The sorting into ethnic identities: Social status and Roma identification in Hungary

Gabor Simonovits* and Gabor Kezdi**

Introduction

Numerous studies point to the immense gap in the standard of living between the Romani and members of the ethnic majority in Central and Eastern Europe (CES) in general and in Hungary in particular. Members of the Romani[1] minority lag behind in all area: there are enormous differences in the school (see Kertesi and Kezdi, 2011b), the labor market and living conditions in general. On the other hand, scholars of discrimination point to the high level of prejudice against the Romani in Hungary and elsewhere. Thus, on the top of poverty and unemployment, the Romani need to get along in the face of hatred and discrimination. The question we seek to answer in this paper is how ethnic identity develops in the context of such marginalization and discrimination. In particular, we would like to see if identification with a marginalized ethnic minority involves any sort of voluntary decision, and if it does, what are the motives of such decision.

While there is some existing empirical research on Roma identity (see Csepeli and Simon, 2004 Koulish, 2005; Prieto-Flores, 2009; Ladanyi and Szelenyi, 2001, 2006) some methodological issues render these studies difficult to inform research on the development of Roma identity. These concerns include the non representative sampling used on these studies, problems associated with the measurement of

* Stanford University, Department of Political Science
** Central European University, Department of Economics

[1] There is some controversy about the name of the Romani ethnic group. In Central and Eastern Europe, the name Roma is used both as a noun (plural: Roma) and as an adjective. It is also used by some international organizations and initiatives, such as the Roma Education Fund. The United Nations, the US Library of Congress and other international associations use the term 'Romani' as an adjective and a noun as well. The name 'Gypsy' is used by many non-Roma, but not by the Roma; it is a name created by outsiders and is derived from the misconception of Egyptian origin.. In this article, we use Roma and Romani interchangeably.

Roma ethnicity and the impossibility to make claims about dynamic processes using cross sectional data.

The goal of our research is to fill this gap and explore the links between social status and Roma both across and within generations. Our approach is based on some recent studies on ethnic and racial identification in the US, Brazil and Germany (see Penner and Saperstein, 2008; Schwartzman, 2007 and Casey and Dustmann, 2010 respectively) that use either panel data about social status and ethnic identification or matched cross sections of families. To our knowledge, our unique panel survey that follows 10000 families for 6 years is the first dataset that allows for the careful description of how socio-economic status (SES) and the environment shapes Roma ethnic identities.

Our results are consistent with recent findings reported in other contexts (Davenport, 2012; Penner and Saperstein, 2008). We find evidence that suggests that Romani ethnicity is a continuous construct that is imperfectly measured in cross sectional surveys. Our analysis about a matched of adolescents and their parents reveals that social status and the ethnic composition of the network in which adolescents grow up has an important role in conditioning the intergenerational transmission of ethnic identity. While in our cross sectional analysis it is very hard to distinguish between the importance of substantive and measurement error based explanations, our longitudinal results supports the former. It appears that changes in social status have genuine effects on self reported ethnicity, with social marginalization leading to an increased likelihood of Romani identification.

The remainder of the paper is structured as follows. The next section describes the case of our analysis: the Roma population of Hungary. We first sketch the history of the minority group, then describe their position in the contemporary Hungarian society with regards to their social status and the discrimination against them. The third section describes our theoretical framework about assimilation and changes in ethnic membership and points to related literature. The fourth section discusses the

merits and shortcomings of our data: we describe the sampling process and the operationalization of the main variables of interests. The fifth section presents our results: first we show some descriptive statistics and then present our main findings. The last section concludes.

The Romani minority in Hungary

The Romani people are one Europe's largest ethnic minorities. While the size and living conditions of the Roma population are both hard to assess, because the general lack and quality of data, the number of Romani in Central and Eastern Europe was estimated to be over 4 million in the early 1990s (Barany, 2002). In Hungary the size of the Roma minority is estimated to be around 5-6% (Kemény and and Janky, 2006). In spite of the uncertainty of population sizes, sources agree about the widespread poverty, low formal employment, low education, poor health and social exclusion in all countries (O'Higgins and Ivanov, 2006; Milcher, 2006; Ringold et al., 2005; UNDP, 2002).

The integration and assimilation of the Roma remained limited throughout modern history. They lived outside mainstream society: they did not own land and most of them made a living as day laborers or sold their own products and services. While the Industrial Revolution brought the Roma minority closer to mainstream society, it also undermined their traditional communities. The Second World War saw members of the Roma minority decimated by deportations and mass executions, similar to the Jews. The communist regime continued the dissolution of local Roma communities and brought about a process of forced assimilation.

The fall of the communist system led to a deep recession and a thorough transformation of labor market in Hungary: while the economy started to grow quickly during the mid-1990s, the demand for unskilled labor collapsed. As a consequence many of the unskilled people who were laid off after the introduction of the market economy have been left unemployed ever since. It is widely accepted (Kemeny and Janky, 2006, Kertesi and Kezdi, 2011a) that the transition to market

economy associated with a dramatic drop in demand for low-skilled workers affected the Roma especially severely (Kertesi and Kezdi, 2011a). Many believe that the primary cause of this gap is the enormous difference between the schooling level of Romani and the ethnic majority.[2] Children born into families ridden with unemployment and poverty, struggling to get along in low quality (and often ethnically segregated) schools end up in low status jobs, or more often unemployed.

Widespread prejudice and discrimination (Kligman, 2001) adds to the miseries of the Roma minority. The majority of Hungarians hold strong prejudices against the Roma (Fabian and Sik, 1996): the associate the Roma with laziness and crime; and tend to see the Roma as responsible for their misfortunes and many are skeptical about the prospect of Roma integration. There is also evidence of discrimination against the Roma in the in law enforcement (Farkas et al, 2004), the coursts (Loss, 2001) and the labor market (Palosi et al., 2007).

The late 2000s saw ethnic tensions elevated even further: murder cases with apparently racist motives committed both by and against Romani were highly publicized. The Hungarian Guard (*Magyar Gárda*), a paramilitary organization, characterized by many as Neo Nazi held marches in localities experiencing conflicts between Roma minority and the ethnic majority. In the general elections of 2010 Jobbik, a newly established extreme right wing party received about 17% of the votes and 10% of the seats in the new parliament. Jobbik was closely associated with the Guard, and its MPs frequently made blatantly racist statements in public. The words of Gyorgy Csepeli (2011), a prominent Hungarian sociologist, are telling: "There is nothing left to be researched. Repeated polls do not provide new information. The tension between Roma and non-Roma has reached its climax. It is high time to act." (p. 86)

[2] The gap in education is striking – even compared to the Black – White gap it the US. 2003, 16% of Roma adults completed secondary (or higher) degrees as compared to 74% among the ethnic majority; 3 in 1000 roma adults had college degrees, against 18% in the general population (Kemeny and Janky, 2006).

Theoretical background and related literature

Despite the prevalence of its use, ethnic identity does not have a generally accepted definition (Margalit and Kuo, 2012). The primordial notion of ethnic identity (Shils, 1957) that views identity as an innate and fixed category has largely been rebutted but there are still lively debates about how identity is developed and shaped. In the theoretical framework that describes ethnicity as a (potentially contested) social construct (see Barth, 1969)[3], two important questions emerged. First, whether changes in identity for instrumental reasons are possible, and second what is the role of situations and social contexts in triggering (or making salient) different identities (Margalit and Kuo, 2012).

