

In the Nation We Trust: National Identity as a Substitute for Religion

Kenneth Harttgen, Matthias Opfinger *

ETH Zurich, Leibniz University Hannover

This version: 2 February 2012

Abstract

We construct an index for national identity using information from the World Values Survey on peoples' attitudes concerning politics and to the state itself. We then analyze the relationship between our new measure of national identity and social heterogeneity. The results indicate that religious diversity is significantly and positively related to national identity, whereas other variables proxying social heterogeneity are not. We argue that national identity is a substitute for religion. At high levels of religious diversity people do not identify with their religious group. They search other objects of identification offering common values and norms. Hence, people identify at the national level. Furthermore, democratic institutions and mobility throughout the country affect national identity positively.

JEL-classification: J15, O1, Z12

Keywords: National Identity, Social Heterogeneity, Religious Diversity, Common Beliefs

*ETH Zurich, Nadel, Voltastrasse 24, 8092 Zurich, Switzerland. Phone: +41-(0)44-6329825. E-Mail: harttgen@nadel.ethz.ch. Leibniz University Hannover, Department of Economics. Koenigsworther Platz 1, 30167 Hannover, Germany. Phone: +49-(0)511-7625655. E-Mail: opfinger@vwl.uni-hannover.de. We thank Carl-Johan Dalgaard, Holger Strulik and Sebastian Vollmer for helpful comments.

1 Introduction

“[National identity] provides the sole vision and rationale of political solidarity today, one that commands popular assent and elicits popular enthusiasm. All other visions, all other rationales, appear wan and shadowy by comparison. They offer no sense of election, no unique history, no special destiny. These are the promises which nationalism for the most part fulfills, and the real reasons why so many people continue to identify with the nation.” Anthony D. Smith (1991), *National Identity*.

Anthony D. Smith devotes a whole book to the concept of national identity. He argues how nation states evolved and developed over time and describes how the population identifies with the nation. The prerequisites for a functioning nation are listed and the formation process of a national identity is explained. However, the whole book is kept on a theoretical, in parts philosophical, and abstract level. As the reader proceeds throughout the book he comes across numerous interesting hypotheses concerning determinants of national identity. Unfortunately, these hypotheses cannot be tested due to a missing measure for national identity. The present paper offers a solution to this problem. We develop an index for national identity, which makes the abstract concept of national identity measurable and comparable across countries. In a later stage of this study we use this index to evaluate the impact of social heterogeneity on national identity and in order to identify other possibly important covariates.

The concept of identity has been introduced into the economics literature by Akerlof and Kranton’s (2000) influential article. They add identity to the utility function and can thereby explain why some outcomes are optimal for a group of people while they might be detrimental to others. Identity can affect economic outcomes through changes in the payoffs from own actions or from the actions of others. Furthermore, the choice of an identity can affect economic behavior or changing social norms might alter identity-based preferences.

Bisin et al. (2010) disentangle the identity formation process and propose two mechanisms. Cultural conformity claims that minority groups adopt inclusive identities and that they integrate into their social surroundings. Contrary, cultural distinctiveness proposes that minorities keep their identities and reduce interactions with individuals from other ethnic groups. The authors find empirical evidence supporting the idea of cultural distinctiveness. Darity et al. (2006) provide an evolutionary model that discusses inter- and intraracial interactions based on identities and explains under which circumstances racist or individualistic identities are formed.

Bodenhorn and Ruebeck (2003) analyze the identity formation process of African Americans in the Antebellum South and find that the size of the community determines the probability of choosing a

mixed-race identity. Similarly, Austin-Smith and Fryer (2005) find that the cost of leaving the peer group explains the education decisions of African Americans. By “acting white”, i.e. becoming better educated, African Americans lose their former identity and choose another one. Battu et al. (2007) come to a very similar conclusion when they investigate the job market decisions of non-whites. Peer pressure and the possible gains of adopting a white identity heavily influence the job market decisions. Constant and Zimmerman (2008) and Constant et al. (2009) develop a measure of ethnic identity and investigate why migrants might choose an identity that favors the country of origin over their host country.

Miles and Rochefort (1991), Calhoun (1993), Jones (1997) and Bond (2006) analyze the relationship between ethnic diversity and national identity and suggest that ethnic diversity and national identity influence each other. However, Masella (2011) finds that ethnic heterogeneity does not have a significant effect on national identity. Smith (1991) proposes that a nation needs a single political culture, a unified economy, and a unified legal code (p.69). Social homogeneity is a prerequisite for all these characteristics. Hence, it is feasible to assume that social heterogeneity has a detrimental effect on the formation process of a national identity. This is the main hypothesis which we want to test. We employ several measures of social heterogeneity in order investigate their impact on national identity.

Ethnic or ethnolinguistic fractionalization has been used as an explanatory variable in many different settings. Easterly and Levine (1997) find that ethnolinguistic fractionalization helps explain Africa’s unfavorable growth experience because ethnic diversity complicates public policies and leads to worse institutions. La Porta et al. (1999) argue that ethnic fractionalization reduces government performance and Alesina et al. (1999) show that higher ethnic diversity leads to a smaller amount of public goods provided.

Collier and Hoeffler (1998), Vanhanen (1999), and Fearon and Laitin (2003) investigate the economic determinants of civil wars. Vanhanen (1999) finds that higher heterogeneity increases the probability of the occurrence of civil wars. Collier and Hoeffler (1998) argue that the effect is not linear. First, the probability of a civil war rises with higher levels of ethnic diversity but after a maximum is reached further increases in ethnic diversity reduce the probability. Thus, Montalvo and Reynal-Querol (2005, 2005a) calculate a measure of ethnic polarization. They follow an idea of Esteban and Ray (1994). This index reaches its maximum if the society consists of two large rivaling ethnic groups. Montalvo and Reynal-Querol (2005) show that ethnic diversity has a direct negative impact on the GDP growth rate, whereas the impact of ethnic polarization and religious polarization is indirect through reduced investment, increased government consumption, or a higher probability of civil wars. In a following paper (2005a) they analyze the direct impact on civil wars and find that increasing ethnic polarization has a significant positive impact on the

occurrence of civil wars.

We will use measures of ethnic and religious diversity, as well as polarization to investigate the impact of social heterogeneity on national identity. Using data from the *World Values Survey* (WVS), we suppose to find that social heterogeneity has a detrimental effect on national identity because a highly fragmented society will find it harder to identify with the same values and norms. If the society consists of different social groups they will be distinct from another. Hence, the members of the separate groups will prefer to identify with their group instead of identifying with their nation. Furthermore, we will test more suggestions from Smith (1991) and try to find other possible correlates of national identity. Testable hypotheses are that democratic institutions and mobility throughout the country have a positive impact on national identity and that geographical factors also influence the formation of a national identity.

The paper is organized as follows. Section 2 describes our index of national identity. We will show the calculation of the index and present the results for those countries used in our analysis. The data and methodology we use to analyze the relationship of our index of national identity to social heterogeneity are presented in Section 3. Section 4 shows the empirical results, which are discussed in Section 5. Section 6 briefly concludes.

2 National Identity Index

This section describes the construction of the index of national identity. The motivation behind constructing an index of national identity is twofold. First, the reason for constructing an index rather than analyzing several potential indicators of national identity separately is that we aim to identify a common underlying factor captured by a set of indicators of political and national interests and orientations, namely the national identity of a person. Second, within a uni-dimensional index, we are directly able to analyze the determinants of national identity.

To derive the national identity index, we apply a principal component analysis. Hereby, we closely follow the approach of Filmer and Pritchett (2001) and Sahn and Stifel (2003) to construct an asset index of material welfare based on the possession of housing durables. The main idea of this approach is to construct an aggregated uni-dimensional index over the range of different dichotomous indicators of political and national interests and orientations capturing the national identity of a person.

The approach of aggregating different variables to a uni-dimensional index is widely used in the economic and social literature. For example, Filmer and Pritchett (2001) and Sahn and Stifel (2001) propose an asset index based on the possession of household assets and dwelling characteristics as a proxy of material welfare of households in cases where no information on household income or expenditure are at

hand. Paldam and Grundlach (2009) use an index approach to derive a measure of religiosity to analyze the religious transition over time.¹

Principal component analysis is, among others, an aggregation technique to identify from a set of variables those linear combinations that best capture the common information behind the variables (Filmer and Scott, 2008). This means that we assume that specific variables on social and political participation and political interest can explain the long-term national identity of a person measured by the maximum variance in the variables:

$$A_i = b_1 a_{i1} + b_2 a_{i2} + \dots + b_k a_{ik} \quad (1)$$

$$a_{ik} = \beta_k c_i + u_{ik} \quad (2)$$

where A_i is the national identity index, the a_{in} 's refer to the respective variable of the person i recorded as dichotomous variables in the data and the b 's are the respective weights for each variable used to aggregate the indicators to a one-dimensional index and that are to be estimated. In our model this means that the identity variable k , identified by a_{ik} is a linear function of an unobserved factor, which in our case is national identity c_i . The relationship between the underlying identity variable k in cv_i is given by β_k plus a noise component u_{ik} , where both terms have to be estimated (Sahn and Stifel 2000).²

For the estimation of the weights we rely on the first principal component as our national identity index.³ The principal component analysis is structured by a set of equations where the identity variable is related to a set of latent factors:

$$\tilde{a}_{1i} = v_{11}A_{1i} + v_{12}A_{2i} + \dots + v_{1k}A_{ki}$$

...

¹A large body of literature exists using an asset index to explain inequalities in educational outcomes (e.g. Ainsworth and Filmer 2006), health outcomes (e.g. Bollen et al. 2002), child mortality (e.g. Sastry 2004) when data on income or expenditure is missing. In addition, asset indices are used to analyze changes and determinants of poverty (e.g. Stifel and Christiaensen 2007).

²The model is based on the following assumptions: (i): persons are distributed iid; (ii): $E(u_i|c_i) = 0$; (iii): $V(u_i) = \text{Diag}(\sigma_1^2, \dots, \sigma_K^2)$.

