CIS WORKING PAPER

Nr. 81 , 2013

published by the Center for Comparative and International Studies (ETH Zurich and University of Zurich)

The Distribution of Ethnic Identities in Post-War Bosnia and Herzegovina: A New Estimation Approach

Daniel Bochsler
NCCR Democracy, University of Zürich

Basil Schläpfer
Centre for Democracy Studies Aarau (ZDA), University of Zurich

Universität Zürich[UZH]

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The Distribution of Ethnic Identities in Post-War Bosnia and Herzegovina: A New Estimation Approach

Daniel Bochsler*
NCCR Democracy, Universität Zürich
Affolternstrasse 56
CH- 8050 Zurich
Switzerland

E-mail: daniel.bochsler@uzh.ch
web: www.bochsler.eu
Tel. +41 44 634 54 53
Fax +41 44 634 52 00

* *corresponding author*

Basil Schläpfer
Centre for Democracy Studies Aarau (ZDA)
Küttigerstrasse 21
5000 Aarau,
Switzerland

E-mail: basil.schlaepfer@uzh.ch

The Distribution of Ethnic Identities in Post-War Bosnia and Herzegovina: A New Estimation Approach

Abstract

To understand political and social dynamics in multi-ethnic societies, we need robust and valid information on patterns of ethnic identification in the population. This paper proposes an indirect approach for estimating the post-war distribution of ethnic identities in the municipalities of the Federation of Bosnia and Herzegovina. Relying on a multiplier approach, we infer the distribution of ethnic identities from the declared identity of the parents of newly born children, correcting for differences in birth-rates as well as war-induced heterogeneity across municipalities and groups. Our approach appears to be the most accurate estimation procedure currently available for the distribution of ethnic identities in municipalities in Bosnia and Herzegovina. It might be useful not only for scholars of political and social processes in Bosnia and Herzegovina, but also for similar estimations in other countries.

Keywords: Ethnic groups; estimation; population; Bosnia and Herzegovina; multiplier method.

Introduction[1]

Ethnicity plays a central role in the social life and political sphere in many societies. The degree of ethnic diversity has a very important impact on social trust (Putnam, 2007) and voting behaviour in elections (Keech, 1968; Key, 1949; Birnir, 2007), and is closely linked to the manifold consequences of clientelism (Franck & Rainer, 2012). Academic studies increasing rely on detailed local information (Buhaug & Rød, 2006; Raleigh & Hegre, 2009) or survey data (Mitchell, Evans & O'Leary 2009) that inform about the ethnic identification of the population. Prominent examples are studies on the relative success of multi-ethnic parties to attract votes from different ethnic groups in ethnically divided societies. Often, however, neither survey data nor local-level aggregate data on the distribution of ethnic identities are available to scholars. Moreover, ethnic diversity itself can become the object of political struggles. Particularly in conflict-affected countries, surveying information on ethnicity in censuses becomes a very sensitive issue, which is subject to political campaigns and manipulation. And yet, in order to improve the understanding of processes of social polarization associated with ethnicity, and in order to advance our knowledge of the institutional designs most conducive to political stability and equality, it is pivotal that scholars can rely on reliable data that inform us about the distribution of ethnic identities in the respective societies.

We provide a new approach for estimating the distribution of ethnic identities based on birth figures that we apply to the case of Bosnian ethnic groups. Due to the political sensitivity of ethnic questions in Bosnia and Herzegovina after the war during the 1990s, no census has been conducted after 1991. This paper provides reliable and transparent estimates of the distribution of ethnic identities for 77 out of 79 municipalities of the Federation of Bosnia and Herzegovina, for the years 2008-2010. Due to the high levels of inter-ethnic violence and ethnic cleansing during and in the aftermath the war, we currently lack reliable information about ethnic diversity in municipalities of Bosnia and Herzegovina. Our contribution is therefore twofold: First, we provide novel estimates of the distribution of ethnic identities in Bosnian municipalities that might serve as valuable information to other scholars. Second, our estimation approach can possibly be fruitfully extended

1 This paper was prepared for presentation at the RRPP Scientific Conference 2012, Sarajevo, Bosnia and Herzegovina, May 24-5, 2012. We would like to express our gratitude to Azer Kurtović from GDi GISDATA d.o.o. Sarajevo, who generously provided us with geographic data on pre- and post-war Bosnia and Herzegovina, and who answered our numerous questions. We also would like to thank Nils Weidmann, Mirsad Tokaca, Nina Caspersen, Carl Dahlman, Gerard Toal, and Ronald Schmidt for their helpful comments and answers to our questions. All errors remain our own.

to other areas of research where data availability might be scarce.[2] State and international agencies have previously estimated the distribution of ethnic identities in the population of Bosnia and Herzegovina. However, these estimates are not accessible to the public, estimation procedures are not available and some provide information only for a small set of municipalities – and, as this paper will show, for some municipalities these existing estimates are not always entirely plausible.

The estimation procedure suggested in this paper relies on a subset of the population. We construct the distribution of ethnic identities in municipalities of Bosnia and Herzegovina from the ethnic identity of birth-giving mothers and fathers. The extrapolation of population data from birth registers might suffer from a number of problems: First, birth rates characteristically vary across different social groups within a population. Thus, some of the groups whose numbers we are interested in estimating using our method may be particularly well represented, while in other groups, we might find considerably less parents. Moreover, birth rates tend to be unevenly distributed across municipalities as well, due to unobserved heterogeneity between villages and towns. In two steps of analysis, our model first provides estimates for group-specific birth rates by municipality, the results of which can be used to obtain an estimate of the full population structure.

Second, scholars widely agree that ethnicity is constructed: the ethnic affiliation can change over time, different persons will perceive the identity of the same group of people differently, and ethnicity is sometimes described as a set of aspects of identities, which can be activated in different ways (Chandra, 2006; Lee, 2009). Ascribed ethnic labels do not necessarily overlap with individuals' self-identification, and the importance they attribute to it – although in ethnically strongly divided societies, ascribed ethnic identities tend to overlap with the individuals' experiences in the public sphere, regardless of their level of self-identification. These problems are intrinsic to most procedures that provide estimates based on a few mutually exclusive ethnic categories. To test for these limits, we carry out a number of validation tests that compare our numbers not only to earlier estimates, but also to an indirectly connected variable. Despite the simplified measure, our estimation results correspond closely to specific patterns of voting behaviour, and when compared to two previous estimation procedures, our results outperform them in the validity tests.