For our purposes the important question is the following. If Roma ethnicity is *not* innate and fixed, what are the conditions under which individuals can and do leave their Roma identity behind? In short, what are the causes of ethnic assimilation? As Laitin (1995, 1998) argues, assimilation should be analyzed as an individual level phenomenon, in which members of a minority group make decisions in the face of cost and benefits associated with leaving their groups. The benefits of leaving Romani identity behind can be substantial in a society where the Romani as an ethnic group are associated with poverty and crime (Fabian and Sik, 1996) and at the same time their ethnic markers are lighter and their vast majority speak the language of the host country.

On the other hand, while ethnic assimilation could be attractive for a number of reasons we do not observe a rapid decline in the number of Romani. One reason for this (apart from the obvious claim that individuals simply do not want to deny their ethnic background) is the intertwinement of ethnicity and social class (Farmoso, 1986). In order to leave the ethnicity behind, Romani would need to climb higher on the social ladder, which is hard for structural reasons outlined above. Nevertheless,

[3] See also Brubaker (2002), Vermeersch (2003), Brubaker, Loveman and Stamatov (2004), Hale (2004) and Wimmer (2008).

it is still possible that we can observe such changes in social status and ethnic identification if we look at the right place.

For existing sociological studies some methodological problems have made virtually impossible to make inferences about the effect of social status on Roma identity. First, since the size of the Romani population varies around 5-10% in CEE countries (Csepeli and Simon, 2004), in order to collect a large enough sample of Romani individuals, researchers needed to use methods that render their sample non-representative of the Roma population. These studies (Ladanyi and Szelenyi, 2001, 2006; Csepeli and Simon, 2004; Koulish, 2005; Prieto-Flores, 2009) either used the help of experts to gather their sample, or based their sample on screening items in national surveys[4]. Thus the samples used in existing studies only include individuals whom *others* regard as Romani. This approach is problematic not only because it individuals from the sample who are potentially Roma, but also because it prevents making causal claims[5].

A related issue is about the dynamics of the process of identity formation. Assimilation and social mobility are obviously dynamic concepts. They could refer to the change of ethnicity and/or social status during the course of a lifetime, or in an intergenerational sense. A straightforward implication is that ideally, the empirical study of social mobility, assimilation and the relationship between the two should involve either repeated measurement (i.e. following a set of individuals over time) or a cross section of families with data on ethnic identity about different generations. To our knowledge, no such research has been conducted and no such data have been collected in the context of the Romani, so we aim to fill this gap with our paper.

[4] These screening items are usually ethnic classifications by the interviewer.

[5] For instance when researcher observe correlation (or the lack of thereof) between Romani self-identification and SES, they fail to reflect on the fact that this implies that their sample is already selected on SES (i.e. rich Roma will not say that they are Roma during the screening, so they do not even make it to the sample).

While for the reasons described below, the systematic analysis of identity formation has been problematic in the context of the Romani, recent research using new data has found some convincing evidence elsewhere. In an influential study, Penner and Saperstein (2008) use the National Longitudinal Study of Youth (NLSY) to show that respondents who had been incarcerated, experienced poverty or lost their job were more likely to identify as Black, even when baseline racial identity is controlled for. Their evidence suggests that "racial propensities can be altered by changes in social position" (p. 19629). An important consequence of that, raised also by the same authors is "that racial stereotypes can become self-fulfilling prophesies: not only does race shape social status, but social status shapes race." (p. 19629).

As for the intergenerational transmission of identity, recent studies by Davenport (2012) and Schwartzman (2007) look at the determinants of ethnic and racial identity of adolescents in mixed-race families in the US and Brazil respectively. Davenport (2012) finds that "children's chosen racial labels are largely a reflection of socioeconomic status, neighborhood context, and religion" (p. 1). Schwartzman (2007) shows that "highly educated nonwhite parents are more likely to classify their children as white than are comparable less-educated nonwhite parents".

Finally, related research in Europe looks at self-identification patterns of first and second-generation immigrants. In particular, Manning and Roy (2010) finds that economic social status does not seem to explain national identities of immigrants in the United Kingdom. Casey and Dustmann (2010) reports similar results based on panel data from Germany. Bisin et al (2011) reports interesting findings on the effect of neighborhood ethnic composition on ethnic identity. The paper show an inverted U shaped relationship between the density of own-ethnic group members in a neighborhood and ethnic identification as measured by the importance of religion and intra group marriage.

After pointing out problems in existing studies of Romani identity and discussing some novel results design and results on identity in different contexts, we now describe our own approach. In the next section we first describe the dataset with

special attention on sampling and the definition of the main variables in the analysis, then describe our empirical strategy and report some descriptive statistics.

Data and empirical strategy

Our analysis uses data from the Hungarian Life Course Survey (HLCS) of the TARKI Research Institute. The HLCS is an individual panel survey administered yearly that follows the model of the National Longitudinal Surveys of Youth in the United States (NLSY79). The original sample is 10,000 students drawn from the population of 8th – grade in 2006. The survey follows these students through their adolescence and interviews them yearly. We use data from six survey waves. The first survey wave was conducted in the fall of 2006, and the last survey wave was conducted in the spring and summer of 2012.

The first two survey waves included questions on the parents' ethnicity, while the second, fourth, fifth and sixth wave included questions on the respondents' ethnicity. Each time and for each individual ethnicity was measured by asking the following two questions. "*In our country people belong to different minorities and ethnic groups. What ethnic group do you consider to belong to primarily? And what ethnic group do you consider to belong to secondarily*?" Respondents could choose between multiple categories, among them Hungarian and Romani. By asking two questions, the survey allowed respondents to state multiple identities. In each survey wave we consider self-identified Roma those who chose Romani as an answer to either of the two questions. Besides self-identification, the first survey wave included a question to the interviewer on whether each parent and the adolescent target person was Roma.

The format of the identification questions was the same in each survey wave. The timing of the question was the same as well: at the beginning of the question sequence for the parents and towards the end of the questionnaire for the adolescent respondents. In wave D, the identity questions were asked earlier and were preceded by a series of questions on social network, in which respondents were asked repeatedly of whether members of their social networks were Romanies.

The other survey waves did not contain questions on the social network, and the identity questions there were preceded by more neutral questions[6].

As the focus of our analysis is on the intergenerational transmission of ethnic identity, our sample consists of respondents who lived with both of their biological parent in the first two survey waves. Table 2 shows the time of each survey wave, the median age of the adolescent survey respondents in each wave, the number of observations, and whether the survey wave contained measures of ethnicity of the parents and the adolescent.

Table 2. Measures of ethnic identity, time of the interview, age of the adolescent at the interview and number of observations in each survey wave of the Hungarian Life Course Survey.

Survey wave	Measures of ethnic identity		Median time of interview	Median age of the adolescent	Number of observations	
	of the parents	of the adolescent			entire sample	lived with both biological parents in wave A and B
A	Yes		Nov. 2006	15	10,011	6,181
B	Yes	Yes	Nov. 2007	16	9,000	5,997
C			Nov. 2008	17	8,648	5,792
D		Yes	Jan. 2010	18	8,100	5,490
E		Yes	Feb. 2011	19	7.621	5,215
F		Yes	June 2012	20	6,974	4,816

For most of the analyses, we work with the sample of adolescent respondents that had non-missing identity measures from each of the four survey waves in which the questions were asked. The size of that sample is 6,394. For the analysis of the intergenerational transmission of ethnic identity, we further restrict our sample to the 4,461 respondents who lived with both of their biological parents in waves A and B. As the survey oversampled students with low test score results and students of special educational need, we use sampling weights throughout the analysis to restore the representative nature of the survey.