³An alternative way to estimate the weights to derive the aggregated index is a factor analysis employed, for example, by Sahn and Stifel (2001) and Paldam and Grundlach (2009). However, the two estimation methods show very similar results. For a systematic overview of different aggregation techniques, see Filmer and Scott (2008).

$$\tilde{a}_{ki} = v_{k1}A_{1i} + v_{k2}A_{2i} + \dots + v_{kk}A_{ki} \quad (3)$$

where the \tilde{a} 's are the k indicators (the a 's in equation 1) normalized by their mean and their standard deviations; A are the k principal components and the v 's are the weights that relate the principal components to the indicators of national identity (Filmer and Scott 2008). After the weights v have been estimated, the inversion of the equation system (3) yields the following set of equations:

$$A_{1i} = b_{11}\tilde{a}_{1i} + b_{21}\tilde{a}_{2i} + \dots + b_{k1}\tilde{a}_{ki}$$

...

$$A_{ki} = b_{1k}\tilde{a}_{1i} + b_{2k}\tilde{a}_{2i} + \dots + b_{kk}\tilde{a}_{ki} \quad (4)$$

The equation for the first principal component is the equation with the maximal variance. The weights that are used to aggregate the identity variables into a one-dimensional national identity index are given by the set $(b_{11}, b_{21}, \dots, b_{k1})$.

Since we are not interested in the analysis of changes in national identity over time, we pool all survey years of the World Values Survey and calculate the national identity index for the whole sample. Table A1 in the appendix shows the results by country for those countries where information on all variables that enter the index are available.

We use micro data from the World Value Survey to calculate the index. As components for the national identity index we include 8 dichotomous variables presented in Table 1 that are assumed to capture the national identity of a person.⁴ Table 1 shows the mean values of the indicators, the standard deviation, the number of observation, and the scoring factors of the principal component analysis. For example, 43.3% of respondents answered to be very interested in politics and 16.6% of the respondents have stated their willingness to fight for their country. The mean value of the identity index is close to zero with a range of around -2 to 2. The distribution of the index is presented in Figure 1 in the appendix. In total, the first component explains 21% of the covariance.

⁴We also tried to derive the index based on more variables. But since this sample size is then reduced a lot and since the results differ not very much, we decide to derive the index for as many countries as possible. With the underlying indicators, we are able to calculate the index for 62 countries in the sample, for almost 100,000 persons. The country specific mean values of the identity index as well as the standard deviation, and the number of observations are presented in Table ??

Table 1: Summary statistics

Indicators	Score	Index		Obs
		Mean	SD	
Politics very important in life (=1)	0.424	0.405	0.491	308,225
Willingness to fight for the country:yes (=1)	0.166	0.732	0.443	256,999
Interest in politics: very or somewhat interested (=1)	0.433	0.467	0.499	309,409
Signing a petition: have done or might do (=1)	0.200	0.663	0.473	310,689
Confidence: parliament: a great deal and quite a lot (=1)	0.290	0.414	0.493	312,863
Confidence: justice system: a great deal and quite a lot (=1)	0.237	0.516	0.500	269,203
Geographical groups belonging to first: country (=1)	0.092	0.337	0.473	254,120
Very proud of nationality (=1)	0.104	0.562	0.496	332,747
Index value (mean)	0.006			95,277
Index value (sd)	1.008			
Index value (min)	-2.030			
Index value (max)	2.000			
% of the covariance explained by the first principal component	0.210			
Eigenvalue of first principal component	1.683			

Source: WVS; calculations by the authors.

Note: Indicators of national identity and index statistics

3 Data and Methodology

3.1 Data

In the previous section we described the construction of our measure of national identity. This index will be the dependent variable throughout the whole analysis.

The aim of the present study is to analyze the relationship between national identity and diversity within the population. For this purpose we employ four different measures of social heterogeneity, ethnic diversity, ethnic polarization, religious diversity, and religious polarization. These four variables are the regressors of main interest. Information on ethnic diversity is taken from Alesina et al. (2003). Montalvo and Reynal-Querol (2005) calculate measures for ethnic and religious polarization. They use these measures to analyze their respective effects on economic development. Diversity and polarization behave similarly in homogeneous societies. Polarization reaches its maximum if the society consists of two equally sized groups, i.e. diversity equals 0.5. Further increases in diversity reduce the index of polarization because conflict between different groups becomes less probable if the society is composed of a large number of small groups. Data on religious diversity is taken from Opfinger (2011) who relies on data from the World Christian Encyclopedia (Barrett, Kurian, Johnson, 2002) and also includes non-religious and atheistic as separate denominations. This method guarantees that the whole population is assigned a denomination which overcomes the weaknesses of previous studies using religious diversity as described in Voas et al. (2002).

Due to data availability, information on income is taken from the Maddison (2010) online database. We choose the year 1973 to reduce the risk of reverse causality and because it was the only year with data for the single Soviet nations prior to the dissolution of the former USSR. We control for income because it is sensible to assume that economic development might have an impact on the formation of a national

identity. Hence, income is included in all regressions.

The second aim of this study is to find other variables that might affect the formation of a national identity. The political environment might be one, which we control for with the Polity score from the Polity IV database and with measures of political rights and civil liberties from the freedomhouse.org webpage. We also use education as a control variable. Information is taken from the Barro and Lee (2010) dataset. The variable we use is the percentage of the population aged 25 years and older that completed secondary education.

Geographic variables might also influence national identity. Consequently, we include a set of geographic variables in the regressions. Data on area size, if the country is landlocked, and the number of neighboring countries is taken from the CIA World Factbook. Information on the population size comes from the UN statistics division. Population density is calculated by dividing population through area. In order to evaluate if the countries under investigation have been a colony or under communist rule we rely on the country information from the CIA World Factbook. Data on soil quality is taken from Nunn and Puga (2009).

As a proxy for openness we use the trade share and constructed trade share from 1985 (Frankel and Romer, 1999). Smith (1991) argues that mobility throughout the country is an important factor for the formation of a national identity. We use kilometers of paved roads as a proxy for physical mobility. Information is once more taken from the CIA World Factbook. We calculate kilometers of paved roads per inhabitant, per square kilometer, and per inhabitant per square kilometer. As a proxy for nonphysical mobility we use the number of phone lines per 100 inhabitants. Information is taken from the World Bank's World Development Indicators.

For the instrumental variables estimations to take into account potential endogeneity problems, we rely on data on the disease environment from Fincher and Thornhill (2008) and on climatic conditions which is taken from Sachs (2001).

3.2 Methodology

The present study is the first attempt to assign a numeric value to the concept of national identity. In order to discover possible correlates we run regressions with a whole set of control variables. Opfinger (2011) finds that religiosity decreases with rising levels of religious diversity but is positively related to ethnic diversity. He argues that religiosity and national identity might be substitutes. As a consequence, ethnic diversity should have a negative impact on national identity. We also use other measures of social diversity to reveal national identity's main explanatory factors.

We cannot use panel data methods since our main explanatory variables, ethnic and religious hetero-

geneity, are observed at only one point in time. Thus, we average our national identity index over the five WVS waves and run cross-country regressions. We use data at the country level because we want to analyze the impact of ethnic and religious heterogeneity. Within one country the deviations would be too marginal in order to gain inclusive insights about the role diversity plays for the formation of a national identity.

In order to reduce the risk of reverse causality, most explanatory variables are used from 1973, the year for which we have income data for all countries. Other control variables, such as the dummy variables for former communist rule, landlocked countries, and former colonial rule, do not change at all over time. The estimated regressions are of the form:

$$NI_i = \alpha + \beta \cdot ethdiv_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i \quad (5)$$

$$NI_i = \alpha + \beta \cdot ethpol_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i \quad (6)$$

$$NI_i = \alpha + \beta \cdot reldiv_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i \quad (7)$$

$$NI_i = \alpha + \beta \cdot relpol_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i, \quad (8)$$

where NI_i is the index of national identity in country i , $ethdiv_i$ is ethnic diversity in country i , $ethpol_i$ is ethnic polarization in country i , $reldiv_i$ is religious diversity in country i , and $relpol_i$ is religious polarization in country i . y_i is income in country i , X_i is a vector of the other control variables and ε_i is the error term.

The coefficient of main interest in each case is β . A positive coefficient means that heterogeneity has a positive effect on national identity, whereas a negative β would imply that national identity decreases if heterogeneity rises.

The explanatory variables of main interest could all be subject to endogeneity bias. Due to the missing time dimension we are not able to use fixed effects models, which would decrease the importance of omitted variable bias. Consequently, we have to rely on two stage least squares estimation with instruments for the variables that we use to measure social heterogeneity. Fincher and Thornhill (2008) propose that the disease environment in a country could explain religious diversity. However, their argumentation seems to fit better to ethnic diversity. Groups that share the same immunity pattern to specific diseases should come together and separate themselves from other groups. This should be more true for ethnic groups than for religious groups. In fact, a common factor of Fincher and Thornhill's (2008) disease and pathogen variables is a suitable instrument for ethnic diversity. The exclusion restriction should hold because it is not obvious how the disease environment should affect the formation of a national identity if not through ethnic diversity.

Ethnic polarization can be instrumented by the percentage of the population living in temperate cli-

matic zones. The relationship is negative, which means that a larger share of people living in temperate climatic zones reduces ethnic polarization. Migration routes of the human population since its beginning in East Africa might explain this pattern. The further ethnic groups moved away from the cradle of mankind the further the different groups drifted apart. As a consequence, areas in temperate climatic zones are typically inhabited by one large ethnic group and some smaller groups which leads to low levels of ethnic polarization. The exclusion restriction should hold in this case as well, as it is plausible to assume that the climatic conditions do not directly affect national identity.

We instrument religious diversity today with past religious diversity. Barrett, Kurian, and Johnson (2002) provide data on religious diversity for the year 1900. It is feasible to assume that past religious diversity influences diversity today and in fact the first stage regressions reveal that past rates of religious diversity are a strong instrument for religious diversity today. The exclusion restriction demands that past rates of religious diversity do not affect national identity today but through present religious diversity. Since many countries of our sample did not exist in their present form in 1900 it is highly unlikely that past rates of religious diversity have a direct impact on national identity. In addition, historic events in the course of the twentieth century might have changed the perception of nationality so that variables that go back further in time should not influence national identity today. As a consequence, religious diversity from the year 1900 is a valid instrument for religious diversity nowadays.

The common factor of diseases and pathogens is also a suitable instrument for religious polarization as it enters very significantly in the first stage regressions. We can assume that the disease environment affects national identity only through social heterogeneity so that the exclusion restriction should not be violated.