While academics and policy-makers would like to be able to rely on ethnic data, in Bosnia and Herzegovina the population census has remained controversial; the question of whether ethnicity

[2] As a consequence, studies of post-war politics and social relations have relied on indirect approaches (Caspersen, 2004: 577), or been limited to the level of the two entities and the Brčko district (Bakke, Cao, O'Loughlin, & Ward, 2009).

should be asked about and how it should be asked is politically sensitive. We share some of these concerns, but argue that the interest in generating unofficial ethnic estimates is legitimate for several reasons. First of all, we believe that valid estimates of the distribution of ethnic identities in ethnically divided countries is important to further our understandings of political and social processes in these countries, such as the effects of particular institutional designs. Our estimation provides for retrospective information in a sensitive period when resettlements took place. As there is a lack of accurate data since the 1991 census, the post-war period risks remaining a black hole for social science studies.[3] In the case of Bosnia and Herzegovina, the provision of ethnicity-related estimates might be less controversial than in other countries, as previous estimates already exist. Yet another estimation might give a clearer and – above all – more transparent procedure, without being new per se. Finally, we stress that – while small and vulnerable groups might gain more political visibility from featuring in population statistics (Williams & Husk, 2013: 297) – the estimation remains unofficial. As such, it does not provide the same legitimacy for political action as 'official' census data, and can and should not be used for the establishment of ethnic quotas. The availability of reliable data on the distribution of ethnic identities can contribute to sound research that improves our understanding of the causes and effects of social and political polarization in ethnically divided societies, as well as potential avenues to its alleviation.

Moreover, our estimation approach can possibly be extended to other social science applications where official information on given subpopulations is scarce, and where population surveys and alternative estimation methods such as enumeration, multiplier, and capture-recapture techniques are not feasible.

We proceed as follows: After a brief review of relevant literature, the second section of this paper explains our group-based birth rate estimation method, and applies it to 77 out of 79 municipalities of the Federation of Bosnia and Herzegovina. In a two-step procedure, we first estimate the birth ratios for different segments of the society, and thereafter use them to estimate the population structure by ethnicity. Results are available on the authors' website.[4] The third section provides a number of tests to examine the validity of our procedure.

[3] The census, which was conducted 1996 by UNHCR was never released, and furthermore, there were considerable population movements in the late 1990s and the 2000s.
[4] www.bochsler.eu/data.html

1. Estimation procedure

The estimation of the size and structure of overall populations from subsamples is a well-established method in various disciplines, such as biology or public health. One of the most common methods to estimate populations is the capture-recapture technique (e.g. Bishop, Fienberg, & Holland, 1975: 229-256; Ball & Asher, 2002).

The capture-recapture technique (and extensions thereof) uses two or more independent lists of individually identifiable subjects (persons, animals, etc.) to estimate the size of the unobserved underlying population. If the independence of the lists cannot be guaranteed, multiple sample recapture methods allow for alternative estimates, based on three or more lists (Bishop et al., 1975: 254-256).

Alternative techniques include enumeration and the multiplier methods (UNAIDS/WHO 2003: 13-23): Enumeration methods are similar to census methods, but limited to pre-defined segments of the target population. Multiplier methods work with two independent records of data that relate to each other and extrapolate certain characteristics from a sub-sample of the population to the overall population (UNAIDS/WHO 2003: 19-23). This method thus relies on the assumption that the sub-sample of the population for which information is available is representative of the overall population.

For previous estimates of the distribution of ethnic identities in the municipalities of Bosnia and Herzegovina, official data on the ethnic structure of the population deviate from each other. The Federal Office of Statistics has estimated the ethnic structure of all 79 municipalities of the Federation of Bosnia and Herzegovina in 2005 for the three constitutive groups (Bosniaks, Serbs, Croats) and others, but it is possible that their estimates might partially still rely on the pre-war census (Federalni zavod za statistiku, 2005). Based on more recent data, the OSCE has estimated the distribution of ethnic identities in 65 municipalities, but their methodology has not been published. Toal and Dahlman (2011) provide estimates for five municipalities of the Federation, whereas the International Crisis Group has estimated the distribution of ethnic identities in Mostar from the entries in the electoral register (2003: 6-7,15). An approach of estimating the distribution of ethnic identities from electoral data (esp. electoral results) seems to be straightforward, as the political landscape of Bosnia and Herzegovina is highly enticized, with most people voting for parties representing their ethnic group. We refrain from using such a strategy, however, as it would not allow the clear identification of voters of multi-ethnic parties, and because data derived from

election results would be highly endogenous for the explanation of political behaviour or political cooperation.

2. A two-stage procedure for the estimation of the distribution of ethnic identities

Our estimation of the distribution of ethnic identities in Bosnia and Herzegovina is based on the multiplier method, but corrects for the selectivity of the sub-sample. While no representative survey or census is available to measure the distribution of ethnic identities at the municipal level, information on ethnic identities is collected upon the registration of newly born children. Every birth registration includes the ethnic identity of both parents of newborns. From this data, we can extrapolate to the overall municipal populations.

Parents of newly born children are, however, not a random or representative sample of the full population. Birth rates vary geographically and across social groups, so that the number of parents cannot directly inform us about the overall size of the group they belong to. Moreover, parents of newborns are not only a selective sub-sample of the overall population, but the selection process might be linked to those attributes of the population which we are particularly interested in, namely ethnicity and geography. For instance, if members of a specific group or municipality are less likely to have children than others, their members will be underrepresented in the birth figures, and thus the overall size of this group risks being underestimated. Therefore, our methodology relies on a multiplier method (extrapolating from a sub-group to the overall population), with a correction procedure introduced for the variance in birth rates across groups and municipalities. This approach takes account of the selective nature of the data. To our knowledge, this represents not only the first publically available, but also the first well-documented estimation of the distribution of ethnic identities in municipalities in the Federation of Bosnia and Herzegovina. The reliance on birth-related baseline data from medical institutions is not uncommon when it comes to the estimation of population figures. For instance, HIV surveillance is commonly based on women attending prenatal care clinics (WHO, 2003: 3-5).