[6] On labor market participation (wave B) and participation in training programs (waves E and F).

Table 3 shows the distribution of the observations in the complete sample among patterns of participation and patterns of valid ethnicity answers. There is considerable attrition in the panel. From the 10,022 respondents interviewed in wave A, 753 dropped out between waves A and B, 893 dropped out between waves B and D, 464 between D and E, and 709 between E and F. As a result, the number of respondents in all four waves in our analysis (B, D, E and F) and the first wave (A) is 6,257. Of them, 6,394 gave valid response to all identity questions, and 4,461 lived with both of their biological parents in waves A and B.

Table 3. Distribution of the participation in survey waves A, B, D, E and F. Entire original sample of 10,022 observations, subsample with valid answer to the ethnicity questions (asked in waves B, D, E and F) and further restricted to those living with both biological parents in wave A and B.

Survey wave pattern	Number of observations		
		With valid answer to ethnic identity	
	Participating	All	Lived with two biological parents in waves A and B
A B D E F	6,527	6,394	4,461
A B D E only	709	792	484
A B D only	464	486	282
A B only	893	896	504
A only	753	-	-
Other patterns*	676	698	423
Total	10,022	9,266	6,154

* Other patterns mean gaps in participation, such as A, B, D and F, etc.

Attrition from the survey is nonrandom. Table 4 shows the mean values of the SES score[7] and interviewer identified Roma variables by the panel categories. According to the figures, those who were identified as Roma by the interviewer in wave A and those with lower SES score are more likely to drop out of the survey at every stage.

[7] We used several variables on parental education, parental employment history and family income to create a one-dimensional measure of socio-economic status. See more on this below.

By restricting the sample to those who participated in all survey waves the results are representative for those with only slightly higher socio-economic status than the entire population. Living with both biological parents is also related to socio-economic status: those with biological parents have substantially higher SES score than those without (the "neither" category in the table). Within each category of survey wave pattern, those living with both biological parents have significantly higher SES scores. However, Roma identification by the interviewer is very weakly related to living with both biological parents. Thus, further restricting the sample to adolescents with two biological parents makes the results relevant for a higher socio-economic status population but one that is not very different in terms of Roma representation.

Table 4. Mean SES score and the fraction of adolescent respondents identified as Roma by the interviewer in wave A by participation and living with both biological parents.

	Mean SES score		Per cent identified as Roma by interviewer in waves A	
Survey wave pattern	Participant	With valid answer to ethnic identity and lived with two biological parents in waves A&B	Participant	With valid answer to ethnic identity and lived with two biological parents in waves A&B
B D E F	0.05	0.45	8.2	8.0
BDE only	-0.07	0.42	8.5	7.3
BD only	-0.09	0.34	11.9	11.0
B only	-0.11	0.36	13.2	12.4
A only	-0.03	-	9.4	-
Other	-0.18	0.27	10.8	11.0
Total	0.00	0.42	9.1	8.7

There is considerable variation in ethnic identification across survey. Table 5 shows the fraction of respondents or the fraction of the parents of the respondents who identified themselves as Roma in each survey wave. There is a slight decrease in

Roma identification of the parents from wave A to wave B, and there is an increase of Roma identification of the adolescent respondents from wave B to subsequent waves. The fraction of adolescent respondents who identified themselves as Roma is highest in wave D. The difference in the timing of the identity questions seems responsible for the particularly high fraction in wave D[8]. We do not have an explanation for the low fraction of Roma identification in wave B.

The fraction of respondents who identified themselves as Roma in any of the survey waves is substantially higher than the fraction who identified themselves as Roma in a particular wave. The difference is largest for the adolescents who had more opportunity to identify themselves as Roma. The fraction of Roma by interviewer identification in wave A is substantially higher than the fraction of self-identified Roma in any given survey wave, but it is somewhat smaller than the fraction of adolescents who identified themselves as Roma at least once in the survey waves combined.

Table 5. Fraction of Roma in each survey wave.

	Fraction Roma (per cent)		
	mother	father	adolescent
Self-identification			
In wave A	5.5	5.3	
In wave B	5.4	4.9	4.5
In wave D			6.0
In wave E			5.6
In wave F			5.7
At least once in any of the survey waves	6.8	6.3	8.3
Interviewer identification in wave A	8.2	7.9	8.0

Sample: Hungarian Life Course Survey. Subsample of adolescent respondents for whom we have non-missing identity measures in waves B, D, E and F. Observations weighted by sampling weights. N = 6,394

[8] This is supported by the fact that the fraction of Roma in the respondents' social network, measured in wave D, is substantially more strongly correlated with Roma identification in wave D than in the other waves.

Our dataset allows for analyzing many variables together with measures of ethnic identification. These include the age and gender of the adolescent respondent, standardized test scores in reading and mathematics measured in grade 8 (before the first survey wave), the complete educational and labor market history of the adolescent respondent, the education attainment and the labor market history of the parents (going back to the time of birth of the adolescent respondent), measures of family income in real terms at household equivalent scale (OECD) as well as measures of the ethnic composition of the network of the adolescent respondents. As indicated above, the network of the respondents was mapped out in wave D by a name generator method (plus some additional questions), and the respondents were asked for indicating if each member of their network was Roma. Moreover, we have data about the ethnic composition of the adolescents neighborhood, school and class, based on information provided by the parents, the school and the interviewer.

In some of the analysis we use a standardized score for the socio-economic status (SES) of the respondents' families. The score combines the log of household equivalent family incomes (separately for all survey waves), household size, the fraction of adults in the household without employment, the number of books in the household (a dummy for zero books and the log of books if positive), dummies for the education of the mother and the father, and the fraction of years the mother and the father worked between the birth of the adolescent respondent and the first survey wave. We combined these measures by estimating a probit regression, predicting the score of the regression and standardizing it using its weighted mean and standard deviation.

Figure 1: The estimated distribution of SES, by ethnicity

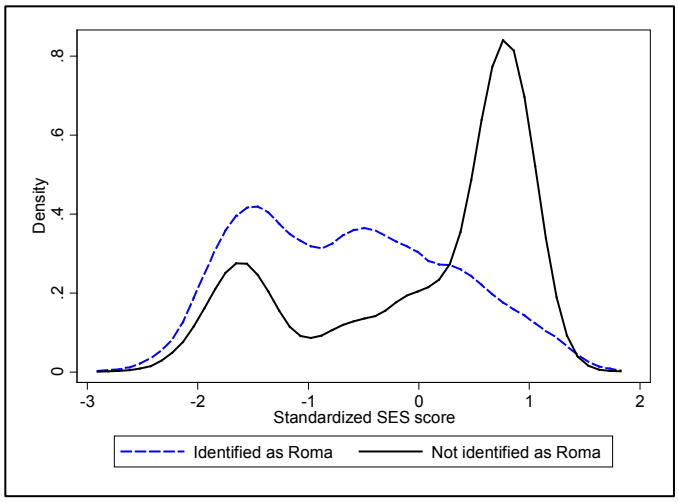

Figure 1 shows the estimated density of socio-economic status by ethnicity based on the one-dimensional measure described above. While there is a considerable overlap in SES of the Roma and the non-Roma, vast majority of the Romani are concentrated at the lowest part of the distribution. That suggests that our measure could – at least in principle - serve its purpose well as an explanatory variable of Roma ethnicity.