3.3 Summary Statistics

Table 2 presents summary statistics for the variables used in this study. We are able to calculate national identity for 62 countries. At the country level, our index ranges from -0.515 to 0.709 index points.⁵ Mean and median are both slightly negative. The lowest value of national identity is observed in Argentina. Bangladesh reveals the highest value of national identity. Taiwan's level comes closest to the mean and the median lies between the observations for Georgia and Croatia.

The index of ethnic diversity is by construction distributed between zero and one, and measures the probability that two randomly drawn persons belong to the same ethnic group. The most homogeneous country in our sample is South Korea with an index value of 0.002. Nigeria is the most ethnically diverse

⁵The mean of the national identity index differs to the mean in Table 1 because, it refers to the country mean and not to the whole micro data sample.

country, 0.851. The index of religious diversity can be interpreted in the same way. Turkey is the most homogeneous country (0.055) whereas South Korea is the most diverse in this case with an index value of 0.848. Ethnic and religious polarization are at their maximum if the society is made up of two large rivaling groups. The indexes reveal low levels if diversity is very low or very high. The country with the lowest value of ethnic polarization in our sample is Portugal (0.02). Belgium is the ethnically most polarized country with an index value of 0.871. There are 20 countries in our sample where religious polarization is zero. This means that the whole population belongs to the same denomination. Religious polarization in the Dominican Republic reaches an index value of 0.999.

Table 2: Summary Statistics

<i>Variable</i>	<i>Number of obs.</i>	<i>Mean</i>	<i>Median</i>	<i>Standard Deviation</i>	<i>Minimum</i>	<i>Maximum</i>
National Identity	62	-0.005	-0.063	0.296	-0.515	0.709
Ethnic Diversity	61	0.315	0.274	0.213	0.002	0.851
Ethnic Polarization	42	0.421	0.385	0.261	0.020	0.871
Religious Diversity	62	0.475	0.496	0.236	0.055	0.848
Religious Polarization	62	0.175	0.024	0.284	0.000	0.999
Log of Income '73	62	8.702	8.755	0.731	6.210	9.810
Polity score '73	56	-0.393	-7	8.263	-9	10
Political Rights '73	59	4.136	4	2.381	1	7
Civic Liberties '73	59	4.136	3	2.278	1	7
Secondary Education	54	23.439	23.485	11.991	0.580	56.470
Area in square km	62	1,186,486	127,438	3,029,922	316	17,098,242
Population in million	61	47.289	10.137	123.078	0.267	915.992
Population Density	61	135.749	86.970	190.013	2.322	1,176.827
Landlocked	62	0.178	0	0.385	0	1
# Neighboring Countries	62	3.790	4	2.847	0	14
% Fertile Soil	62	47.504	49.789	22.383	0.073	100.000
Former Colony	62	0.258	0	0.441	0	1
Former Communist Country	62	0.371	0	0.487	0	1
Trade Share '85	48	64.631	57.185	40.150	15.040	211.940
Constructed Trade Share '85	48	24.171	16.165	41.170	2.560	281.290
Roads per 1000 inh	55	8.027	6.859	6.670	0.193	25.945
Roads per sq km	56	0.871	0.378	1.146	0.011	6.373
Roads per inh per sq km	55	11,227.75	894.334	36,024.88	1.711	162,287.9
Phone Lines per 100 inh	56	16.786	10.000	14.621	0.000	58.000

Source: Calculations by the authors

Note: The table shows the number of observations, mean, median, standard deviation, minimum, and maximum for all variables.

Income in 1973 is measured in logarithmic terms. It ranges from 6.21 which equals 497.7 1990 US-Dollars in Bangladesh to 9.81 which equals 18,215 1990 US-Dollars in Switzerland. The polity score by construction lies between -10 for total autocracies to +10 for full democracies. There are four countries where the polity score in 1973 is -9. These are Albania, Brazil, the Philippines, and Portugal. There are 17

countries with a polity score of +10. These are the Western European countries, the Western off-shoots, and Japan. The indexes of political rights and civil liberties are rescaled so that a higher score correlates to higher political rights and civil liberties, respectively. The distribution over the countries is similar to that of the polity score.

Secondary education measures the percentage of the population aged 25 years and older that completed secondary education. In India only 0.58 percent of the population completed secondary education. The highest value is observed in Armenia with 56.47 percent. The smallest country in our sample is Malta with an area of 316 square kilometers compared to the largest country Russia, more than 17 million square kilometers. Concerning the population the smallest country is Iceland with 267,000 inhabitants, India is the largest country with slightly less than 916 million inhabitants. Population density is lowest in Australia with 2.32 inhabitants per square kilometer and reaches 1176.83 in Malta. There are eight islands without land boundaries in our sample. The Russian Federation has the highest number of neighboring countries (14). In Norway only 0.07 percent of the soil is fertile whereas the value for Malta is 100 percent. Landlocked, former colony, and former Communist country are dummy variables.

Trade share and the constructed trade share measure the openness of a country and are reported in Frankel and Romer (1999). The actual trade share is lowest in India, whereas for the constructed trade share the value for the United States is smallest. For both measures the highest value is reported for Luxembourg.

Bangladesh has only 0.19 kilometers of paved roads per 1000 inhabitants. The highest value in this category is reported in Ireland with 25.95 kilometers per 1000 inhabitants. In Brazil there are only 0.01 kilometer of paved roads per square kilometer compared to 6.34 kilometers in Malta. The lowest value of paved roads per inhabitant per square kilometer comes from Malta with 1.71 compared to the highest value of more than 162,287 kilometers per inhabitant per square kilometer in the United States. The number of phone lines per 100 inhabitants ranges from 0 in Bangladesh and India to 58 in Sweden.

4 Results

In this section, we present the results of various regressions with which we want to analyze the effect of social heterogeneity on national identity. The explanatory variables of main interest are ethnic and religious diversity and polarization. Furthermore, we add a large set of control variables in order to investigate what else might influence the formation of a national identity.

4.1 Ethnic Diversity

First, we analyze the effect of ethnic diversity on national identity. As described before, a larger value of ethnic diversity represents higher heterogeneity. A positive β implies that higher ethnic diversity leads to higher levels of national identity. In the next subsection we present the results of cross-country OLS regressions before we turn to instrumental variable estimations in section 4.1.2.

4.1.1 OLS Results

Table 3: OLS regression results, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic Diversity	-0.32*	-0.22	-0.17	-0.23	-0.33	-0.35*	-0.34*	-0.25	-0.32*	-0.28
	(-1.80)	(-1.26)	(-0.95)	(-1.27)	(-1.60)	(-1.89)	(-1.92)	(-1.25)	(-1.73)	(-1.53)
Log of Income '73	-0.03	-0.14**	-0.08	-0.07	-0.05	-0.03	0.01	0.00	-0.03	-0.03
	(-0.54)	(-2.44)	(-1.50)	(-1.27)	(-0.83)	(-0.63)	(0.11)	(0.00)	(-0.50)	(-0.60)
Polity score '73		0.02***								
		(3.72)								
Pol. Rights '73			0.05**							
			(2.56)							
Civ. Liberties '73				0.04**						
				(2.03)						
Secondary Educ.					0.00					
					(0.69)					
Area in square km						0.00				
						(0.72)				
Population in mill							0.00*			
							(1.81)			
Pop. Density								0.00		
								(0.95)		
Landlocked									-0.04	
									(-0.38)	
# neighboring coun.										-0.02
										(-1.44)
cons	0.35	1.25**	0.60	0.55	0.49	0.38	0.02	0.05	0.33	0.43
	(0.75)	(2.56)	(1.29)	(1.16)	(0.95)	(0.82)	(0.05)	(0.08)	(0.71)	(0.93)
N	61	56	59	59	54	61	60	60	61	61
R ² adj.	0.02	0.22	0.10	0.06	0.01	0.01	0.06	0.02	0.01	0.04

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Ethnic Diversity	-0.32*	-0.32*	-0.28	-0.32	-0.33	-0.25	-0.31	-0.31	-0.10
	(-1.78)	(-1.76)	(-1.58)	(-1.48)	(-1.55)	(-1.25)	(-1.54)	(-1.61)	(-0.55)
Log of Income '73	-0.05	-0.03	-0.04	-0.04	-0.04	-0.04	-0.01	-0.04	-0.30***
	(-0.90)	(-0.44)	(-0.80)	(-0.60)	(-0.66)	(-0.54)	(-0.14)	(-0.73)	(-4.00)
% fertile soil	0.00								
	(-1.27)								
Former Colony		0.00							
		(-0.01)							
Former Communist coun.			-0.13*						
			(-1.72)						
Trade Share '85				0.00					
				(0.24)					
Constr. Trade Share '85					0.00				
					(0.61)				
Roads per 1000 inh.						0.00			
						(0.59)			
Roads per sq km							-0.03		
							(-0.74)		
Roads per inh. per sq km								0.00	
								(1.65)	
Phone Lines per 100									0.02***
									(4.48)
cons	0.63	0.34	0.49	0.42	0.45	0.38	0.19	0.46	2.39***
	(1.23)	(0.59)	(1.07)	(0.80)	(0.85)	(0.64)	(0.38)	(0.87)	(3.97)
N	61	61	61	47	47	54	55	54	55
R ² adj.	0.03	0.00	0.05	-0.01	0.00	-0.01	-0.01	0.03	0.28

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

Table 3 shows the results of the OLS regressions. Ethnic diversity and income are kept in all estimations. With regards to ethnic diversity the main result is easily observed at first sight. The coefficient on ethnic diversity fails to reach statistical significance in more than half of the estimations. However, the sign is negative, which hints in the direction that higher levels of ethnic diversity lead to lower values in our index of national identity. The size of the coefficient is, except for column 19, always between -0.2 and -0.35 which implies that an increase in the index of ethnic diversity by 0.1 reduces our measure of national identity by only 0.02 to 0.035 index points.

Income does also not enter statistically significantly in 17 out of the 19 regressions. Only in columns 2 and 19 income has a significant negative effect on national identity. In columns 2 through 4 we control for different variables that proxy democratic institutions. All these variables have a positive and significant impact on national identity at the five percent level. The polity score from the Polity IV database reaches the highest level of statistical significance. In column 5 we control for education and find that this variable does not seem to affect national identity.