The model

Information related to the distribution of ethnic identities in municipalities in Bosnia and Herzegovina is generally scarce. The only relevant information which might be indicative of the ethnic identification of the general population stems from birth registers. The registration of every newly born child includes information on the ethnic identity of both its parents. This information is

thereafter reported in the birth statistics of the Federation of Bosnia and Herzegovina, for each of the three constituent groups, and several other groups. We also have data on the overall size of the population of the municipalities. Our model will thus be based on the following information (and a set of further control variables, such as numbers of refugees/returnees and GDP estimates):

- p_{Ti} total population size of municipalities *i*
- b_i total number of newly born children in municipality *i*
- $b_{e,i}$ parents of newly born children in municipalities *i*, by ethnic identity *e*, where *e* distinguishes Bosniaks (B), Croats (C), Serbs (S), and others (O)

We will use this information to derive an estimation of the following variable:

$p_{e,i}$ population of municipalities by ethnic groups
($p_{S,i}$ = population with Serbian ethnic identification in municipality *i*, etc.)

Naive Birth Rate Model

A first, direct application of the multiplier method allows us to obtain an estimate of the ethnic population structure. We could assume that the parents of newly born children are a non-selective sample of the overall population, and thus treat them as a representative sub-group. Hence, we estimate a single birth rate for each municipality ($\gamma_i = \frac{2 \cdot b_i}{p_i}$). Note that our birth rate refers to the overall population, and not to a specific age segment of the population. We derive the population structure by ethnic group by applying the same single birth rate to each of the ethnic groups, $p_{e,i} = \frac{b_{e,i}}{\gamma_i}$.

This direct application would lead to valid results under the assumption that there are no significant differences between the birth rates and the age structures of different ethnic groups. In addition, such a direct estimation would rely on the assumption that birth rates only co-vary with local characteristics, and that they are not connected to ethnicity. We call this the *naive birth rate model*, which assumes that those conditions hold. If, however, the birth rate varies across ethnic groups, either due to demographic or value-related differences between the ethnic groups, the estimates will be biased.

Full Birth Rate Model

The assumptions underlying the naive model are empirically implausible: war and migration have significantly altered the structure of the population, and not only particularly affected the ethnic, but

also the age structure of the society. For this reason, a valid estimation should consider possible variation in birth rates as well as differences in the distribution of age groups across municipalities.

We allow the birth rate $\gamma_{e,i}$ to vary between municipalities *i* and between ethnic groups *e*. We define the birth rate as the quotient of the number of newly-born children $b_{e,i}$ from ethnic group *e,* in municipality *i* and the number of inhabitants declaring themselves as belonging to this ethnic group $p_{e,i}$. We do not have any information of γ_e or of $p_{e,i}$.

$$\gamma_{e,i} = \frac{b_{e,i}}{p_{e,i}} \tag{1}$$

We know, however, that the sum of the population of all ethnic groups $p_{e,i}$ residing in a municipality *i* is equal to the total size of the population, p_{Ti}.

$$p_{Ti} = p_{Bi} + p_{Ci} + p_{Si} + p_{Oi} \tag{2}$$

The two equations (1) and (2) can be re-arranged as follows:

$$p_{Ti} = \frac{b_{B,i}}{\gamma_{B,i}} + \frac{b_{C,i}}{\gamma_{C,i}} + \frac{b_{S,i}}{\gamma_{S,i}} + \frac{b_{O,i}}{\gamma_{O,i}}$$

We divide the overall municipal population p_{Ti} by the overall local birth rate b_i, and this results in the inverse overall birth rate, $\frac{1}{\gamma_i} = \frac{p_{Ti}}{b_i}$. We add an arrow $\beta_x X_i$, which stands for further covariates of the local birth rate.

$$\frac{1}{\gamma_i} = \frac{p_{Ti}}{b_i} = \left(\frac{b_{Bi}}{\gamma_B \cdot b_i} + \frac{b_{Ci}}{\gamma_C \cdot b_i} + \frac{b_{Si}}{\gamma_S \cdot b_i} + \frac{b_{Oi}}{\gamma_O \cdot b_i} \right) + \beta_x X_i \tag{3}$$

The resulting equation (3) is a linear function, where each of the five unknown variables - the birth rates γ_B, γ_C, γ_S, γ_O, and the parameter for the control variables, β_x – is part of one summand. We can estimate the function with a linear regression model, where $\frac{1}{\gamma_i}$ is our dependent variable.

Note that the equation is not fully free of assumptions. We allow for birth rates to vary by ethnic groups, i.e. we estimate birth rates for each ethnic group, γ_B, γ_C, γ_S, γ_O. Also, we allow birth rates to vary across municipalities, by adding further control variables X, which explain some of the unexplained variation. This means, however, that birth rates need to vary similarly for all ethnic groups across municipalities. For instance, if rural municipalities have higher birth rates than urban municipalities, this can be considered, by including a variable for rural municipalities in the vector

of control variables *X*. However, we expect the birth rate is higher for all four ethnic groups in rural municipalities.

In a nutshell, this first step of our estimation procedure derives estimates of group-specific birth rates by municipality.

Model specification and control variables

The difficulty of operationalising models with demographic variables in Bosnia and Herzegovina is the scarcity of data – due to the lack of up-to-date census data. This paragraph discusses the data on which we rely in our estimation and several issues which we consider to be crucial for estimating population structures in a post-war society.

The main information relies on the birth registration statistics of the Federation of Bosnia and Herzegovina, and on the population statistics. As explained earlier, the birth registration includes a question about the ethnic identity of both parents of newborns. It should be self-reported, and registered by the state official ('matičar'). In practice, however, state officials occasionally do register parents who do not claim to belong to any of the official identity groups by selecting a group based on the parents' names.[5] Data was obtained from the Statistical Office of the Federation (2003-2010).