Empirical approach

We conceptualize Roma identity as a latent personal trait that is one element of a multi-dimensional vector that characterizes a person's identity. Similarly to other elements in this multi-dimensional identity, Roma identity is characterized by a continuous measure, which we denote by $r_{it} \geq 0$. Index i refers to potential heterogeneity in Roma identity, and index t refers to potential variation in Roma identification through time. The intensity of Roma identification can vary from zero to positive numbers, with higher values representing stronger identification.

We hypothesize that Roma identity is malleable, and individuals have a room for choosing their identity. In particular, we assume that ethnic identity is composed of inherited elements (i.e. the ethnicity of parents), and elements related to social status and the environment. On one, it is possible, that these elements are not congruent with each other, for instance someone can raised by parents who are

middle class, but identify as Roma, or by parents who do not identify as Roma, but in a neighborhood with many Roma. On the other hand, it is also possible that some of the constitutive elements of ethnicity as a latent trait change in time, thereby causing changes in reported ethnicity. The main goal of our analysis is to shed light on how these elements map to ethnic identification empirically.

Roma identity (r_{it}) is a latent variable that is not measured directly by surveys. Instead, surveys can provide imperfect and potentially noisy measures of r_{it}. The data used in our analysis asks two questions in each survey wave about whether respondents belong to particular national or ethnic groups, including the Roma group. In this setup let p_{its} denote the probability that individual i choses the Roma answer to any of the two questions in survey wave administered at time t. The additional index s denotes the circumstances of the survey situation. p_{its} is related to latent Roma identity of individual i at time t but it may be altered by random noise and circumstances specific to survey s. In particular, the probability of choosing the Roma answer (p_{its}) is a mapping from the intensity of Roma identity (r_{it}) and survey circumstances (d_{its}):

$$p_{its} = p(r_{it} , d_{its})$$

We assume that the probability of choosing the Roma answer is zero if the respondent has no Roma identity (i.e., if r_{it}=0), it may be zero or positive if the respondent has oositive Roma identity (i.e., if r_{it}>0), and the probability is higher if the respondent's Roma identity is stronger. In formulae,

$$p_{its} = 0 \text{ if } r_{it} = 0$$
$$p_{its} \geq 0 \text{ if } r_{it} > 0$$
$$\partial\, p_{its} / \partial\, r_{it} > 0 \text{ if } r_{it} > 0$$

In the empirical analysis we shall estimate linear probability models of the following structure:

$$p_{its} = \beta' x_{it} + u_{it}$$

The interpretation of β will depend on the content of the right-hand-side variable x: for variables that are related to the survey situation (e.g., dummies for survey waves) β will capture the effect of the survey situation on the probability of choosing the Roma answer, and for the other variables β will capture the relationships with the Roma identity.

With repeated measures of p_{its} across multiple survey waves, we shall estimate three different kinds of regressions: linear probability models for whether the respondent chose the Roma answer at least once, linear regressions for the fraction of the survey waves in which the respondent chose the Roma answer, and linear panel regressions for the repeated observations. The first type of regressions approximate the probability of having Roma identity (r_i>0); the second type of regressions approximate the intensity of Roma identity (r_i), while the panel regressions can address time series variation in the Roma identity (r_{it}).

Characteristics of the Roma identity in Hungary

In this section we present descriptive evidence on some important characteristics of the Roma identity. We first address the relationship of Roma identity and Hungarian identity. Second we show evidence on the continuous nature of Roma identity by analyzing our measure of the intensity of identity. Third we show some evidence on the relationship of self-identification and interviewer identification. Throughout this section we restrict the sample to adolescents who answered the identity questions in all of the four survey waves in which the question was administered. We do not put further restrictions on the sample here.

Bi-ethnic identity

First, we show that in Hungary, virtually everyone with a Roma identity also has a Hungarian identity. This finding is in line with approaches that put no restriction on the relationship between the minority identity and majority identities (eg. Casey and Dustmann, 2010).

Table 5 shows the distribution the of adolescent respondents by whether they chose the Roma answer and whether they chose the Hungarian answer to the identity questions at least once in the survey (recall that the respondents received the pair of identity questions in four waves). All but one of the respondents in the sample who chose the Roma identity in at least one of the four survey waves chose the Hungarian identity in at least one of the survey waves, too[9].

The fact that respondents with Roma identity also have a Hungarian identity has an important implication for the measurement of Roma identity. When inquiring about self-identification, many surveys, including most of the national censuses in the region, ask a single question on national or ethnic identity.[10] By doing so these surveys force respondents with a Roma identity to choose between a Hungarian identity and a Roma identity even though they have both identity. That choice most likely depends on the strength of the respondents' Roma identity but will have many idiosyncratic elements. It will also lead to a vast underrepresentation of the number of respondents with Roma identity as many of them would choose to answer some other identity that also characterizes them. By asking two questions our data allows for stating dual identities and helps respondents self-identify themselves as Roma together with another identity.[11]

Table 5. Hungarian and Roma identification by adolescent respondents.

[9] Whether the choice of the Hungarian answer category by all Roma respondents reflect their dual ethnic identity or their identification with the political nation of Hungarians is impossible to answer with the data at hand.

[10] Up to 2011 this was true for the Hungarian census, and this is true for the census in all other countries of Central and Eastern Europe with sizeable Roma minority. For example, the Hungarian census of 2001 asked "what nationality the respondent belongs to" and offered answer categories that included Hungarian and Roma. The Romanian census of 2011 asked "what ethnic group does the person consider he/she belongs to", allowed for any answer and then used a coding system to put respondents into one category that included Romanian, German, Hungarian, Roma etc.

[11] Similarly to our survey, the Hungarian census of 2011 asked the following two questions, with slightly different wording: "What nationality do you belong to?" and "Besides the one indicated before, do you belong to another nationality?" The number of respondents choosing the Roma answer category to one of the two questions was 308,957. This can be compared to 189,984, the number of respondents who choose the Roma answer category instead of other categories in the Hungarian census of 2001.

Ever chose Hungarian identity	Ever chose Roma identity		
	No	Yes	All
No	0.00	0.03	0.03
Yes	91.64	8.33	99.97
All	91.64	8.36	100.00

Sample: Hungarian Life Course Survey. Subsample of respondents for whom we have non-missing identity measures in all waves. Observations weighted by sampling weights. N = 6,394

The intensity of Roma identification

In the theoretical framework we conceptualized Roma identity as a latent variable that has an intensity element. We think of people's Roma identity as a continuous latent variable that can take on value 0 (not characterized by Roma identity at all) to any positive number (the larger number the more intensive Roma identity). When faced the question-pair on identity in our survey, respondents choose the Roma identity as an answer to one of the two questions with some probability that is a function of their latent Roma identity and the characteristics of the survey situation. In particular, we hypothesized that respondents without a Roma identity never choose the Roma answer, and the probability of choosing the Roma answer increases with the intensity of latent Roma identity.

Our data has repeated measures of identity that allows for a direct investigation into the intensity of Roma identity. Using data from the four survey waves with ethnic identity question asked from the adolescent respondents, we have created a measure of the intensity of Roma identification by counting the number of waves in which the respondents identified themselves as Roma. The distribution of the measure is shown in Table 6 together with the socio-economic status, reading test score and the fraction of Roma in the friendship network of the respondents.