We control for geographical factors that might influence the formation of a national identity in columns 6 through 11 and find that only population size has a positively significant impact at the ten percent level. Thus, countries with a larger population reveal higher levels of national identity. In columns 12 and 13 we test whether the history of the country matters for national identity. In column 12 we find, that the colonial past does not affect national identity. In contrast, countries that have been under communist rule exhibit lower levels of national identity which is statistically significant at the ten percent level.

We control for the openness of the economy in columns 14 and 15 with the trade shares calculated by Frankel and Romer (1999). Neither variable comes close to statistical significance at conventional levels. We include different measures of mobility in columns 16 through 19 to test Smith's (1991) proposition that mobility throughout the country is important for the formation of a national identity. We account for physical mobility in columns 16 through 18 in which we add different measures of the amount of paved roads within a country. Paved roads per inhabitant per square kilometer is the only variable that comes close to statistical significance as it falls short of the ten percent significance level only very slightly. The effect is positive. In column 19 we control for the number of phone lines per 100 inhabitants as a proxy for non-physical mobility which turns out to be highly statistically significant. The estimated effect is positive, an increase by ten phone lines per 100 inhabitants raises national identity by 0.2 index points.

As a first general finding, we can state that ethnic diversity does not seem to be an important factor in explaining national identity. With regards to the other control variables, democratic institutions and mobility across the country had the strongest positive effects on national identity. The impact of population

size is also slightly positive whereas a communist past reduces national identity.

4.1.2 Instrumental Variable Results

Since the OLS regressions might suffer from endogeneity issues we rely on instrumental variables in order to gain unbiased results and to establish possible causal effects of ethnic diversity on national identity.

Table 4: Instrumental Variable regression results, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic Diversity	-0.28 (-0.61)	-0.19 (-0.42)	-0.11 (-0.25)	-0.12 (-0.27)	-0.06 (-0.10)	-0.48 (-0.95)	-0.66 (-1.51)	-0.04 (-0.07)	-0.31 (-0.73)	-0.04 (-0.09)
Log of Income '73	-0.03 (-0.51)	-0.13** (-2.49)	-0.08 (-1.56)	-0.07 (-1.29)	-0.06 (-0.90)	-0.04 (-0.71)	0.00 (-0.04)	0.02 (0.22)	-0.03 (-0.50)	-0.03 (-0.50)
Polity Score '73		0.02*** (3.43)								
Pol. Rights '73			0.05** (2.35)							
Civ. Liberties '73				0.04** (2.04)						
Secondary Educ.					0.00 (0.82)					
Area in square km						0.00 (0.78)				
Population in mill							0.00* (1.91)			
Pop. Density								0.00 (0.85)		
Landlocked									-0.04 (-0.39)	
# neighboring coun.										-0.02 (-1.58)
cons	0.32 (0.62)	1.24** (2.40)	0.57 (1.21)	0.50 (1.00)	0.43 (0.82)	0.47 (0.86)	0.19 (0.36)	-0.19 (-0.21)	0.33 (0.65)	0.33 (0.64)
N	60	56	59	59	54	60	60	60	60	60

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Ethnic Diversity	-0.49 (-1.19)	-0.26 (-0.56)	-0.51 (-1.34)	-0.39 (-1.13)	-0.36 (-1.03)	-0.18 (-0.33)	-0.45 (-0.78)	-0.58 (-1.15)	0.56 (0.93)
Log of Income '73	-0.06 (-1.06)	-0.03 (-0.45)	-0.05 (-0.89)	-0.04 (-0.68)	-0.04 (-0.70)	-0.04 (-0.58)	-0.01 (-0.15)	-0.05 (-0.83)	-0.36*** (-3.72)
% fertile soil	0.00 (-1.36)								
Former Colony		-0.01 (-0.05)							
Former Communist coun.			-0.12 (-1.51)						
Trade Share '85				0.00 (0.26)					
Constr. Trade Share '85					0.00 (0.64)				
Roads per 1000 inh.						0.01 (0.62)			
Roads per sq km							-0.03 (-0.76)		
Roads per inh. per sq km								0.00* (1.78)	
Phone Lines per 100									0.02*** (3.93)
cons	0.78 (1.39)	0.33 (0.56)	0.61 (1.24)	0.47 (0.88)	0.47 (0.88)	0.37 (0.65)	0.25 (0.46)	0.59 (1.04)	2.68*** (3.84)
N	60	60	60	46	46	54	54	54	55

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

Ethnic diversity is instrumented by a common factor of the disease and pathogen variables proposed by Fincher and Thornhill (2008). The results of the instrumental variable regressions are presented in

Table 4. The control variables remain in the same order as in Table 3. The results from the instrumental variable estimations support the main insights from the OLS regressions. The significance level on the coefficients on ethnic diversity are reduced further. Obviously, ethnic diversity does not have a causal effect on national identity. Income is again only significant in columns 2 and 19. The presence of democratic institutions, however, seems to be an important factor for the formation of a national identity. The polity score and the indexes of political rights and civil liberties remain positive and statistically significant. The polity score is significant at the one percent level, the other two variables at the five percent level.

Population size and roads per inhabitant per square kilometer also stay positive and significant at the ten percent level. Compared to Table 3 communist past loses statistical significance at the ten percent level, but the sign is still negative and the coefficient of a similar magnitude. The number of phone lines per 100 inhabitants maintains its positive and significant impact on national identity. Raising the number of phone lines by 10 per 100 inhabitants results in an increase in national identity by 0.2 index points.

We can conclude that ethnic diversity does not seem to have a causal effect on the level of national identity. It fails to gain statistical significance in the instrumental variable regressions. The presence of democratic institutions and mobility throughout the country seem to be important factors for the formation of a national identity. In the next subsections we will explore if these relationships hold when we substitute ethnic diversity for other variables of social divide.

4.2 Ethnic Polarization

In this subsection we investigate the relationship between ethnic polarization and national identity. As described before, ethnic polarization reaches its maximum if a majority faces a large minority group. If the society is ethnically very homogeneous or very heterogeneous ethnic polarization is small. Similar to the subsection on ethnic diversity we will present OLS results first which will be followed by instrumental variable estimations. The control variables remain the same as in Tables 3 and 4.

4.2.1 OLS Results

Table 5 shows the results of OLS regressions in which ethnic polarization is the variable of main interest. A positive coefficient implies that rising levels of ethnic polarization increase national identity.

In contrast to ethnic diversity ethnic polarization appears to have a significant impact on national identity. The coefficient is negative and statistically significant at least at the ten percent level in 18 out of 19 regressions, in 16 cases at the five percent level. The size of the coefficient varies between -0.36 and -0.63 in the regressions in which ethnic polarization is significant. Raising ethnic polarization by 0.1 index points then implies a reduction in our measure of national identity by 0.036 to 0.063 index points.

This amounts to the difference in national identity for example between the United States and Bosnia and Herzegovina or between Germany and Iceland.

Table 5: OLS regression results, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic Polarization	-0.47** (-2.51)	-0.38* (-1.94)	-0.36* (-1.79)	-0.42** (-2.12)	-0.40** (-2.20)	-0.61*** (-3.22)	-0.48** (-2.56)	-0.46** (-2.24)	-0.47** (-2.49)	-0.45** (-2.36)
Log of Income '73	-0.02 (-0.40)	-0.13* (-1.93)	-0.07 (-1.06)	-0.06 (-0.84)	-0.14** (-2.04)	-0.04 (-0.65)	0.02 (0.26)	-0.01 (-0.21)	-0.02 (-0.42)	-0.02 (-0.43)
Polity score '73		0.02** (2.66)								
Pol. Rights '73			0.04 (1.48)							
Civ. Liberties '73				0.03 (1.09)						
Secondary Educ.					0.01** (2.43)					
Area in square km						0.00** (2.16)				
Population in mill							0.00 (1.46)			
Pop. Density								0.00 (0.22)		
Landlocked									0.03 (0.20)	
# neighboring coun.										-0.01 (-0.70)
cons	0.44 (0.87)	1.25** (2.25)	0.60 (1.14)	0.56 (1.04)	1.22** (2.17)	0.56 (1.16)	0.07 (0.12)	0.35 (0.55)	0.45 (0.88)	0.49 (0.95)
N	42	37	40	40	40	42	41	41	42	42
R ² adj.	0.10	0.24	0.10	0.08	0.21	0.18	0.13	0.08	0.08	0.09

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Ethnic Polarization	-0.53*** (-2.86)	-0.43** (-2.08)	-0.47** (-2.49)	-0.47** (-2.48)	-0.47** (-2.48)	-0.40* (-1.80)	-0.47** (-2.35)	-0.54** (-2.72)	-0.28 (-1.58)
Log of Income '73	-0.05 (-0.93)	-0.05 (-0.57)	-0.03 (-0.47)	-0.02 (-0.35)	-0.03 (-0.46)	-0.05 (-0.51)	0.00 (-0.03)	-0.05 (-0.75)	-0.28*** (-3.08)
% fertile soil	0.00* (-1.74)								
Former Colony		-0.07 (-0.42)							
Former Communist coun.			-0.10 (-0.62)						
Trade Share '85				0.00 (-0.15)					
Constr. Trade Share '85					0.00 (0.41)				
Roads per 1000 inh.						0.01 (0.62)			
Roads per sq km							-0.04 (-1.04)		
Roads per inh. per sq km								0.00** (2.21)	
Phone Lines per 100									0.02*** (3.22)
cons	0.91 (1.62)	0.68 (0.89)	0.49 (0.95)	0.43 (0.85)	0.46 (0.90)	0.55 (0.77)	0.30 (0.53)	0.66 (1.14)	2.32*** (3.28)
N	42	42	42	42	42	37	38	37	39
R ² adj.	0.14	0.08	0.09	0.08	0.08	0.05	0.07	0.17	0.29

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

Similar to the results of the previous subsection, income does not appear to be an important factor for national identity. It reaches statistical significance in only three regressions. The polity score remains statistically significant at the two percent level. The relationship between democratic institutions and national identity is still positive. However, the coefficients on political rights and civil liberties are no

longer significant.

In column 5 we control again for secondary education. In this setting it has a positive and significant relationship to national identity. An increase in the secondary education completion rate by ten percentage points implies that our measure of national identity rises by 0.1 index points. In columns 6 to 11 we add again the geographical variables. We find that population size is no longer significant, but instead, area size becomes significant. The coefficient is positive which implies that national identity is higher in larger countries. Concerning mobility we find, once more, that physical mobility, proxied by kilometers of paved roads per inhabitant per square kilometer, as well as non-physical mobility, proxied by the number of phone lines, to positively affect the level of national identity. The other control variables do not enter significantly.