Further variables were included to control for possible biases. First, due to migration and war casualties, the age structure might vary across municipalities, and might possibly co-vary with the ethnic composition of the municipalities. The estimation might thus relate to a specific age group, rather than to the overall population. In this case, the variable p_{Ti}, which measures the overall size of the population, would need to be restricted to those particular age groups. The municipality data available allow us to restrict the population to those aged 18 to 64. This comes at a cost, however, as we thus estimate the distribution of ethnic identities only for those aged 18-64. (To test for the robustness of the estimates, we have run the model for the full and the restricted age sample, with almost identical results.)

Second, the ethnic and the age structure are affected by refugees and returnees. In both regards, they are a highly selective social group, as people were usually expelled from their homes based on their ethnic identity, and their return depends on the economic and social context of the municipalities. This means that they are not representative of the local population. While we do not focus on refugees and returnees in particular, the age and ethnically selective nature of both groups

[5] Interview with a political analyst, Sarajevo on 21 May 2012 (kindly provided by Elena Stavrevska).
A template of the registration card can be found in the appendix of Federalni zavod za statistiku (2009).

will also have an impact on local birth rates. Furthermore, the groups of refugees and returnees might not be homogeneous across the territory of Bosnia and Herzegovina, so that controlling for the size of the two groups is not sufficient. Instead, we expect heteroscedasticity effects: high numbers of refugees and returnees might lead to biases, so that the results of our models will be less precise in municipalities heavily affected by wartime migration. To control for such a heteroscedasticity effect, we run linear models with separate variance-predictors, and include the share of refugees and returnees in the variance part of the estimation.

Data on refugees and returnees is available only for 68 out of 77 municipalities; we have imputed missing information for the remaining 9 municipalities. Estimation results remain stable with/without imputation, but the imputation procedure allows us to produce estimates for all municipalities.

We further control for the local GDP estimate, the overall population size of the municipality, and for population density (all logarithmised) to capture some of the socio-economic characteristics of the municipalities and their economic wealth. Given that we use the ethnic identity of parents as a sample for the overall ethnic diversity of the population, the estimates should be more accurate in large municipalities (large samples) than in small municipalities. We therefore include the municipal size in the variance part of the estimation.

Variance models simultaneously estimate not only the mean of the distribution of the dependent variable, but also its variance σ^2. They thus allow us to model the effects of heteroscedasticity if we know to which variables the variance is related (Davidian & Carroll, 1987; Braumoeller, 2006). In this particular case, the main model (equation 3) and one category of the control variables relate to the mean of the dependent variable, whereas the share of refugees and returnees among the population, the size of the municipality, and the share of children (<15 years) are likely to have an impact on the variance.

In two municipalities, there were no births registered at all (Ravno), or only very few births registered (Neum). While Ravno is a rather small municipality, Neum is surrounded by Croat territory, and is according to all sources overwhelmingly inhabited by ethnic Croats. Mostly, parents with residence in Neum give birth in hospitals in neighbouring Croatia, where they can receive health care as domestic citizens.[6] While the exclusion of the two cases does not substantially alter the results of the estimation as a whole (results available from the authors), the resulting estimates for the two municipalities are unreliable.

[6] http://wikileaks.org/cable/2003/04/03ZAGREB866.html (last accessed on 11 July 2013).

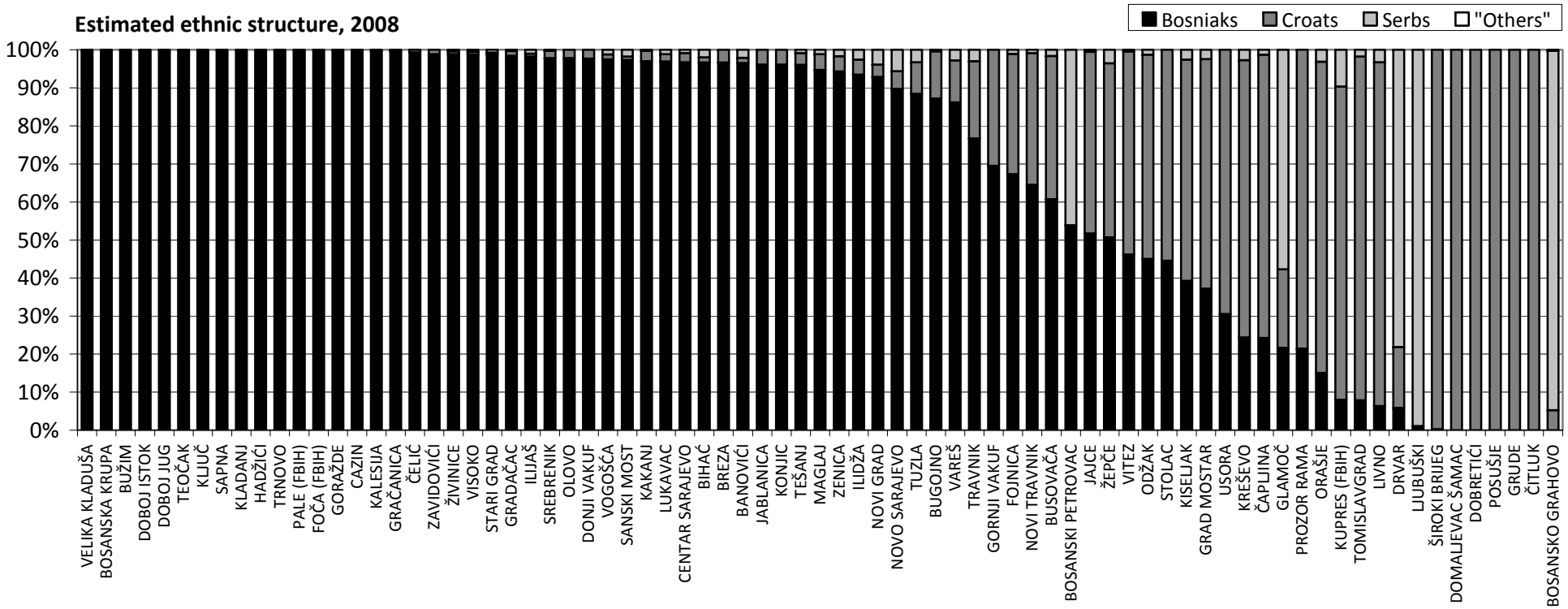
Results of the estimation of the birth rates are presented in table 1 (dependent variable: inverted birth rates). In the model, ethnic Bosniaks are our remote category, meaning that the coefficients measure the differences between the birth rates of the Bosniaks and of other ethnic groups. Birth rates tend to be higher in municipalities with a low population density, and with a lower share of non-returned refugees. After controlling for these factors, we find that the overall birth rate lowers as the percentage of Serbian parents increases. Among Croats, birth rates tend to be higher than among Serbs, but lower than among Bosniaks. This is in line with our expectation that there are differences in the birth rates between ethnic communities. As expected, the size of the municipality, age structure and share of refugees and returnees affect the accuracy of the estimation.