Table 6. Intensity of Roma identification: Percentage of times the respondents identified themselves as Roma.

Number of waves the	Distribution	Mean	Mean	Per cent

respondents identified themselves as Roma		standardized SES	standardized reading test score	Roma in network
0	91.6	0.2	0.0	1
1	1.9	– 1.1	– 0.7	18
2	1.8	– 1.8	– 0.7	30
3	2.4	– 2.3	– 1.1	44
4	2.3	– 2.3	– 1.2	51
Total	100.0	0.0	0.0	4

Sample: Hungarian Life Course Survey. Subsample of respondents with non-missing identity measures in all waves. Observations weighted by sampling weights. N = 6,394

The patterns documented in Table 6 are consistent with our approach to conceptualize Roma identity as a continuous variable. There is substantial variation in the frequency of the adolescent respondents identifying as Roma, and that variation is very strongly related to socio-economic status, reading test scores and ethnic composition of the respondents' networks. The stronger the Roma identification the lower the socio-economic status, the lower the test scores and the higher the fraction of Roma in the network are.

If ethnic identification was binary (everyone being either Roma, or not) and those with a Roma identity would exhibit the same intensity of identity, we would expect the observed fraction of waves respondents identified themselves as Roma to be independent of important individual characteristics. That is clearly rejected by the evidence. On the contrary, we see patterns that our consistent with the hypothesis of social status being an important element of the existence and intensity of Roma identity.

Intergenerational transmission of Roma identity

The role of the family cannot be understated in the development of ethnic identity. A variety of different approaches highlight the importance of childhood socialization in the formation of ethnicity (see Laitin, 1995; Marks et al., 2007). It seems plausible to assume that the ethnicity of parents provides a "natural default" to that of their children. If this is the case, two interesting issues emerge. On one hand, if the transmission of ethnicity in the family is the rule, how can we explain the exceptions,

that is incongruence in terms of ethnicity across generations? On the other hand, how does ethnicity play out in families with mixed-ethnic parents? Does the "one-drop rule" apply for Roma ethnicity (i.e. are children with *any* Romani parent considered members of the minority)? If not, what determines the "dominant ethnicity" between parents (see Davenport, 2012 for a case study on biracial children in the USA)?

In this section we take a closer look at the relationship between the Roma identity of adolescents and their parents. We analyze three questions: first, whether the socio-economic status of the family and the ethnic composition of the environment are related to intergenerational links of Roma identity, and second, whether, in the case of mixed-ethnic parents, children of the same gender than the Roma parent are more likely to have Roma identity than children of the opposite gender.

Table 7 shows the ethnic identification of adolescents and their parents.. Both parents are identified as Roma in 5.3 per cent of the families with two biological parents, and in an additional 2.6 per cent only one parent is identified as Roma. Not surprisingly, the adolescents' ethic identification is very strongly related to their parents' ethnic identification. Less than 2 per cent of the adolescents identified themselves as Roma if neither of their parents identified themselves as Roma, compared to 96 per cent if both of the parents identified themselves as Roma.

Table 7. Distribution of the adolescents' and their parents' ethnic identification

Parents' self-identification	Fraction of adolescents that identified themselves as Roma in one of the four survey waves	Distribution of parents' self-identification
Neither identified as Roma	1.8	92.1
One identified as Roma	64.4	2.6
Both identified as Roma	96.1	5.3
Total	8.4	100.0

Sample: Hungarian Life Course Survey. Subsample of respondents who lived with both of their biological parents in waves A and B and for whom we have non-missing identity measures in all waves.
Observations weighted by sampling weights. N = 4,461

However, the correlation is not perfect: there are quite a few adolescents who identified themselves as Roma even though neither of their two biological parents identified themselves as Roma, and, conversely, there are some adolescents who did not identify themselves as Roma even though both of their two biological parents identified themselves as Roma. The lack of perfect correlation may reflect genuine changes of Roma identity across generations, but it may also reflect measurement error[12]. In any case, the results suggest that the transmission of Roma identity in one-Roma-parent families is significantly weaker than in two-Roma-parent families which is evidence against the application of the "one drop rule".

Separating child-parent differences in observed ethnic identification due to measurement error from genuine changes in ethnic identification is hard, and we do not have a bulletproof empirical strategy for that. The main concern is that even if we find a correlation between social status and adolescents' ethnic identity *conditional on their parents' identity,* there is an alternative explanation. According to the measurement error interpretation, adolescents from low-SES families where parents did not identify as Roma may have Roma identity because their parents had Roma identity, too, only the survey failed to detect that. Paradoxically, even if this is the case, we can still treat such results as evidence for an effect of social status on identity: we can interpret such results such that *conditional on the ethnicity of their children* more affluent parents are less likely to identify as Roma.

Table 8 reports the results of four linear regressions. Two versions are shown for two left-hand-side variables: whether the adolescent chose the Roma answer in at least one of the four survey waves and the fraction of the survey waves the adolescent chose the Roma answer. Thus, the second dependent measure can be thought of as a measure of the intensity of Roma identification. For each left-hand-side variable, the first regression (in columns 1 and 3) contains two dummy variables, one for both parents identified as Roma and one for only one parent

[12] People with Roma identity may not choose the Roma answer category to the identity questions in some survey waves. As a result, even if both the parents and their children have Roma identity (to some degree) some of them may not be identified as Roma in our data

identified as Roma. The left out category is neither parent identified as Roma. The second regression for each left-hand-variable (in columns 2 and 4) adds the standardized SES variable and its interactions with the two dummies.

The coefficient estimates of column (1) reproduce figures in the previous table. The coefficients in column (3) show the corresponding figures for the intensity of Roma identification. The results are qualitatively the same as in column (1): there is a very strong but imperfect link between the ethnic identity of the parents and their adolescent children, and children of only one Roma parent have substantially weaker Roma identity than children with both parents Roma.

According to the results in columns (2) and (4), Socio-economic status has a small but statistically significant role in measured Roma identity. Adolescents from families with one standard deviation *lower* SES are 6 per cent more likely to identify themselves as Roma in one of the survey waves if neither of their parents identified themselves as Roma (column 2), and they identify themselves as Roma in 3 per cent lower fraction of the waves (column 4). When one of the two parents is Roma, the effect of SES is stronger, but when both parents are Roma, ethnic identification depends on SES in a way that is similar to adolescents with parents that did not identify themselves as Roma.

Table 8. Roma identification of adolescents and their parents and socio-economic status (SES).

	Adolescent identified as Roma in one of four waves		Fraction of the four waves in which adolescent identified as Roma	
	(1)	(2)	(3)	(4)
One parent identified as Roma	0.62	0.49	0.31	0.21
	(0.06)**	(0.08)**	(0.03)**	(0.04)**
Both parents identified as Roma	0.94	0.91	0.73	0.66
	(0.01)**	(0.03)**	(0.02)**	(0.05)**
SES		-0.06		-0.03
		(0.01)**		(0.01)**
One parent Roma * SES (interaction)		-0.09		-0.09
		(0.06)		(0.04)*
Both parents Roma * SES		0.03		-0.01

(interaction)				
		(0.02)		(0.03)
Constant	0.02	0.05	0.01	0.03
	(0.00)**	(0.00)**	(0.00)**	(0.00)**
R^2	0.69	0.71	0.72	0.74
N	4,461	4,461	4,461	4,461

Sample: Hungarian Life Course Survey; subsample of respondents who lived with both of their biological parents in waves A and B and for whom we have non-missing identity measures in all waves. Observations weighted by sampling weights. N = 4,461. Heteroskedasticity robust standard error estimates in parentheses. * significant at the 5% level. ** significant at the 1% level.

