349,766	1,297,507	1,135,075	20,478	29,471	35,282

Notes: Standard errors computed using the delta method in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variable is salary/wage employment, i.e. excluding unpaid and self-employment. Group 1 are non-Roma and group 2 are Roma in the first three columns, whereas group 1 are Roma males and group 2 are Roma females in the last three columns. The sample includes all individuals aged 16–64 not enrolled in education. *Education* contains the aggregated effect of a dummy for literacy as well as categorical dummies for completion of primary, secondary, and tertiary education. *Age* contains both age and age squared.
Datasource: IPUMS International.

TABLE 5. Share of Roma and non-Roma by occupational category

	1992			2002			2011		
Occupation	R	O	Δ	R	O	Δ	R	O	Δ
White collar									
Legislators, senior officials, managers	0.00	0.02	−0.02	0.01	0.03	−0.02	0	0.03	−0.03
Professionals	0.00	0.08	−0.08	0.01	0.12	−0.11	0.03	0.21	−0.18
Technicians and associate professionals	0.01	0.14	−0.13	0.01	0.15	−0.14	0.08	0.11	−0.03
Clerks	0.02	0.06	−0.04	0.01	0.07	−0.06	0.02	0.06	−0.04
Blue collar									
Service workers, shop and market sales	0.03	0.06	−0.03	0.07	0.12	−0.05	0.15	0.18	−0.03
Agricultural and fishery workers	0.07	0.02	0.05	0.14	0.02	0.12	0.04	0.01	0.03
Crafts and related trades workers	0.30	0.34	−0.04	0.17	0.26	−0.09	0.23	0.19	0.04
Plant and machine operators, assemblers	0.17	0.2	−0.03	0.08	0.14	−0.06	0.16	0.11	0.05
Elementary occupations	0.39	0.07	0.32	0.49	0.09	0.40	0.30	0.09	0.21
Armed forces	0.01	0.01	0.00	0.01	0.01	0.00			
Duncan dissimilarity index		0.37			0.53			0.32	

Notes: This table presents the share of salary/wage employed Roma and non-Roma by occupation in each year. The columns headed by "R" are for Roma, whereas "O" represents others. Δ is the difference of the two. The Duncan dissimilarity index represents the share of workers that would have to change occupation in order to equalize distributions. Therefore, 0 indicates perfect similarity and 1 indicates perfect dissimilarity in occupations.
Datasource: IPUMS International.

5. CONCLUSIONS

The Roma are a disadvantaged and understudied minority across Europe. I fill the gap concerning the empirical evidence by exploring the Romanian censuses of 1977, 1992, 2002, and 2011. I show that in Romania, the Roma form a rapidly growing but segregated community with persistently weak socio-economic outcomes and higher fertility rates. Battaglia et al. (2017) show that residential segregation is causally linked to higher fertility among Roma in Serbia suggesting that de-segregation could lower the high levels amongst the Roma to some extent.[15] The Roma lag behind the rest of the population in terms of all observed categories, be it dwelling characteristics, education, or the labor market. Even controlling for observables and parental characteristics, I find lower school attendance among the Roma, in particular among Roma girls. These attendance gaps are worrying for the future of the many young Roma in the Romanian population as they point to persistent inequality. Remedial programs, such as the Teaching Assistant Program in Serbia, have been shown to have positive effects on attainment [Battaglia and Lebedinski (2015)]. At least the literacy gap has been narrowing for both male and female Roma, both of whom are now close to universal levels for most recent cohorts.

I present descriptive evidence that lower rates of employment and higher rates of unemployment of the Roma cannot be explained by their lower levels of educational achievement or other observable characteristics. An Oaxaca–Blinder decomposition of employment suggests that Roma might be suffering from discrimination, as their outcomes are even worse than their qualifications and circumstances suggest, especially for female Roma. Given that Roma form a growing share of the population, this is not only important for ethical and equity considerations, but also for the rest of society, as the Roma will form a large potential tax base and source of productivity to the Romanian and European economy. More causal evidence is required to gain a better understanding of why Roma remain in such a disadvantaged position in order to design appropriate policy measures.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit https://doi.org/10.1017/dem.2017.19

NOTES

1 When Yale University researchers in 2000 tried different approaches when asking respondents about their ethnicity in Romania, 61% of those that the interviewer identified as Roma did not self-identify [Revenga et al. (2002)]. However, Ladányi and Szelényi (2001) provide evidence that labels by outsiders are unstable across time and tend to omit middle-class Roma, thereby potentially not providing a superior methodology. Although indeed self-reporting could induce issues through a selection bias, I provide evidence that it is stable across time. In Online appendix A, I show that the aggregate numbers as well as the level of education are consistent across the censuses.