Table 1: Explanatory model for birth rates γ_i in the Federation BiH, 2008-2010
Variance model (Linear regression with multiplicative heteroscedasticity)
[a] Variable imputed for 9 municipalities.

	year=2008			year=2009			year=2010		
	Coef.		Robust Std. Err.	Coef.		Robust Std. Err.	Coef.		Robust Std. Err.
main									
b_{Ci} / b_i	10.02	**	3.22	10.27	*	4.18	11.93	**	4.42
b_{Si} / b_i	49.98	**	7.63	76.69	**	22.34	-23.14	*	10.54
b_{Oi} / b_i	-49.16		30.97	-126.26	**	32.43	-99.83	**	23.82
% employed women	-10.99		12.24	5.53		11.53	-15.35		18.46
population density (log)	.54		.85	1.69		1.08	1.88	*	.85
GDP per capita (log)	.33		1.99	-1.69		1.99	-2.83		2.50
population (log)	-.23		.1.49	-.31		1.53	1.31		1.94
returnees (log)[a]	.09		4.54	-4.58		4.93	-1.64		5.27
refugees non-returned (log)[a]	21.77	**	2.61	47.03	**	5.41	45.31	**	3.12
constant	30.94		19.17	39.86	*	18.59	39.23	*	19.57
vare									
returnees (log)[a]	1.96	*	.93	1.85		1.14	2.72	**	.91
refugees non-returned (log)[a]	-8.03	**	1.26	-4.26	**	1.23	-8.08	**	1.22
population (log)	-.89	**	.27	-.78	**	.30	-1.02	**	.26
% children	-34.54	**	6.02	-32.96	**	6.59	-27.2	**	6.23
constant	31.92	**	8.84	18.34	**	3.02	20.10	**	2.78
Number of obs	77			77			77		
Model $\chi^2(12)$	83.9			121.2			72.035		
Prob > χ^2	0			0			0		
Pseudo R^2	0.1309			0.1736			0.1093		
VWLS R^2	0.975			0.5641			0.9897		

In the second step, based on the models presented in table 1, we calculate predicted values for the birth rates in 77 municipalities, for each ethnic group (γ_B, γ_C, γ_S, γ_O, for Bosniaks, Croats, Serbs, 'others'), whilst considering of the socio-economic context. Equation 1 allows us to predict the overall distribution of ethnic identities in these municipalities, which we display for 2008 in figure 1. Certainly, the distinction of municipalities by the largest group after the war (Bosniaks majorities are located on the left-hand side, Croat and Serb majorities right-hand side) reproduces well-established knowledge. More importantly however, our results also provide an estimation of the actual size of each group. To our knowledge, this is the first result of this kind that relies on an objective and public procedure. The next section will assess the extent to which our estimates are also robust and valid.

Figure 1: Population estimates by ethnic group, 2008, municipalities of the Federation of Bosnia and Herzegovina. Full model estimation.



Reliability and validity

In what follows we provide the results of a number of tests that assess the reliability and the validity of our estimates reported above.

Reliability test: Although the age structure - and, especially in small municipalities -- the number of births can change over time, the distribution of ethnic identities should correlate strongly from one year to the next in the post-war period. Hence, if reliable, our estimation results should almost be stable over the years. As a test of the reliability of our results, we have therefore correlated our estimates for the available years (2008-2010).

Validity tests: Testing for validity is more challenging. We need to assess the extent to which our population estimates correspond to the actual distribution of ethnic identities in municipalities in Bosnia and Herzegovina – which due to the absence of official data can only be done indirectly. We perform two well-established tests of validity. First, we test whether the new measures reproduce the results of other estimates (criterion validity). Second, we employ the new estimates to test for theoretically expected relationships (construct validity).

Criterion validity refers to whether a given measure correlates with other indicators for the same construct – in our case, the ethnic composition of municipalities. We could only find unofficial and not publically available previous estimates which were made by the Federal Statistical Office (for all municipalities, in 2005) and the OSCE (covering only 31 of the municipalities of the Federation, in 2008). We correlated them with our estimates for the year closest in time to these, namely 2008. Ideally, the correlation coefficient (according to Pearson) should amount to 1. Our estimates of the full model are all above 0.96 for the ethnic Bosniaks and the ethnic Croats, and around 0.83-0.91 for the estimates of ethnic Serbs. Our results estimate the share of Serbs to be higher than the Federal Statistical Office; the difference is particularly pronounced for the municipality of Drvar (32% versus 78%). Given these differences in the estimates of ethnic Serbs, it is crucial to test which of the measures is more precise in explaining theoretically expected patterns. Compared to the previous estimations, our full model fits the alternative estimates only marginally better than the naive birth-rate model. For 'other' minorities, no previous measurements are available.

Table 2: Results of reliability and criterion validity tests

		Bosniaks		Croats		Serbs	
Type of test	*Indicator*	*naive birth-rate model*	*full model*	*naive birth-rate model*	*full model*	*naive birth-rate model*	*full model*
Reliability	mean correlation across years	0.975	0.947	0.962	0.966	0.984	0.978
Criterion validity	correlation with OSCE estimates (N=31)	0.956	0.962	0.964	0.976	0.831	0.829
	correlation with FZS estimates (N=77)	0.962	0.970	0.956	0.961	0.881	0.907

Construct validity refers to whether a given measure is able to predict a theoretically expected relationship. In many ethnically divided countries, we can find a strong correlation between ethnic identities and votes for ethno-nationalist parties. In Bosnia and Herzegovina, elections are regularly perceived as de-facto ethnic censuses. We use the vote share of ethno-nationalist parties in the 2006 parliamentary elections and the 2008 local elections to assess our estimates. Ethno-nationalist parties obtained 73%-75% of the votes in the Federation of Bosnia and Herzegovina in these elections.