We next ran regression similar to the ones above, but now we replaced social status with the ethnic composition of the adolescents' social network. Our prediction was that conditional on parents' ethnicity, adolescents having more Roma in their network would be more likely to identify as Romani themselves. This is exactly what we find: there is a very strong correlation between the proportion of Romani in one's network and their likelihood of identifying as Roma. (of course the same caveats about measurement issues apply here, too).

Table 9. Roma identification of adolescents and their parents and network composition

	Adolescent identified as Roma in one of four waves		Fraction of the four waves in which adolescent identified as Roma	
	-1	-2	-3	-4
One parent identified as Roma	0.62	0.5	0.31	0.23
	(0.06)**	(0.07)**	(0.03)**	(0.04)**
Both parents identified as Roma	0.94	0.9	0.73	0.58
	(0.01)**	(0.03)**	(0.02)**	(0.04)**
Percent Roma in network		1.15		0.66
		(0.10)**		(0.07)**
One parent Roma * Network (interaction)		-0.4		-0.19
		(0.18)*		-0.14
Both parents Roma * Network (interaction)		-1.02		-0.31
		(0.11)**		(0.09)**
Constant	0.02	0.01	0.01	0
	(0.00)**	(0.00)**	(0.00)**	(0.00)**
R^2	0.69	0.76	0.72	0.79
N	4,461	4,461	4,461	4,461

Apart of the sheer magnitude of the effect (in "non Roma" families there is a roughly one-to-one mapping from network composition and the likelihood of reported Roma ethnicity) there are some interesting interaction results (see column 2 and 4). We see that the effect of network composition is the highest for families with no parents identified as Roma and lowest where both parents did. When interpreting these patterns, to mechanisms should be considered. On one hand, the "room" for any variable to have an effect is the highest for respondents with no Roma parents, since their baseline probability of being Roma is very small and it is likely that some of these respondents' parents are *actually Roma*. On the other hand, we suspect that if no measurement error would be present, we would see the largest effects for mixed-ethnic parents, since they are the ones for which ethnicity is the least fixed. The patterns that we see in the interaction models are consistent with the interplay of these two mechanisms.

Finally, we focus on the gender aspect of the transmission of ethnic identity, with a particular focus on mixed-ethnic couples to their children. The first question we ask is whether the way Roma identity is transmitted to daughters and sons is different. The second question we ask is whether daughters or sons are more likely to identify themselves as Roma in families where only the mother is Roma or only the Father is Roma. Table 9 shows the results on the fraction of adolescents who identified themselves as Roma in at least one of the four survey waves. The numbers indicate that in mixed-ethnic families the ethnicity of the same-sex parent has stronger effect on the child's ethnicity than that of the opposite sex parent.

Table 9. Gender and the transmission of Roma ethnicity in the family

	Per cent of adolescents identified as Roma		
Of the two biological parents	Female	Male	Total
Neither identified as Roma	2	2	2
Only mother identified as Roma	**69**	**53**	60
Only father identified as Roma	**60**	**79**	70
Both identified as Roma	96	96	96
Total	9	8	8

Sample: Hungarian Life Course Survey. Subsample of respondents who lived with both of their biological parents in waves A and B and for whom we have non-missing identity measures in all waves. Observations weighted by sampling weights. N = 4,461

Table 10 shows the corresponding results in a linear regression (column 1), controlling for SES and the ethnic composition of the adolescents' network (column 2), and analogous linear regressions with the intensity of Roma identity as in the previous regression (columns 3 and 4). Besides allowing for control variables, the regressions allow for testing whether the observed differences are statistically significant. For each specification we test whether the coefficient on the only mother Roma X female interaction variable (the gender difference in ethnic identification if the mother is Roma) is different from coefficient of the only father Roma X female interaction variable (the gender difference in ethnic identification if the father is Roma).

Table 10: Roma identification of adolescents and their parents by gender

	Adolescent identified as Roma in one of four waves		Fraction of the four waves in which adolescent identified as Roma	
Only mother identified as Roma	0.52	0.38	0.24	0.12
	(0.11)**	(0.10)**	(0.06)**	(0.05)**
Only father identified as Roma	0.77	0.59	0.37	0.24
	(0.09)**	(0.07)**	(0.07)**	(0.07)**
Both parents identified as Roma	0.95	0.57	0.77	0.51
	(0.01)**	(0.05)**	(0.03)**	(0.04)**
Only mother Roma * female (interaction)	0.15	0.17	0.12	0.17
	(0.15)	(0.14)	(0.08)	(0.08)*
Only father Roma * female (interaction)	-0.19	-0.13	-0.11	-0.05
	(0.15)	(0.14)	(0.10)	(0.09)
Both Roma * female (interaction)	-0.01	0.03	-0.08	-0.05
	(0.03)	(0.04)	(0.04)*	(0.04)
Adolescent is female	-0.00	-0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
SES		-0.04		-0.02
		(0.01)**		(0.00)**
Fraction Roma in Neighborhood		0.02		0.02
		(0.01)**		(0.00)**
Fraction of Roma in personal network		0.35		0.36
		(0.07)**		(0.05)**
Fraction of Roma of best friends in school		0.23		0.09
		(0.04)**		(0.03)**
Constant	0.02	0.03	0.01	0.01

	(0.00)**	(0.01)**	(0.00)**	(0.00)**
R^2	0.69	0.76	0.73	0.80
N	4,461	4,075	4,461	4,075
F-test: Only mother Roma * female interaction)= Only father Roma * female (interaction)				
test statistic	2.61	2.52	3.29	3.16
p-value	0.106	0.118	0.069	0.076

Sample: Hungarian Life Course Survey; subsample of respondents who lived with both of their biological parents in waves A and B and for whom we have non-missing identity measures in all waves. Observations weighted by sampling weights. N = 4,461. Heteroskedasticity robust standard error estimates in parentheses. * significant at the 5% level. ** significant at the 1% level.

The results show that there are no gender differences in the Roma identification of adolescents in families with zero or two parents identified as Roma. However, daughters are somewhat more likely to identify themselves as Roma if only their mother is Roma than if only their father is Roma, whereas sons are substantially more likely to identify themselves as Roma if only their father is Roma than if only their mother is Roma.

According to the regression results and the F-test carried out on their coefficients, these latter differences are only marginally statistically significant and only when Roma identification is measured by its intensity. Because relatively few of the couples are of mixed ethnicity, the sample size is small, which may be responsible for the low statistical significance. Nevertheless, the consistency of the results across specifications suggests that the difference may be real and not just due to random sampling variation.[13]

Some panel evidence

The previous section provided evidence on how Roma identity is transmitted across generations and how social status, gender and the ethnic composition of te adolescents' network conditions assimilation. Now we take a different approach and look at how ethnic identification *changes* during the adolescence. Development psychologists argue that ethnic identity changes with age from early adolescence

[13] The differential transmission of ethnic identity by gender is the characteristic of self-identification only: in the first wave, interviewers identified adolescents with one Roma parent only as Roma with the same likelihood regardless of their gender.

and that the impact of the family and early environment shape the ethnic through a process of development (see Akiba et al., 2004; Marks et al., 2007).