4.2.2 Instrumental Variable Results

Again, we have to take into account that ethnic polarization might suffer from endogeneity bias. We use the percentage of the population in each country living in temperate climatic areas as an instrument for ethnic polarization. The results are presented in Table 6.

The instrumental variable regressions show important differences to the OLS results. When ethnic polarization is instrumented by the percentage of the population in each country living in temperate climatic areas, ethnic polarization does no longer have a significant impact on national identity. Apparently the OLS results are biased away from zero. In all our instrumental variable regressions ethnic polarization fails to reach significance at the ten percent level.

Once more, income enters significantly in four out of 19 regressions. The results on democratic institutions hold. In column 2 the polity score enters significantly at the one percent level. In columns 3 and 4 the indexes of political rights and civil liberties are significant at the five and ten percent levels, respectively. The findings on secondary education are also supported by the instrumental variable results. Secondary education has a positive and significant impact on the level of national identity. An increase in the secondary education completion rate by 10 percentage points implies a rise in national identity by 0.14 index points.

Concerning the geographical variables we find area size to be significantly related to our measure of national identity. In contrast to the OLS regressions population density now also has a positive and significant effect on national identity. National identity rises by 0.14 points if population density increases by one standard deviation. Once more, we find support for the idea that mobility throughout the country is an important factor. In columns 18 and 19 kilometers of paved roads per inhabitant per square kilometer and the number of phone lines per 100 inhabitants have a positive and significant effect on national identity.

Increasing kilometers of paved roads per inhabitant per square kilometer by one standard deviation raises national identity by 0.1 index points, or a third of a standard deviation. 10 phone lines more per 100 inhabitants raise national identity by 0.19 index points, almost two-thirds of a standard deviation.

Table 6: Instrumental variable regression, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic Polarization	-0.33 (-1.01)	-0.13 (-0.43)	-0.13 (-0.41)	-0.28 (-0.94)	-0.24 (-0.85)	-0.63* (-1.72)	-0.32 (-1.07)	-0.08 (-0.23)	-0.32 (-1.02)	-0.33 (-1.03)
Log of Income '73	-0.04 (-0.68)	-0.13** (-2.13)	-0.13* (-1.84)	-0.11 (-1.51)	-0.17** (-2.51)	-0.05 (-0.98)	0.00 (-0.03)	0.02 (0.36)	-0.04 (-0.69)	-0.04 (-0.79)
Polity score '73		0.02*** (2.94)								
Pol. Rights '73			0.07** (2.23)							
Civ. Liberties '73				0.05* (1.71)						
Secondary Educ.					0.01*** (2.70)					
Area in square km						0.00* (1.93)				
Population in mill							0.00 (1.31)			
Pop. Density								0.00* (1.85)		
Landlocked									0.02 (0.10)	
# neighboring coun										-0.02 (-1.06)
cons	0.53 (1.01)	1.23** (2.28)	0.90* (1.67)	0.84 (1.52)	1.41** (2.53)	0.72 (1.44)	0.17 (0.29)	-0.23 (-0.34)	0.53 (1.01)	0.64 (1.23)
N	41	37	39	39	39	41	40	40	41	41

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Ethnic Polarization	-0.47 (-1.47)	-0.08 (-0.15)	-0.42 (-1.44)	-0.27 (-0.88)	-0.28 (-0.90)	-0.13 (-0.34)	-0.28 (-0.78)	-0.49 (-1.44)	-0.11 (-0.42)
Log of Income '73	-0.06 (-1.03)	-0.11 (-1.04)	-0.04 (-0.80)	-0.05 (-0.80)	-0.05 (-0.85)	-0.10 (-1.08)	-0.02 (-0.30)	-0.07 (-1.12)	-0.37*** (-4.15)
% fertile soil	0.00 (-1.39)								
Former Colony		-0.17 (-0.80)							
Former Communist coun			-0.12 (-0.76)						
Trade Share '85				0.00 (0.46)					
Constr. Trade Share '85					0.00 (0.79)				
Roads per 1000 inh						0.01 (1.08)			
Roads per sq km							-0.02 (-0.28)		
Roads per inh. per sq km								0.00** (2.19)	
Phone Lines per 100									0.02*** (4.12)
cons	0.89 (1.59)	1.06 (1.27)	0.63 (1.23)	0.55 (1.04)	0.58 (1.10)	0.91 (1.24)	0.37 (0.60)	0.84 (1.44)	3.02*** (4.35)
N	41	41	41	41	41	36	37	36	38

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

The OLS results suggested that ethnic polarization might have an important effect on the formation of a national identity. The instrumental variable regressions revealed that the results were due to endogeneity and that ethnic polarization does not affect national identity. The results of this subsection imply

that democratic institutions, mobility throughout the country, education, and country size appear to be important correlates of national identity.

4.3 Religious Diversity

In the previous two subsections, we analyzed the impact of ethnicity on national identity. It appears that ethnic heterogeneity does not affect the formation of a national identity. But other levels of social differences might cause variation in national identity across countries. Opfinger (2011) proposes that religious diversity affects levels of religiosity negatively. We follow the argumentation of Bruce (2000) who suggests that religiosity and national identity might be substitutes. Therefore, we explore the effect of religious diversity on national identity in this subsection. We expect to find a positive relationship between religious diversity and national identity, as religious diversity decreases religiosity and religiosity and national identity are supposed to be substitutes.

4.3.1 OLS Results

We repeat the estimations of the previous subsections, but replace the variables measuring ethnic heterogeneity by religious diversity. A positive sign on the coefficient implies that higher levels of religious diversity increase national identity. The results are presented in Table 7.

Table 7: OLS regression results, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Religious Diversity	0.24 (1.50)	0.33** (2.16)	0.32** (2.06)	0.30* (1.88)	0.29 (1.53)	0.24 (1.43)	0.22 (1.34)	0.27* (1.68)	0.25 (1.56)	0.24 (1.51)
Log of Income '73	-0.03 (-0.55)	-0.15*** (-2.80)	-0.11* (-1.99)	-0.10 (-1.66)	-0.06 (-0.92)	-0.03 (-0.54)	0.00 (0.04)	0.01 (0.19)	-0.03 (-0.49)	-0.03 (-0.65)
Polity score '73		0.02*** (4.41)								
Pol. Rights '73			0.06*** (3.38)							
Civ. Liberties '73				0.05*** (2.69)						
Secondary Educ.					0.00 (0.24)					
Area in square km						0.00 (0.04)				
Population in mill							0.00 (1.55)			
Pop. Density								0.00* (1.68)		
Landlocked									-0.07 (-0.68)	
# neighboring coun.										-0.02 (-1.61)
cons	0.13 (0.28)	1.18** (2.51)	0.58 (1.29)	0.49 (1.06)	0.36 (0.68)	0.13 (0.28)	-0.15 (-0.31)	-0.28 (-0.54)	0.11 (0.24)	0.25 (0.55)
N	62	56	59	59	54	62	61	61	62	62
R ² adj.	0.01	0.26	0.15	0.09	0.01	-0.01	0.03	0.04	0.00	0.03

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Religious Diversity	0.27 (1.67)	0.24 (1.47)	0.33** (2.05)	0.40** (2.07)	0.44** (2.25)	0.25 (1.49)	0.22 (1.28)	0.17 (0.92)	0.29** (2.06)
Log of Income '73	-0.05 (-0.99)	-0.04 (-0.61)	-0.05 (-1.00)	-0.05 (-0.85)	-0.05 (-0.94)	-0.07 (-0.92)	-0.02 (-0.35)	-0.04 (-0.66)	-0.33*** (-4.68)
% fertile soil	0.00 (-1.53)								
Former Colony		-0.03 (-0.29)							
Former Communist coun.			-0.18** (-2.33)						
Trade Share '85				0.00 (0.64)					
Constr. Trade Share '85					0.00 (1.17)				
Roads per 1000 inh.						0.01 (1.06)			
Roads per sq km							0.00 (-0.12)		
Roads per inh. per sq km								0.00 (1.09)	
Phone Lines per 100									0.02*** (5.07)
cons	0.46 (0.93)	0.23 (0.40)	0.35 (0.79)	0.23 (0.47)	0.26 (0.53)	0.39 (0.67)	0.08 (0.15)	0.25 (0.48)	2.46*** (4.32)
N	62	62	62	48	48	55	55	55	56
R ² adj.	0.03	-0.01	0.08	0.03	0.05	0.00	-0.02	0.00	0.34

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

With regards to the role of religious diversity the regressions do not deliver a clear result. The coefficient is positive as we would expect if national identity and religiosity were indeed substitutes. Hence, higher levels of religious diversity appear to be correlated with higher levels of national identity. But this finding is statistically significant at least at the ten percent level only in 9 out of 19 estimations. In those regressions in which religious diversity enters significantly the size of the coefficient varies between 0.27 and 0.44. An increase in religious diversity by 0.1 index points therefore correlates with a rise in national identity by 0.027 to 0.044 index points.

As before, income does not seem to have an important effect on national identity. It enters significantly only in four regressions. In columns 2 through 4 we control again for democratic institutions. All three variables reveal a positive coefficient which is statistically significant at the one percent level. In contrast to the findings from the previous subsection secondary education and the geographical variables do not enter significantly. But instead, countries that have formerly been under communist rule reveal lower levels of national identity. This finding is significant at the five percent level.

Openness, proxied by the trade share, is again insignificant. Concerning the hypothesis of the importance of mobility we find in this setting that only the number of phone lines per 100 inhabitants is significant. The variables on paved roads do not enter significantly.

4.3.2 Instrumental Variable Results

As before, we have to account for possible endogeneity. We use rates of religious diversity in 1900 as instrument for religious diversity today. The results of these estimations are shown in Table 8.