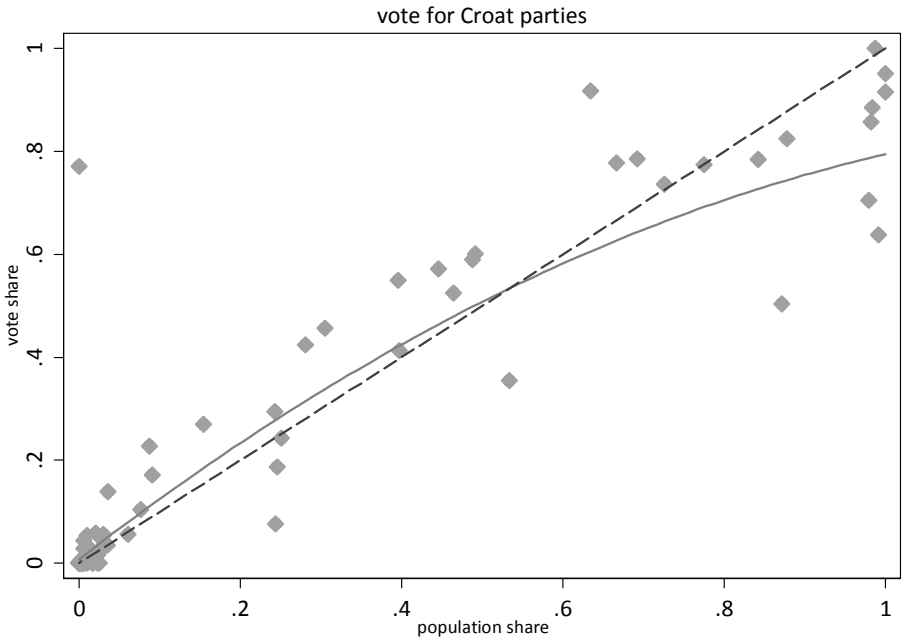
Two aspects are tested for. First, we expect that the municipality-level voting results can inform us to a limited extent about the distribution of ethnic identities Due to the contextual effects of ethnic diversity on the propensity of voters to vote for nationalists, the relationship is often curvilinear.[7] Under strong patterns of ethnic voting, the model should have high explanatory power for the vote shares of the ethno-nationalist political parties. The classification of parties according to their ethnic orientation is included in appendix A. Figure 2 shows such a model exemplarily for the Croat parties, based on our full estimation model. The curvilinear line is the best fit, and the graph demonstrates that our population estimates can explain the vote share of Croat parties reasonably well, although some variation remains.

This unexplained variation might reveal that our population estimates are not very precise. However, the variation is at least partly due to the fact that not all ethnic Croats vote for ethno-nationalist parties, but instead cast their votes for non-ethnic or multi-ethnic parties. This might weaken the explanatory power of the model. The second test we apply for the data is less

[7] For models explaining the curvilinearity of this relationship, see (Bochsler, 2013; Grofman & Handley, 1995).

demanding and considers the possibility that not all votes will be cast for ethno-nationalists. Considering this, and assuming equal turnout across all groups, the vote share of ethno-nationalist parties should be somewhere between 0% and the percentage of the ethnic group they claim to represent among the local population. Hence, if the population estimates are accurate, the vote share of the Croat parties should be located below the diagonal line in figure 2. Any points above the diagonal line indicate that we have under-estimated the share of ethnic Croats in these particular municipalities. Our second indicator of construct validity counts all votes which were cast for ethno-nationalist parties above the diagonal line, i.e. deviations, which cannot be explained with the population structure.

Figure 2: Estimated share of Croat population and vote share for Croat nationalist parties, 2008 municipal elections. (Full estimation model)



We thus use two indicators to measure the construct validity of our estimates. The first indicator, the adjusted R^2, measures to what degree our estimates explain the vote share of ethno-nationalist parties. High values, close to 1, indicate a high degree of validity. Second, we test whether the population figures underestimate the voting potential for ethno-nationalist parties. This is the caseif ethno-nationalist parties of an ethnic group win higher percentages of the vote than constituted by their ethnic group in a particular municipality. As we are less concerned with minor deviations (which might, for instance, be due to differences in turnout across ethnic groups) than with larger

ones, we use a measure of squared deviations.[8] Small deviations (on a scale from 0 to 10000) indicate a high degree of validity.

Based on these two indicators, our results clearly outperform previous estimates of the ethnic map of Bosnia and Herzegovina. This holds for two previous estimates which we use as our benchmark: the estimates of the Statistical Office of the Federation of 2005, and – restricting the sample to 30 available municipalities for which numbers are available – the estimates provided by the OSCE. Considering that the Statistical Office estimates date back to 2005, we have performed our validity tests not only on the results of the 2008 municipal elections, but also for the 2006 parliamentary elections.

The corrected birth-rate procedure presented in this paper appears to provide a much more accurate estimation of the size of the Serb and the Bosniak communities, compared to the estimates of the Statistical Office: validation models, which are based on the figures of the Statistical Office, result in only moderate explanatory power for the success of Serb parties, with an adjusted R^2 around 0.49 (figures for the 2008 elections). Our figures allow a fairly accurate explanation (adj R^2: 0.96). The Statistical Office appears to under-estimate the size of the Serbian community (dev_e: 23.6), and of the ethnic Bosniak community (dev_e: 20.6) in certain municipalities, which our full model does not. In contrast, we under-estimate the size of the Croat community in several municipalities,[9] but with a deviation of dev_e=13.9, this is still relatively more accurate than the previous measures for the Bosniaks and the Serbs. In part, the inaccuracies of the Statistical Office's figures for the Serb community might stem from the considerable migration between 2005 (when the estimation was done) and 2008 (when the elections were held). However, if we apply the same models to the results of the 2006 parliamentary elections, our measure considerably outperforms the population figures of the Statistical Office for two indicators (and is only moderately less accurate regarding a third indicator).