2 Masseria et al. (2010) find that being a Roma is the main determinant of feeling threatened by illness due to unhygienic circumstances.

3 See Ivanov et al. (2006), O'Higgins and Ivanov (2006), and O'Higgins (2010).

4 According to the Council of Europe (2012), in relative terms, Romania has the fourth largest Roma community in Europe with 8.63% after Bulgaria (9.94%), Macedonia (9.56%), and the Slovak Republic (9.02%).

5 The only other samples available at IPUMS International, which contain "Roma" as an ethnic category, are Hungary (1980 and 1990) and Belarus (1999). Although these data points could provide additional snapshots, they do not allow me to document the evolution to the present day. Moreover, in the Hungarian census, the Roma identity is only available as an alternative to other nationalities, i.e., to immigrants not specifying the country they are from [Kertesi and Kézdi (2011a)]. Therefore, this presented study focuses on Romania.

6 Unfortunately, no information on income or financial wealth is available. The smallest geographical unit in the data is at county level of which 42 exist in Romania. In the census, Bucharest additionally is split into six sectors, such that the analysis considers 47 counties.

7 However, in Online appendix A, I provide evidence that neither the aggregate level of reporting nor the composition of those self-reporting seems to have changed since 1992.

8 Age 40 has been chosen assuming that by this age most of child-rearing has been completed.

9 These differences are not merely driven by differences in fertility as the gap remains when dropping children under the age of 18 from the sample. A gap also persists when only considering related individuals. Moreover, Roma households on average span more generations.

10 The results are robust to the inclusion of age dummies instead of the polynomial. The sample is restricted to 2011 in order to give an outlook on the future. For a large part of the generation of school-aged children in the 2002, sample educational achievement is already observable in 2011.

11 Battaglia and Lebedinski (2014) find that parental perceived returns of Roma in Serbia are low and respond positively to a Roma Teaching Assistant Program.

12 In Figure B.7 of Online appendix, I add the distribution of type of employment (salary/wage employment, self-employment, unpaid) for all Roma (conditional on employment).

13 By restricting the sample to only salary/work employment, the estimates are more likely to reflect actual labor market associations, as informal work and self-employment are likely to be based on subjective broad definitions. Nonetheless, measurement error might still be an issue due to self-reporting. The results are very similar when restricting the sample to individuals aged 25–64 and are available upon request.

14 Occupations were categorized into blue and white collar according to classifications in http://www.eurofound.europa.eu/surveys/ewcs/2005/classification.htm.

15 Although social differences play the greatest role, Kertesi and Kézdi (2016) also hint at residential segregation as an important factor for the unequal distribution of Roma students across schools and classes.

REFERENCES

Battaglia, Marianna and Lara Lebedinski (2014) The Curse of Low Aspirations: Remedial Education and Perceived Returns to Education of Roma People. Instituto Valenciano de Investigaciones Económicas, SA (Ivie). Working papers no. WP-AD 2014-04.

Battaglia, Marianna and Lara Lebedinski (2015) Equal access to education: An evaluation of the Roma teaching assistant program in Serbia. *World Development* 76, 62–81.

Battaglia, Marianna et al. (2017) Segregation and fertility: The case of the Roma in Serbia. Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES). Discussion Paper 20177-11.

Beck, Sam (1984) "The Romanian Gypsy Problem." In Papers from the Fourth and Fifth Anuual Meetings, Gypsy Lore Society, North American Chapter, edited by Joanne Grumet, pp. 100–109. New York: Gypsy Lore Society 1985.

Blinder, Alan S. (1973) Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* 8(4), 436–455.

Borooah, Vani K. and Sriya Iyer (2005) The decomposition of inter-group differences in a logit model: Extending the Oaxaca-Blinder approach with an application to school enrolment in India. *Journal of Economic and Social Measurement* 30(4), 279–293.