Table 8: Instrumental variable regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Religious Diversity	0.65*** (2.83)	0.72*** (3.25)	0.72*** (3.23)	0.72*** (3.13)	0.64** (2.53)	0.69*** (2.80)	0.65*** (2.76)	0.69*** (2.97)	0.67*** (2.88)	0.65*** (2.87)
Log of Income '73	-0.04 (-0.82)	-0.18*** (-3.17)	-0.14** (-2.47)	-0.13** (-2.17)	-0.06 (-0.96)	-0.04 (-0.77)	-0.02 (-0.37)	0.00 (-0.01)	-0.04 (-0.75)	-0.05 (-0.93)
Polity score '73		0.02*** (4.54)								
Pol. Rights '73			0.07*** (3.77)							
Civ. Liberties '73				0.06*** (3.12)						
Secondary Educ.					0.00 (-0.42)					
Area in square km						0.00 (-0.62)				
Population in mill							0.00 (1.22)			
Pop. Density								0.00* (1.88)		
Landlocked									-0.09 (-0.90)	
# neighboring coun.										-0.02 (-1.56)
cons	0.07 (0.15)	1.22** (2.55)	0.64 (1.39)	0.55 (1.17)	0.28 (0.54)	0.04 (0.09)	-0.14 (-0.28)	-0.38 (-0.72)	0.05 (0.10)	0.19 (0.41)
N	62	56	59	59	54	62	61	61	62	62

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Religious Diversity	0.73*** (3.14)	0.65*** (2.82)	0.82*** (3.44)	0.80*** (2.95)	0.83*** (3.05)	0.68*** (2.87)	0.68*** (2.78)	0.65** (2.44)	0.58*** (3.01)
Log of Income '73	-0.08 (-1.35)	-0.05 (-0.74)	-0.08 (-1.40)	-0.07 (-1.24)	-0.08 (-1.31)	-0.10 (-1.32)	-0.04 (-0.66)	-0.04 (-0.69)	-0.35*** (-4.87)
% fertile soil	0.00* (-1.76)								
Former Colony		-0.01 (-0.10)							
Former Communist coun.			-0.24*** (-2.83)						
Trade share '85				0.00 (1.08)					
Constr. Trade Share '85					0.00* (1.69)				
Roads per 1000 inh.						0.01 (1.36)			
Roads per sq km							0.02 (0.39)		
Roads per inh per sq km								0.00 (0.19)	
Phone lines per 100									0.02*** (5.01)
cons	0.46 (0.90)	0.11 (0.18)	0.36 (0.76)	0.23 (0.46)	0.27 (0.55)	0.44 (0.73)	0.01 (0.01)	0.06 (0.11)	2.48*** (4.33)
N	62	62	62	48	48	55	56	55	56

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

The instrumental variable regression results show very important and remarkable differences to the OLS findings. If religious diversity is instrumented by past rates of religious diversity it is significant in 19 out of 19 estimations. It is significant at the one percent level in 17 regressions and at the five percent level in two more.

Furthermore, the coefficient is larger compared to the OLS results. It varies between 0.58 and 0.83. The average value of the coefficient is 0.7. Hence, an increase in religious diversity by 0.1 index points

raises national identity on average by 0.07 index points. If religious diversity changes by one standard deviation, national identity reacts with a change by 0.165 index points in the same direction which is more than half of a standard deviation. The difference in religious diversity between the most homogeneous and most diverse countries amounts to 0.793 index points. This value implies a difference in national identity of more than half an index point which is the difference between Argentina, the country with the lowest value of national identity, and Switzerland which is in the upper half of countries in ascending order of national identity. It also amounts to the difference in national identity between, for example, the United States and France. Apparently, religious diversity has a strong impact on the formation of a national identity. Countries that are religiously very homogeneous reveal low levels of national identity whereas national identity is high if the population is religiously diverse.

Once again, income is only significant in four out of 19 regressions. Columns 2 through 4 show the by now well-known pattern. Democratic institutions have a positive and significant impact on the level of national identity. All three variables are significant at the one percent level.

In this setting secondary education does not have a significant effect on national identity. In column 5 the sign of the coefficient even becomes negative. Concerning the geographical variables in columns 6 through 11 we find that only two of them are statistically significant. Population density seems to increase national identity whereas the percentage of fertile soil has a significant negative relationship to national identity. The size of the coefficient is fairly small. Ten percentage points more fertile soil reduces national identity by only 0.03 index points.

Again, a communist past has a strong and negative effect on national identity, as shown in column 13. For the first time, openness has a significant effect on our measure of national identity. In column 15 the constructed trade share of Frankel and Romer (1999) enters positively. Increasing trade openness by ten percentage points raises national identity by 0.02 index points, a rather small effect. Supporting the findings from Table 7 only non-physical mobility appears to have an important effect on national identity, as becomes apparent in column 19. Kilometers of paved roads fall short of statistical significance at conventional levels.

The findings of this subsection, especially the instrumental variable results, support the idea that social heterogeneity affects the formation of a national identity. Religious diversity enters significantly and the size of the effect is also remarkable. Apparently, national identity is higher if the society is religiously highly fragmented. People seem to choose to identify on a national level only if this is not possible through religion. We will come back to this point in the discussion of the results. Furthermore, democratic institutions and non-physical mobility have a robust positive effect on national identity. A communist

past appears to have a detrimental effect on our measure of national identity.

4.4 Religious Polarization

As a last possible measure of social heterogeneity we use religious polarization. This concept is comparable to ethnic polarization. It reaches its maximum if the society consists of only two large religious groups. Religious polarization is small if a large majority faces a number of small minorities or if a large number of equally sized groups coexist. As before, a positive sign on the coefficient implies that higher levels of religious polarization increase national identity.

4.4.1 OLS Results

We follow the same procedure as before. First, we show the results of OLS regressions of national identity on religious polarization and the remaining control variables. The results are presented in Table 9.

Apparently, religious polarization does not affect the level of national identity. Since, we found religious diversity to have a significant impact on national identity this result is not surprising. Religious polarization does not reach statistical significance in any of the 19 regressions. Income is also statistically significant in only two estimations.

Table 9: OLS regression results, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Religious Polarization	0.13 (0.84)	-0.02 (-0.14)	0.06 (0.36)	0.04 (0.25)	0.22 (1.28)	0.12 (0.78)	0.08 (0.51)	0.16 (0.99)	0.12 (0.76)	0.10 (0.67)
Log of Income '73	0.00 (0.02)	-0.14** (-2.09)	-0.07 (-1.16)	-0.06 (-0.92)	-0.03 (-0.49)	0.00 (-0.03)	0.02 (0.41)	0.04 (0.68)	0.00 (0.02)	-0.01 (-0.14)
Polity score '73		0.02*** (3.92)								
Pol. Rights '73			0.05*** (2.73)							
Civ. Liberties '73				0.04** (2.04)						
Secondary Educ.					0.00 (1.06)					
Area in square km						0.00 (0.28)				
Population in mill							0.00 (1.51)			
Pop. Density								0.00 (1.56)		
Landlocked									-0.04 (-0.37)	
# neighboring coun.										-0.02 (-1.51)
cons	-0.04 (-0.07)	1.19** (2.06)	0.44 (0.82)	0.37 (0.66)	0.15 (0.27)	-0.01 (-0.03)	-0.26 (-0.48)	-0.45 (-0.79)	-0.03 (-0.06)	0.12 (0.24)
N	62	56	59	59	54	62	61	61	62	62
R ² adj.	-0.02	0.20	0.09	0.04	-0.01	-0.04	0.00	0.01	-0.03	0.00

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Religious Polarization	0.05 (0.28)	0.19 (1.13)	0.01 (0.09)	0.06 (0.32)	0.06 (0.36)	0.26 (1.45)	0.21 (1.20)	0.18 (1.00)	0.13 (0.90)
Log of Income '73	-0.03 (-0.48)	-0.02 (-0.35)	-0.03 (-0.53)	-0.02 (-0.22)	-0.02 (-0.24)	-0.02 (-0.34)	0.02 (0.29)	-0.01 (-0.17)	-0.29*** (-3.78)
% fertile soil	0.00 (-1.06)								
Former Colony		-0.10 (-0.84)							
Former Communist coun.			-0.14 (-1.64)						
Trade Share '85				0.00 (0.23)					
Constr. Trade Share '85					0.00 (0.55)				
Roads per 1000 inh.						0.01 (1.13)			
Roads per sq km							-0.01 (-0.27)		
Roads per inh. per sq km								0.00 (1.28)	
Phone Lines per 100									0.02*** (4.79)
cons	0.36 (0.57)	0.18 (0.32)	0.32 (0.59)	0.13 (0.20)	0.13 (0.22)	0.10 (0.17)	-0.19 (-0.33)	0.06 (0.09)	2.23*** (3.52)
N	62	62	62	48	48	55	55	55	56
R ² adj.	-0.02	-0.02	0.01	-0.06	-0.06	0.00	-0.03	0.00	0.29

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

Concerning the remaining covariates, we find again that democratic institutions have a positive and significant impact on national identity. The polity score and the index of political rights are significant at the one percent level, while the index of civil liberties is significant at the five percent level. The number of phone lines, our proxy variable for non-physical mobility, is also positively and significantly related to national identity. The other control variables do not reach significance at conventional levels.

4.4.2 Instrumental Variables Results

Also in this last case we have to deal with possible endogeneity issues concerning religious polarization. Again, the common factor of the disease and pathogen variables from Fincher and Thornhill (2008) might be used as an instrument. The first stage statistics show that this is a very strong instrument. The results of the instrumental variable regressions are presented in Table 10.

Our main insight on religious polarization is preserved in the instrumental variable regressions. It does not have a significant impact on the level of national identity. Income is again significant in only two out of 19 regressions.

Once more, we find democratic institutions to have a significant effect on national identity. The coefficient is positive and significant at the one percent level for the polity score and political rights and at the five percent level for civil liberties. The instrumental variable estimations reveal significance of some of the geographical variables. In column 7 population size has a positive impact on national identity which is significant at the five percent level. Also, the number of neighboring countries and the percentage of the land that is fertile are significant at the ten percent level. Both variables have a negative impact on the

level of national identity.