To compare our model to the OSCE estimates, we restricted the validation procedure to the 30 municipalities for which OSCE estimates are available. The OSCE appears to under-estimate the number of ethnic Bosniaks in some municipalities, and our results out-perform the OSCE on both indicators for this group. We also seem to have a slightly better estimation of the ethnic Serbs,

[8] The measure is calculated separately for each ethnic group, as follows: $dev_e = \frac{\sum_{i=1}^{N} \left[\max\left(vote_{e,i} - p_{e,i}, 0\right)\right]^2}{N}$, where N is the number of municipalities, $vote_{e,i}$ is the vote share of the ethno-nationalist party of group e in municipality i.

[9] Especially in Usora, where we count 69% ethnic Croats and 31% ethnic Bosniaks, but 92% of the votes are cast for ethnic Croat parties. However, the number of registered births is very low, with 26 births on 7000 inhabitants in 2008.

judging by the adjusted R^2. Both the OSCE and our figures seem to under-estimate the number of the ethnic Croats in certain municipalities, but here the OSCE slightly outperforms our measures.

Quite clearly, our *full model* improves the estimation, compared to the *naive model*, which does not correct for different birth rates across municipalities and ethnic groups. The *full model* clearly outperforms the *naive model* on the squared deviation measure. Not correcting for the different birth rates across ethnic groups, the naive model tends to under-estimate the size of the ethnic Croat and the ethnic Serb community. The underestimation of the naive model is systematic and the degree similar for all municipalities. This also explains why the naive model and the full model have almost the same explanatory power (adjusted R^2) – even though the naïve model leads to contradictions when we employ the squared deviation measures.

Table 3: Results of construct validity tests

			Bosniaks			**Croats**			**Serbs**			
			previous est.	*naive model 2008*	***full model 2008***	*previous est.*	*naive model 2008*	***full model 2008***	*previous est.*	*naive model 2008*	***full model 2008***	N
Construct validity I	adj R^2	elections 2008	0.729	0.699	**0.692**	0.922	0.916	**0.921**	0.485	0.956	**0.956**	77
		elections 2006	0.820	0.810	**0.801**	0.942	0.915	**0.913**	0.423	0.744	**0.754**	77
		30 OSCE municipalities, 2008	0.680	0.741	**0.728**	0.955	0.977	**0.975**	0.835	0.935	**0.935**	30
Construct validity II	squared deviation (dev_e)	elections 2008	15.5	0.1	**0.7**	7.3	33.9	**13.9**	23.6	5.1	**0.8**	77
		elections 2006	20.6	0.8	**3.5**	14.8	30.3	**20.0**	0.3	0.7	**0.5**	77
		30 OSCE municipalities, 2008	9.1	0.0	**0.8**	13.9	67.7	**30.0**	0.0	2.1	**1.0**	30

3. Conclusions

Even though it has been argued that ethnicity plays a major role in a variety of social and political processes, data on the ethnic identification of the population is not available in many countries. In the context of the post-war situation of Bosnia and Herzegovina, the question has been so sensitive and politically controversial that it has led to a substantial delay of the population census. Bosnia and Herzegovina is divided between ethnic Bosniaks, Serbs, Croats, and 'others', and the ethnic identity of the citizens is crucial in many domains of public life. The lack of information on the distribution of ethnic identities within the country and its municipalities complicates studies on the political and social processes of the post-war period, such as for instance studies on the relative success of multi-ethnic parties to attract votes from different ethnic groups.

This paper estimates the distribution of ethnic identities at the level of municipalities of Bosnia and Herzegovina. To our knowledge, this is the first attempt to provide such estimates based on a transparent and objective procedure, and it provides publically available results. Results are available for 77 out of 79 municipalities of the Federation of Bosnia and Herzegovina, one out of the two main political units of the country. The proposed estimation procedure is based on the ethnic identity of newly born children's parents. Our estimation procedure applies the multiplier method, extrapolating from a sub-sample to the whole population. Going beyond this literature, it corrects for unequal birth rates across ethnic groups and across municipalities. Out estimates might thus provide valuable information for quantitative scholars of various research areas, such as post-conflict politics, party politics, and electoral behaviour. Multiple tests of robustness and validity indicate that our procedure leads to a good approximation of the distribution of the ethnic identities in the municipalities of Bosnia and Herzegovina. Most importantly, our numbers also seem to outperform previous measures by the OSCE and the Statistical Office of the Federation. However, in municipalities with very low birth figures, the measurements are less accurate, as they seem to under-estimate the size of the Croat population.

Research based on reliable data can not only improve our understanding of the political and social causes and consequences of this situation, but also possibly contribute to its eventual alleviation, for instance by providing insights into the circumstances under which post-war social divides between social groups can be overcome.

4. References

Bakke, K. M., Cao, X., O'Loughlin, J., & Ward, M. D. (2009). Social distance in Bosnia-Herzegovina and the North Caucasus region of Russia: Inter and intra-ethnic attitudes and identities. *Nations and Nationalism, 15*(2), 227-253.

Ball, P., & Asher, J. (2002). Statistics and Slobodan: Using Data Analysis and Statistics in the War Crimes Trial of Former President Milosevic. *Chance, 15*(4), 17-24.

Birnir, J. K. (2007). *Ethnicity and Electoral Politics*. Cambridge: Cambridge University Press.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge (MA): MIT Press.

Bochsler, D. (2013). Radicalizing Electoral System Effects on Support for Nationalist Hardliners. *Nations and Nationalism, 19*(1), 21-43.

Braumoeller, B. F. (2006). Explaining Variance; Or, Stuck in a Moment We Can't Get Out Of. *Political Analysis, 14*, 268-290.

Buhaug, H., & Rød, J. K. (2006). Local determinants of African civil wars, 1970–2001. *Political Geography, 25*(3), 315-335.