Following the research of Penner and Saperstein (2008) we look at how significant changes in the life of adolescents affect their ethnic identification. We hypothesize that changes that are characteristic (or on the contrary, atypical) of the Romani population can map to ethnic identification through the internalization of typologies and prejudices (Kligman, 2001). Moreover, life changes might have an indirect effect on ethnic identification of adolescents through changes in the ethnic composition of their peer group (cf. Telles, 2002)

In particular, we look at three events that are characteristic to the Roma population empirically, and are intertwined with the public perception of Roma identity (Kligman, 2001). First, we predict that the decrease of household income could translate to an increase in the likelihood of Roma identification. The association of poverty and ethnic identification has been found elsewhere (Penner and Saperstein, 2008) and the enormous ethnic gap in the prevalence of poverty is a well-established fact in Hungary (see Kertesi and Kezdi, 2011).

Second, we analyze the effect of teenage pregnancy on Roma identity. While the Roma population shows a significantly higher fertility rate as compared to the rest of the population, this ethnic gap is particularly striking among adolescents. In our sample less then 2% of non-Roma adolescent had kids before the age of 20, while this number is above 20% among the Romani.

Finally, we look at differences in ethnic identification as a function of participation in higher education. As it was argued by existing studies of the Roma identity, staying in formal education is one of the few potential pathways for social mobility for the Roma (see Kertesi and Kezdi, 2011). If escaping from social marginalization would lead to leaving behind Roma identity, we would expect that staying in school decrease the likelihood of Roma identification. Moreover, since staying in school after the age required by the law is very much atypical among the Roma, higher

education could leave to a decrease in Roma identification through a change in the ethnic composition of peers.

Table 10: The ethnic gap in the dependent variables

Ethnicity	Household income	Still in school	Has children
Non Roma	70446	61%	3%
Roma	37691	8%	36%

All variables are measured in the last wave, N = 6,394 Observations weighted by sampling weights. Only respondents living with both biological parents.

We test our hypotheses by estimating fixed effects linear regressions explaining Roma identification with the variables describe above. The merit of this strategy is that including respondent fixed effects removes all the unobserved heterogeneity that is time invariant. In particular, with this approach we can make sure that our results are not driven by measurement error in the ethnic identification of parents (see more on this above). Thus, we identify the potential effect of changes from temporal variation only. We estimate each or our regressions on the full sample then we split the sample based on the ethnic identification of the parents. Our rationale for doing that is that we expect variables have different effects for adolescents growing up in families with different ethnic composition.

Apart of the main variables of interest, we include the age of the respondent (measured with monthly precision). On one hand, this allows for estimating the effect of aging *per se* on ethnic identification: (see Casey and Dustmann, 2010). On the other hand including age as a control variable is necessary since two of our explanatory variables of interest (school participation and having a children) is, for obvious reason, trending. Finally, we also include an indicator variable for the third wave because of the difference in the context in which the ethnic identity item was asked (see above).

Table 11 shows parameter estimates that are in line with our hypothesis, though the effects are not estimated very precisely. In particular, we see that giving birth to children, dropping out of school and getting poorer both correspond to a lower

likelihood of identifying as roma (though only the second one achieves statistical significance). A reason why the effects are not estimated precisely might be that children raised in families where the parents identify as Romani do not leave their minority identity behind, since the socialization is simply too strong. At the same time, using fixed effects reduces the effective sample size, since we only use those respondents that *changed* their identities between waves. In any case, we try to reduce the uncertainty in the estimates by splitting the sample based on the ethic composition of the respondents' parents (with similar considerations as in the interaction models of the previous section).

Table 11: Life changes and Roma identity (linear panel model, with fixed effects)

Dependent variable: Identifies as Roma	(1)	(2)	(3)	(4)	(5)
Age	0.02				0.02
	(0.02)				(0.02)
Has children		0.04**			0.04*
		(0.02)			(0.02)
Still in school			-0.01**		-0.01*
			(0.01)		(0.01)
Household income (log)				-0.00	-0.00
				(0.01)	(0.01)
Constant	-0.44	0.06***	0.07***	0.10*	-0.32
	(0.37)	(0.00)	(0.00)	(0.06)	(0.43)
Observations	20,743	21,202	21,211	19,094	18,651
R-squared	0.003	0.004	0.004	0.003	0.005
Number of respondents	5,941	6,085	6,085	5,946	5,807

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1, Wave fixed effects included, but not reported Observations weighted by sampling weights. Only respondents living with both biological parents

Table 12 presents the results for the split samples. Apparently, in "non-Romani" families, giving birth to children, dropping out of school and getting poorer corresponds to an increase in the likelihood of Roma identification. The estimated effects are surprisingly large as compared to Penner and Saperstein (2008). For instance, giving birth to a baby increases the likelihood of identifying as Romani by 7%. Those who drop out of school, or experience a 1% decrease in household

income are about 1% more likely to say that they are Roma. At the same time, we find null-results in the case of adolescents where some of the parents have identified themselves as Romani.

A probable reason for that is that the small number of respondents with self identified Roma parents becomes even more of a problem here as compared to the cross sectional models, since we base all our inferences on respondents whose reported ethnicity *changes* in time. At the same time, a more substantive explanation is that the effects of the studied life changes are changes are asymmetric. Adolescents whose parent *did not* identify as Romani, might start to "feel Roma" when they experience changes characteristic of the minority, but those with Roma parents will not change their identity if they *not* get poorer, *not* drop out of school or *do not* give birth to children. These interpretation paints a dark picture of a one way street: while social status affects ethnic identity, social mobility does not lead to ethnic assimilation.

Table 12: Life changes and Roma identity (linear panel model, with fixed effects)

Dependent variable: Identifies as Roma	(1)	(2)	(3)
Age	0.02	-0.14	0.13
	(0.01)	(0.42)	(0.27)
Has children	0.07***	0.07	-0.03
	(0.02)	(0.15)	(0.06)
Still in school	-0.01***	0.01	-0.04
	(0.00)	(0.07)	(0.04)
Household income (log)	-0.01*	0.05	-0.01
	(0.00)	(0.08)	(0.05)
Constant	-0.23	2.08	-1.34
	(0.19)	(6.65)	(4.54)
Number of Roma parents	None	One	Two
Observations	16,763	522	1,366
R-squared	0.019	0.021	0.007
Number of respondents	5,217	158	432

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Wave fixed effects included, but not reported Observations weighted by sampling weight

Only respondents living with both biological parents

Conclusion

Despite the importance and intensity of the discourse about sources and consequences of ethnic identity, empirical evidence on the matter seems surprisingly scarce on the matter. In this paper we provided some empirical evidence that could inform researchers of Romani identity formation, and more generally empirical research involving ethnicity.

We believe that our research provides three important insights about Romani ethnicity that, to some degree, probably apply to other settings, too. First, we have shown strong evidence supporting a continuous (as opposed to categorical) definition of ethnic identity. Using repeated measures on self-reported ethnicity, we find meaningful variation in the frequency by which respondents identify as members of the ethnic minority. As a consequence, existing studies of ethnic identity using single cross sectional samples probably failed to identify individuals as Romani, even though some of them might have identified as Romani at a different time.