Table 10: Instrumental Variable regression results, dependent variable: national identity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Religious Polarization	-0.15 (-0.55)	-0.11 (-0.41)	-0.06 (-0.25)	-0.07 (-0.27)	-0.03 (-0.10)	-0.26 (-0.88)	-0.44 (-1.37)	-0.01 (-0.05)	-0.19 (-0.65)	-0.02 (-0.08)
Log of Income '73	-0.04 (-0.62)	-0.15** (-1.98)	-0.10 (-1.31)	-0.09 (-1.07)	-0.06 (-0.83)	-0.06 (-0.87)	-0.04 (-0.53)	0.02 (0.24)	-0.04 (-0.65)	-0.03 (-0.46)
Polity Score '73		0.02*** (3.90)								
Pol. Rights '73			0.05*** (2.85)							
Civ. Liberties '73				0.05** (2.12)						
Secondary Educ.					0.00 (0.79)					
Area in square km						0.00 (0.67)				
Population in mill							0.00** (2.04)			
Pop. Density								0.00 (1.53)		
Landlocked									-0.07 (-0.69)	
# neighboring coun.										-0.02* (-1.66)
cons	0.39 (0.62)	1.36* (1.92)	0.64 (1.02)	0.58 (0.85)	0.46 (0.70)	0.57 (0.87)	0.35 (0.55)	-0.19 (-0.30)	0.43 (0.68)	0.34 (0.57)
N	61	56	59	59	54	61	61	61	61	61

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Religious Polarization	-0.39 (-1.09)	-0.18 (-0.40)	-0.44 (-1.20)	-0.54 (-1.07)	-0.36 (-0.89)	-0.10 (-0.28)	-0.24 (-0.72)	-0.37 (-1.04)	0.25 (1.08)
Log of Income '73	-0.12 (-1.33)	-0.04 (-0.59)	-0.11 (-1.32)	-0.11 (-1.07)	-0.09 (-0.97)	-0.06 (-0.73)	-0.04 (-0.56)	-0.10 (-1.15)	-0.27*** (-3.30)
% fertile soil	0.00* (-1.73)								
Former Colony		0.02 (0.11)							
Former Communist coun.			-0.26** (-2.09)						
Trade Share '85				0.00 (-0.51)					
Constr. Trade Share '85					0.00 (0.17)				
Roads per 1000 inh.						0.01 (0.69)			
Roads per sq km							-0.02 (-0.53)		
Roads per inh. per sq km								0.00* (1.67)	
Phone Lines per 100									0.02*** (4.75)
cons	1.32 (1.43)	0.36 (0.59)	1.13 (1.39)	1.14 (1.13)	0.87 (1.00)	0.46 (0.67)	0.42 (0.61)	0.88 (1.14)	2.03*** (2.95)
N	61	61	61	47	47	55	55	55	56

Source: calculations by the authors

Note: t-statistics in parentheses; *, **, *** denotes statistical significance at the ten, five, and one percent levels.

In column 13 we find, as already mentioned before, that a communist past decreases national identity, which is significant at the five percent level. With regards to mobility throughout the country, column 18 shows that kilometers of paved roads per inhabitant per square kilometer, as a proxy for physical mobility positively affects national identity. Column 19 supports our result that non-physical mobility, as measured by phone lines, significantly increases the level of national identity. The other covariates again do not enter

significantly.

Obviously, religious polarization does not determine national identity. This result is not surprising, as we found before that religious diversity has a significant impact on national identity. The other results are in line with our previous findings. Democratic institutions, mobility throughout the country, and size are positively related to national identity. A communist past reduces national identity.

To sum up, we find that social heterogeneity is one important factor in explaining different levels of national identity. Religious diversity is the force driving these results. It is positively and significantly related to national identity. The implication of this finding will be discussed in the next section. In addition, we found that democratic institutions and mobility throughout the country have positive and significant effects on the formation of a national identity. The country size also seems to have an impact on national identity, as in several regressions either country size or population size entered significantly. A communist past appears to decrease national identity. Income, education, openness, colonial past, and other geographical variables seem to play, if at all, only a very minor role.

5 Discussion

The present study pursues two objectives. It is a first attempt to assign numeric values to the idea of national identity. Second, we try to reveal, which factors might drive the formation of a national identity. To achieve our first goal we use data from the World Values Survey. We extract questions that refer to national identity. Due to data availability on some of the questions we are restricted to those eight items presented in Table 1 which relate to the respondents' attitudes towards politics and the state itself. On the whole we have more than 95,000 observations included in our index, that is on average 1,500 respondents per country.

In the empirical section, we investigate the relationship between national identity and social heterogeneity. In addition, we also look for other covariates that might influence the formation of a national identity. Bruce (2000) argues that religiosity and national identity might be substitutes and Opfinger (2011) finds that ethnic diversity has a large positive impact on religiosity which lets him conclude that people identify with their religious group if they are not able to identify with their nationality if the society is ethnically too fragmented.

We find that ethnic diversity does not have a significant impact on our measure of national identity. Instead, religious diversity has a strong and causal positive effect on national identity. It appears that religiosity and national identity are in fact substitutes. Religion seems to be the most favorite object of identification in a society. Opfinger's (2011) finding of ethnic diversity's impact on religiosity appears after

controlling for the level of religious diversity. This means that when holding the level of religious diversity constant, increasing levels of ethnic diversity raise the level of religiosity.

We interpret our findings on social heterogeneity in the following way. If the society is religiously homogeneous it chooses to identify with their religious group. By sharing the same belief people send out signals that they also share a common set of values, which is probably the most important object to identify with. Only if the society is religiously highly fragmented the people do not identify with their religion. They start to doubt that their neighbors share the common set of religious values and norms and hence, decrease their religious involvement. On a second level, people still feel some kind of closeness to the people in their environment. Since they cannot identify with their religious group when people adhere to different denominations they choose to identify on another level with broader common norms. This is the level of national identity. People of the same nationality can identify with their country which might be due to political, social, or cultural factors. Consequently, religious diversity has a direct impact on national identity, whereas ethnic diversity does not.

Two simple examples can make this more easily understandable. First, take two persons of the same nationality, say German. These persons will identify with their religious group as long as they adhere to the same denomination. They share a common set of values, which is based on their religious beliefs. Two persons of the same nationality might not be able to identify with their religion if they adhere to two different denominations, say Protestant and Catholic. Hence, higher religious diversity decreases the importance of religion. But still these people share a broader set of values or cultural beliefs which are based on their national heritage and lets them form a national identity. As a consequence, higher religious diversity, which leads to less importance of religion, increases national identity.

As a second example, consider two US American citizens where one is Caucasian and the other is African American. No matter what their religion is these persons can at least identify on a national level. They share a common set of values which is based on being a US national. This example can help understand why ethnic differences might not affect the formation of a national identity.

Masella (2011) does also not find a significant effect of ethnic diversity. However, he does not offer other explanatory factors for national identity. This role is taken by religious diversity in our approach. Miles and Rochefort (1991) also find in their survey study that religion is the most important factor of social identification. If the society is religiously sufficiently homogeneous people choose to identify with their religion because religiosity offers the narrowest set of common values and norms. Only if religious fragmentation is too strong the society looks for other objects of identification, which offers common values on a broader level. In this case people identify with their nationality. Apparently, national identity is a

substitute for religion if identification on a religious level is not possible due to social heterogeneity.

Concerning the other possible influential variables on national identity, we find further interesting results. The most robust and probably most important finding is that democratic institutions have a positive impact on national identity. Democracy gives people freedom they need in order to find their object of identification. A limitation to the liberal rights of the population reduces the closeness of the people to their country. If they are granted political rights and civil liberties people feel comfortable in their country and can identify with the values that are established by the society.

We also find support for the proposition that mobility throughout the country has an important influence on national identity. We estimate the separate impact of physical and non-physical mobility and find that both have a significant impact on national identity. However, non-physical mobility, which we proxied by the number of phone lines per 100 inhabitants, seems to be more important than physical mobility as it always enters highly significantly. Apparently the contact between individuals is important to form a national identity. This contact can be established through personal interaction for which physical mobility is needed. But it can also be established through non-physical mobility. Longer distances can be more easily covered by a phone call or an e-mail than by physical travel. Non-physical mobility saves time and enables people to be in contact with a lot of persons at very low costs.

Furthermore, country size and secondary education revealed significant results in part of the regressions. Secondary education might have a positive impact when the students are taught that they have a good government and should be proud of their country. It might also be the case that a better educated society can find more compromises on which similar values they can identify. In poorly educated societies fanaticism might spread more easily which could reduce the set of common beliefs. It has also been found before that the importance of religion decreases as people become better educated. Since national identity is a substitute for religion the decreasing role of religiosity could foster the importance of national identity. Country size might have a positive impact because a larger population might increase the probability that people discover a commonness to their neighbors with which they can identify.

A past under communist rule has a detrimental effect on national identity. Probably the disappointment about the negative consequences communism had for the population reduces the bonds the population has towards its nation. Surprisingly, we did not find a significant impact of income on national identity. Raising the level of economic development does not affect national identity. The concept of national identity seems to be above the level of material well-being. People can identify with the rest of the society due to shared values and beliefs. This is independent of the economic circumstances.

6 Conclusion

The present paper is a first attempt to make the concept of national identity numerically measurable. For this purpose we use information from the World Values Survey. The survey delivers information on peoples' attitudes concerning politics and the state itself. We use common factor analysis in order to receive one index of national identity. Our index consists of eight single indicators. In the combination we have more than 95,000 respondents to the different questions. We calculate our index on the country level which lets us work with 62 country observations for national identity.

The second main contribution of this study is to analyze the relationship between our new measure of national identity and variables measuring social heterogeneity. We use ethnic diversity, ethnic polarization, religious diversity, and religious polarization as proxy variables. Furthermore, we control for income, democratic institutions, geographical factors, education, openness, and mobility throughout the country. We conduct a cross country analysis. First, we use OLS estimation and due to endogeneity issues test the robustness of the results by using instrumental variables.

We find that only religious diversity has a direct significant effect on the level of national identity. Raising religious diversity by one standard deviation increases our index of national identity by more than half a standard deviation. Democratic institutions and mobility throughout the country are positively related to national identity, a past under communist rule has a negative impact. The other variables showed no clear pattern of significant effects.

We argue that people identify with the group that share the most common set of values and norms which is, in general, probably the religious community. If people cannot identify with their religious group because religious diversity is too high they choose another object of identification that offers common values on a broader level. Consequently, people identify with their nationality. Only if people cannot identify with their religion they choose their nationality as an object of identification.

We can conclude that religiosity and national identity indeed appear to be substitutes. If religious diversity is too high the importance of religion decreases. As a consequence, people look for another object of identification which offers a common set of values and norms which can be found in a national identity.