Caspersen, N. (2004). Good Fences Make Good Neighbours? A Comparison of Conflict-Regulation Strategies in Postwar Bosnia. *Journal of Peace Research, 41*(5), 569-588.
Chandra, K. (2006). What is Ethnic Identity and Does It Matter? *Annual Review of Political Science, 9*, 397-424.
Davidian, M., & Carroll, R. J. (1987). Variance Function Estimation. *Journal of the American Statistical Association, 82*(400), 1079-1091.
Federalni zavod za statistiku. (2003-2010). Živorođeni prema nacionalnoj - etničkoj pripadnosti majke i oca. Sarajevo: Federalni zavod za statistiku.
Federalni zavod za statistiku. (2005). *Procjene ukupnog broja prisutnih stanovnika, po nacionalnoj strukturi, stanje 30.06.2005.* Sarajevo: FZS.
Federalni zavod za statistiku. (2009). *Demografska statistika 2008*. Sarajevo: FzS.
Franck, R., & Rainer, I. (2012). Does the Leader's Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharian Africa. *American Political Science Review, 106*(2), 294-325.
Grofman, B., & Handley, L. (1995). Racial context, the 1968 Wallace vote, and southern presidential dealignment: evidence from North Carolina and elsewhere. In M. Eagles (Ed.), *Spatial and Contextual Models in Political Research.* (pp. 151-162). London: Taylor and Francis.
International Crisis Group. (2003). *Building Bridges in Mostar*. Brussels: International Crisis Group.
Keech, W. R. (1968). *The Impact of Negro Voting. The Role of the Vote in the Quest for Equality.* Chicago: Rand Mc Nailly.
Key, V. O. (1949). *Southern Politics in State and Nation*. New York: Vintage.
Lee, T. (2009). Between Social Theory and Social Science Practice. Toward a New Approach to the Survey Measurement of 'Race'. In R. Abdelal, Y. M. Herrera, A. I. Johnston & R. McDermott (Eds.), *Measuring Identity* (pp. 113-144). Cambridge: Cambridge University Press.
Putnam, R. D. (2007). E Pluribus Unum: Diversity and Community in the Twenty-first Century. The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies, 30*(2), 137-174.
Raleigh, C., & Hegre, H. (2009). Population size, concentration, and civil war. A geographically disaggregated analysis. *Political Geography, 28*(4), 224-238.
Toal, G., & Dahlman, C. T. (2011). *Bosnia Remade. Ethnic Cleansing and its Reversal*. Oxford: Oxford University Press.
UNAIDS/WHO Working Group on Global HIVAIDS and STI Surveillance. (2003). *Estimating the Size of Populations at Risk for HIV: Issues and Methods*: Family Health International (FHI).
WHO. (2003). *HIV/AIDS Epidemiological Surveillance Update for the WHO African Region 2002*. Harare: WHO.
Williams, M., & Husk, K. (2013). Can we, should we, measure ethnicity? *International Journal of Social Research Methodology, 16*(4), 285-300.

Supplementary material (online):
Categorization of political parties and electoral lists

Bosniak	Bosanskohercegovačka Patriotska Stranka-Sefer Halilović
	Bošnjačka Seljačka Stranka
	BOSS - Bosanska Stranka-Mirnes Ajanović
	Demokratska Narodna Zajednica BiH
	Lista za Čapljinu SBiH-BPNS
	Narodna Bošnjačka Stranka
	Savez SDA-S BiH
	Socijaldemokratska Unija Bosne i Hercegovine
	Stranka Demokratske Akcije
	Stranka Demokratske Aktivnosti za Evropsku BiH A-SDA
	Stranka za Bosnu i Hercegovinu
Croats	HDZ 1990 HSS-NHI HSP
	Hrvatska Demokratska Zajednica 1990
	Hrvatska Demokratska Zajednica BiH
	Hrvatska Koalicija – HSS - NHI HDZ BiH HDZ 1990 HNZ
	Hrvatska Koalicija (HDZ BiH HDZ 1990 HKDU BiH)
	Hrvatska Koalicija HDZ 1990 HSS – NHI
	Hrvatska Koalicija HDZ 1990 HSS-NHI Za Žepče
	Hrvatska Koalicija HDZ 1990-HKDU
	Hrvatska Koalicija Kantona Sarajevo
	Hrvatska Koalicija za Drvar HDZ 1990-HDZ BiH-HSP Đapić-Jurišić
	Hrvatska Koalicija za Pougarje HSP-Đ-J HDZ1990 HSS-NHI
	Hrvatska Koalicija za Ravno HDZ1990 HDZBiH
	Hrvatska Koalicija za Travnik
	Hrvatska Koalicija za Zenicu HDZ BiH HDZ 1990 HSS-NHI HNZ
	Hrvatska Koalicija-HDZ BiH-HSS-NHI-HDZ 1990
	Hrvatska Kršćanska Demokratska Unija BiH
	Hrvatska Narodna Zajednica
	Hrvatska Stranka Prava Bosne I Hercegovine - Đapić – Dr. Jurišić
	Hrvatsko Zajedništvo Herceg-Bosne
	Hsp BiH Đapić-Dr.Jurišić-HNZ BiH
	Hss – NHI
	Koalicija Hrvatskih Stranaka HDZ BiH HSS-NHI HDZ 1990 HKDU BiH
	Koalicija: HNZ-HDU
	Koalicijska Lista HDZ BiH i HDZ 1990
Serbian	Demokratska Stranka Naroda Srpske
	Demokratski Narodni Savez
	Demokratski Pokret Srpske-Depos
	Koalicija Za Bolji Život-DNS i DS
	Narodna Demokratska Stranka
	Partija Demokratskog Progresa
	Savez Nezavisnih Socijaldemokrata - SNSD - Milorad Dodik
	Savez Srpskih Stranaka Srebrenice: SDS-PDP-SP i SRS RS
	SDS-DNS-PDP-SRS RS
	SNSD-PDP
	Socijalistička Partija
	Srpska Demokratska Stranka
	Srpska Demokratska Stranka 1990 - Izvorna - Pokret za Srpsku
	Srpska Napredna Stranka Republike Srpske
	Srpska Narodna Radikalna Stranka - Banja Luka
	Srpska Radikalna Stranka Dr Vojislav Šešelj
	Srpska Radikalna Stranka Republike Srpske
	Srpska Stranka Republike Srpske
	Srpski Pokret Obnove Republike Srpske
Mixed-ethnic	Građanska Demokratska Stranka BiH
	Narodna Stranka Radom za Boljitak
	Naša Stranka
	Socijaldemokratska Partija Bosne i Hercegovine