Second, we found strong evidence against the "one drop rule": a non negligible number of adolescents reports ethnic identities *different* to their parents and only about 60% of the children of mixed-ethnic parents identify as Romani. Interestingly, we found evidence suggesting that gender has some role in the transmission of ethnic identity in mixed-ethnic families, with the ethnic identity of the same-sex parent being more consequential.

Third, we found evidence consistent with the hypothesis that social status, and the ethnic composition of ones network play an important role in the intergenerational transmission of ethnic identity. In particular, we showed that holding parents' ethnicity fixed, adolescents in less affluent families, and with more Romani in their social network are more likely to identify as Romani. To be sure, this finding is hard

to interpret, since it is possible that in some families we fail to "detect" the ethnicity of some parents, but not that of the adolescent and this measurement error is correlated with social status or network composition.

However, even in the face of our concerns about measurement error, we allow ourselves some confidence in interpreting our results based on our finding in the longitudinal analysis. Looking at the determinants of temporal variation of ethnic identity "within individuals" we find patterns that are consistent with the claim that social status shapes ethnic identities, and not just the other way round. We find that adolescents who experience life changes associated to low social status (poverty, teenage pregnancy and dropping out of school) are more likely to identify as Roma, even when family background is controlled for via fixed effects. Interestingly, the effects described above only manifest themselves in families where the parents *did not* identified as Romani, suggesting an asymmetric relationship between social status and Romani ethnicity.

While our paper points at the potential pitfalls of using non-representative populations to study of Romani ethnicity, some parts of our analysis reach the point where we simply lack the sample size to give credible answers to questions about interesting subpopulations (such as mixed ethnic families). In this sense, we believe different sampling methods, if done carefully, can complement each other in answering important questions about the formation of ethnic identities.

References

Bárány, Zoltán (2002). *The East European Gypsies*. Cambridge: Cambridge University Press

Barth, Fredrik (1969) *Introduction*, in Fredrick Barth (ed.), *Ethnic groups and boundaries. The social organization of culture difference,* Boston: Little Brown & Co

Bisin, Alberto, Eleonora Patacchini, Thierry Verdier, and Yves Zenou (2011): Formation and persistence of oppositional identities. *European Economic Review* 55(8) 1046-1071.

Brubaker, Rogers 2002 'Ethnicity without groups', *Archives européennes de sociologie,* vol. XLIII, 163-89

Brubaker, Rogers, Mara Loveman, and Peter Stamatov (2004): Ethnicity as cognition. *Theory and Society* 33(1) 31-64.

Casey, Teresa., and Christian Dustmann (2010). Immigrants' identity, economic outcomes and the transmission of identity across generations. *The Economic Journal*, 120(542), 31-51.

Csepeli, György: Gypsies and gadje: The perception of Roma in Hungarian society, *Central European Political Science Reiview* 11(40) 62-78

Csepeli, György and Dávid Simon (2004) Construction of Roma Identity in Eastern and Central Europe: Perception and Self-identification' *Journal of Ethnic and Migration Studies* 30(1), 129-150

Davenport, Lauren (2012): *Identification Between Black and White: The Politics of Racial Labeling Among Biracial Americans,* mimeo

Fábián, Zoltán and Endre Sik (1996): „Előítéletesség és tekintélyelvűség." In: Andorka Rudolf, Kolosi Tamás, Vukovich György. (szerk.) Társadalmi riport 1996. 381–413. old.

Farkas, Lilla, Gábor Kézdi, Sándor Loss, and Zsolt Zádori (2004): A rendőrség etnikai profilalkotásának mai gyakorlata. *Belügyi Szemle* 2-3.

Farmoso, Bernard (1986) *Tsiganes et Sédentaires. La reproduction culturelle d'une société*, Paris: L'Harmattan

Hale, Henry (2004): Explaining Ethnicity, *Comparative Political Studies* 37 458-485

Kemény, István. and Béla Janky (2006). Roma population of Hungary 1971–2003, in Kemény, I.,(ed.), Roma of Hungary, New York, NJ: East European Monographs, CO –

Atlantic Research and Publications, pp. 70–225.

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 Kezdi (2011b): The roma/non-roma test score gap in Hungary. *American Economic Review: Papers & Proceedings* 101 (3), 519-525

Kligman, Gail (2001). On the social construction of "otherness": Identifying "the Roma" in post-socialist communities. *Review of Sociology, 7,* 61–78.

Koulish, Robert (2005): Hungarian Roma Attitudes on Minority Rights: The Symbolic Violence of Ethnic Identification' *Europe-Asia Studies*, 57(2) 311-326

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

Ladányi, János and Iván Szelényi (2006): *Patterns of Exclusion: Constructing Gypsy Ethnicity and the Making of an Underclass in Transitional Societies of Europe*, New York: Columbia University Press

Laitin, David D. (1995): Marginality A Microperspective. *Rationality and Society* 7(1) 31-57.

Laitin, David D. (1998): *Identity in Formation*. Ithaca: Cornell University Press

Loss, Sándor. (2001): A romákkal szembeni diszkrimináció Magyarországon *Szabolcs-Szatmári-Beregi Szemle*, 36 (2), 190–201.

Manning, Alan and Sanchari Roy (2010). Culture Clash or Culture Club? National Identity in Britain. *The Economic Journal*, 120(542), 72-100.

Margalit, Yotam and Alex Kuo (2012): Measuring Individual Identity: Experimental Evidence. *Comparative Politics*, 44(4) 459-479

Marks, A. K., Szalachab, L. A., Lamarre, M., Boydd, M. J. and García Coll, C. (2007): Emerging ethnic identity and interethnic group social preferences in middle childhood: Findings from the children of immigrants development in context (CIDC)

study. *International Journal of Behavioral Development*, 31(5), 501-513.

Milcher, Susanne (2006). Poverty and the determinants of welfare for roma and other vulnerable groups in southeastern Europe, *Comparative Economic Studies*, 48(1), 20–35.

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

Palosi Éva, Endre Sik, and Bori Simonovits. (2007): Discrimination in shopping centers." *Szociologiai Szemle* 17, no. 3-4.

Penner, Andrew. M. and Alyia Saperstein (2008). How social status shapes race. *Proceedings of the National Academy of Sciences*, 105(50), 19628-19630.

Prieto-Flores, Òscar (2009) Does the canonical theory of assimilation explain the Roma case? Some evidence from Central and Eastern Europe, *Ethnic and Racial Studies*, 32(8) 1387 – 1405

Ringold, Dena, Mitchell. A. Orenstein and Erika Wilkens (2005). *Roma in an Expanding Europe*, Washington, DC: The World Bank.

Schwartzman, Luisa F. (2007): Does Money Whiten? Intergenerational Changes in Racial Classification in Brazil. *American Sociological Review* 72(6):940.

Shils, Edward (1957): Personal, Primordial, Sacred and Civil Ties. *British Journal of Sociology* 8 130-45.

Telles, Edward E. (2002): Racial Ambiguity Among the Brazilian Population *Ethnic and Racial Studies* 25(3) 415–441.

UNDP (2002): *Avoiding the Dependency Trap*, Bratislava: United Nations Development Program.

Vermeersch, Peter (2003): Ethnic minority identity and movement politics: the case of the Roma in the Czech Republic and Slovakia *Ethnic and Racial Studies*. 26(5) 879-901

Wimmer, Andreas (2008): The Making and Unmaking of Ethnic Boundaries: A

Multilevel Process Theory, *American Journal of Sociology,* 113. 970–1022