Brüggemann, Christian (2012) Roma education in comparative perspective. In *Analysis of the UNDP/World Bank/EC Regional Roma Survey*. Bratislava: UNDP.

Council of Europe (2012) Estimates on Roma population in European countries. rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680088ea9

Crowe, David (2007) *A History of the Gypsies of Eastern Europe and Russia*. New York: Palgrave Macmillan US.

Cukrowska, E. and A. Kóczé (2013) *Interplay between Gender and Ethnicity: Exposing Structural Disparities of Romani Women*. Analysis of the UNDP/World Bank/EC regional Roma survey data. Roma Inclusion Working Papers. Bratislava: UNDP.

De Laat, Joost (2010) *Economic Costs of Roma Exclusion*. Berlin: World Bank.

ERRC and Savelina Danova-Rusinova (2004) *Stigmata: Segregated Schooling of Roma in Central and Eastern Europe*. Budapest: European Roma Rights Center (ERRC).

Gilberg, Trond (1974) Ethnic minorities in Romania under socialism. *East European Quarterly* 7(4), 435.

Ivanov, Andrey, Mark Collins, Claudia Grosu, Jaroslav Kling, Susanne Milcher, Niall O'Higgins, Ben Slay and Antonina Zhelyazkova (2006) *At Risk: Roma and the Displaced in Southeast Europe*. Bratislava: United Nations Publications.

Kertesi, Gábor and Gábor Kézdi (2011a) Roma employment in Hungary after the post-communist transition. *Economics of Transition* 19(3), 563–610.

Kertesi, Gábor and Gábor Kézdi (2011b) The Roma/non-Roma test score gap in Hungary. *American Economic Review* 101(3), 519–525.

Kertesi, Gábor and Gábor Kézdi (2016) On the test score gap between Roma and non-Roma students in Hungary and its potential causes. *Economics of Transition* 24(1), 135–162.

Ladányi, János and Iván Szelényi (2001) The social construction of Roma ethnicity in Bulgaria, Romania and Hungary during market transition. *Review of Sociology* 7(2), 79–89.

Liégeois, Jean-Pierre (2007a) Roma education and public policy: A European perspective. *European Education* 39(1), 11–31.

Liégeois, Jean-Pierre (2007b) *Roma in Europe*. Strasbourg: Council of Europe.

Masseria, Cristina, Philipa Mladovsky and Cristina Hernández-Quevedo (2010) The socio-economic determinants of the health status of Roma in comparison with non-Roma in Bulgaria, Hungary and Romania. *European Journal of Public Health* 20(5), 549–554.

McKee, Martin (1997) The health of gypsies. *British Medical Journal* 315(7117), 1172.

Minnesota Population Center (2017) Integrated Public Use Microdata Series, International: Version 6.5 [dataset]. Minneapolis.

Oaxaca, Ronald (1973) Male-female wage differentials in urban labor markets. *International Economic Review* 14(3), 693–709.

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* 31(2), 163–187.

O'Higgins, Niall and Andrey Ivanov (2006) Education and employment opportunities for the Roma. *Comparative Economic Studies* 48(1), 6–19.

O'Higgins, Niall and Christian Brüggemann (2014) The consequences of cumulative discrimination: How special schooling influences employment and wages of Roma in the Czech Republic. *European Educational Research Journal* 13(3), 282–294.

Pan, Christoph and Beate Sibylle Pfeil (2003) *National Minorities in Europe: Handbook*, vol. 63. Braumüller, Vienna: Braumüller.

Rat, Cristina (2005) Romanian Roma, state transfers, and poverty: A study of relative disadvantage. *International Journal of Sociology* 35(3), 85–116.

Revenga, Ana, Dena Ringold and William Martin Tracy (2002) *Poverty and ethnicity: A Cross-Country Study of Roma Poverty in Central Europe*, vol. 531, Washington, DC: World Bank Publications.

Ringold, Dena, Mitchell Alexander Orenstein and Erika Wilkens 2005. *Roma in an Expanding Europe: Breaking the Poverty Cycle*. Washington, DC: World Bank Publications.

Yun, Myeong-Su (2004) Decomposing differences in the first moment. *Economics Letters* 82(2), 275–280.