References

- Ainsworth, Martha, Deon Filmer (2006). Children's Schooling: AIDS, Orphanhood, Poverty, and Gender. *World Development* 34(6):1099-1128.
- Akerlof, George A., Rachel E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics* 115(3): 715-753

- Alesina, Alberto, Reza Baqir, William Easterly (1999). Public Goods and Ethnic Divisions. *Quarterly Journal of Economics* 114(4): 1243-1284
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, Romain Wacziarg (2003). Fractionalization. *Journal of Economic Growth* 8: 155-194
- Austen-Smith, David, Roland G. Fryer, Jr. (2005). An Economic Analysis of “Acting White”. *Quarterly Journal of Economics* 120 (2): 551-583
- Barrett, David B., George T. Kurian, Todd M. Johnson (2001). World Christian Encyclopedia. *Oxford University Press*
- Barro Robert J., Jong-Wha Lee (2010). A New Data Set of Educational Attainment in the World, 1950-2010. *NBER Working Paper, 15902*
- Battu, Harminder, McDonald Mwale, Yves Zenou (2007). Oppositional identities and the labor market. *Journal of Population Economics* 20: 643-667
- Bisin, Alberto, Eleonora Patacchini, Thierry Verdier, Yves Zenou (2010). Bend It Like Beckham: Ethnic Identity and Integration. *NBER Working Paper, 16465*
- Bodenhorn, Howard, Christopher S. Ruebeck (2003). The Economics of Identity and the Endogeneity of Race. *NBER Working Paper, 9962*
- Bollen, Kenneth A., Jennifer L. Glanville, Guy Stecklov (2002). Economic Status Proxies in Studies of Fertility In Developing Countries: Does the Measure matter?. *Population Studies* 56(1): 81-96.
- Bond, Ross (2006). Belonging and Becoming: National Identity and Exclusion. *Sociology* 40 (4):609-626
- Bruce, Steve (2000). The Supply-Side Model of Religion: The Nordic and Baltic States. *Journal for the Scientific Study of Religion* 39: 32-46
- Calhoun, Craig (1993). Nationalism and Ethnicity. *Annual Review of Sociology* 19: 211-239
- CIA (2009). The World Factbook 2009. *Central Intelligence Agency, Washington, DC*
- Collier, Paul, Anke Hoeffler (1998). On economic causes of civil war. *Oxford Economic Papers* 50: 563-573
- Constant, Amelie F., Klaus F. Zimmermann (2008). Measuring Ethnic Identity and its Impact on Economic Behavior. *Journal of the European Economic Association* 6(2-3): 424-433
- Constant, Amelie F., Liliya Gataulina, Klaus F. Zimmermann (2009). Ethnosizing immigrants. *Journal of Economic Behavior and Organization* 69: 274-287
- Darity, William A., Jr., Patrick L. Mason, James B Stewart (2006). The economics of identity: The origin and persistence of racial identity norms. *Journal of Economic Behavior and Organization* 60: 283-305
- Easterly, William, Ross Levine (1997). Africa’s Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics* 112(4): 1203-1250

- Esteban, Joan-María, Debraj Ray (1994). On the Measurement of Polarization. *Econometrica* 62(4): 819-851
- Fearon, James D., David D. Laitin (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97(1): 75-90
- Filmer, Deon, Lant H. Pritchett (2001), Estimating Wealth Effects without Expenditure Data - or Tears: An Application to Educational Enrollments in States of India, *Demography*, 38(1): 115-132.
- Filmer, Deon, Kinnon Scott (2008), Assessing Asset Indices, *World Bank Policy Research Working Paper No. 4605*, World bank.
- Fincher, Corey L., Randy Thornhill (2008). Assortative sociality, limited dispersal, infectious disease and the genesis of the global pattern of religious diversity. *Proceedings of the Royal Society B: Biological Sciences*
- Frankel, Jeffrey A., David Romer (1999). Does Trade Cause Growth? *American Economic Review* 89(3): 379-399
- Freedom House (2010). Freedom in the World.
- Gurr, Ted R., Monty G. Marshall, Keith Jagers (2011). Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010.
- Jones, F.L. (1997). Ethnic diversity and national identity. *Journal of Sociology* 33(3): 285-305
- La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, Robert Vishny (1999). The Quality of Government. *The Journal of Law, Economics and Organization* 15(1): 222-279
- Maddison, Angus (2010). Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD
- Mauro, Paolo (1995). Corruption and Growth. *Quarterly Journal of Economics* 110(3): 681-712
- Miles, William F.S., David A. Rochefort (1991). Nationalism Versus Ethnic Identity in Sub-Saharan Africa. *American Political Science Review* 85(2): 393-403
- Montalvo, Jose G., Marta Reynal-Querol (2005). Ethnic diversity and economic development. *Journal of Development Economics* 76: 293-323
- Montalvo, Jose G., Marta Reynal-Querol (2005a). Ethnic Polarization, Potential Conflict, and Civil Wars. *American Economic Review* 95(3): 796-816
- Nunn, Nathan, Diego Puga (2009). Ruggedness: The Blessing of Bad Geography in Africa. *NBER Working Paper, 14918*
- Opfinger, Matthias (2011). Religious Market Theory vs. Secularization: The Role of Religious Diversity Revisited. *University of Hannover Discussion Paper, 475*
- Paldam, Martin, Erich Gundlach (2009). The Religious Transition: A long-run perspective. *Kiel Working*

Paper, 1576, Kiel Institute for the World Economy

Sachs, Jeffrey D. (2001). Tropical Underdevelopment. *NBER Working Paper, 8119*

Sahn, David E., David Stifel (2003), Exploring Alternative Measures of Welfare in the Absence of Expenditure Data, *Review of Income and Wealth* 49(4): 463-489.

Sahn, David E., David Stifel (2000), Poverty Comparisons over Time and Across Countries in Africa, *World Development* 28(12):2123-2155.

Sastry, Narayan (2004). Trends in Socioeconomic Inequalities in Mortality in Developing Countries: The case of Child Survival in Sao Paulo, Brazil. *Demography* 41(3): 443-464.

Smith, Anthony D. (1991). National Identity. *University of Nevada Press, Reno*

Stifel, David and Luc Christiaensen (2007). Tracking Poverty Over Time in the Absence of Comparable Consumption Data. *The World Bank Economic Review* 21(2): 317-341.

Vanhanen, Tatu (1999). Domestic Ethnic Conflict and Ethnic Nepotism: A Comparative Analysis. *Journal of Peace Research* 36(1): 55-73

Voas, David, Alasdair Crockett, Daniel V.A. Olson (2002). Religious Pluralism and Participation: Why Previous Research Is Wrong. *American Sociological Review* 67(2): 212-230

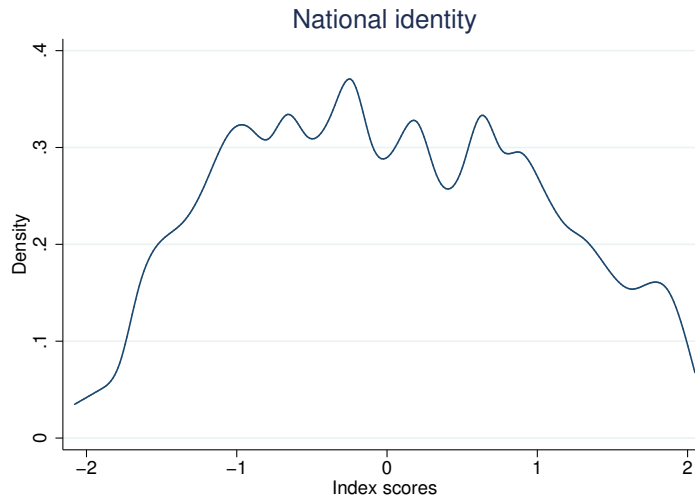
Appendix

Table A 1: Index values by country

country	Index value			country	Index value		
	Mean	SD	N		Mean	SD	N
Albania	-0.037	0.946	505	Latvia	-0.208	0.982	693
Azerbaijan	0.077	0.978	1521	Lithuania	-0.271	1.012	1036
Argentina	-0.515	0.924	803	Luxembourg	0.279	0.984	453
Australia	0.200	0.984	1718	Malta	-0.149	1.015	218
Austria	0.163	1.012	2023	Mexico	0.079	0.964	2543
Bangladesh	0.709	0.961	1022	Moldova	-0.331	0.966	809
Armenia	-0.035	1.018	1524	Netherlands	0.376	1.001	831
Belgium	-0.480	0.947	1849	New Zealand	0.202	0.933	1181
Bosnia and Herzegovina	0.412	1.033	925	Nigeria	-0.244	1.006	2243
Brazil	-0.118	0.920	2691	Norway	0.613	0.906	2146
Bulgaria	0.145	0.950	1235	Peru	-0.350	0.847	950
Belarus	-0.125	0.965	1966	Philippines	0.189	0.908	1098
Canada	0.258	0.964	1368	Poland	0.208	0.998	1974
Chile	-0.240	0.944	2116	Portugal	-0.403	0.898	855
Taiwan	-0.004	0.955	565	Puerto Rico	-0.181	0.974	1005
Croatia	-0.061	0.961	821	Romania	-0.284	0.905	776
Czech Republic	0.062	0.968	2427	Russia	-0.288	0.969	4389
Denmark	0.276	0.969	896	Slovakia	-0.072	0.960	1415
Dominican Republic	-0.323	0.880	288	Slovenia	-0.182	0.892	2245
Estonia	-0.079	0.983	608	South Africa	0.485	1.049	3876
Finland	-0.092	0.894	2125	Spain	-0.488	0.928	3297
France	-0.132	0.968	1862	Sweden	0.307	1.003	1644
Georgia	-0.065	1.008	1687	Switzerland	0.060	1.001	847
Germany	0.080	0.939	2627	Turkey	0.382	0.893	2614
Hungary	-0.082	0.963	1315	Ukraine	-0.253	0.979	2050
Iceland	0.125	0.919	602	Macedonia	-0.327	0.890	621
India	0.358	0.966	3032	United Kingdom	0.136	1.006	1219
Ireland	-0.096	0.941	858	United States	0.417	0.912	2620
Italy	-0.391	0.996	3702	Uruguay	-0.094	1.047	762
Japan	0.194	1.000	965	Venezuela	-0.458	0.828	975
South Korea	0.652	0.825	1165	Serbia	-0.294	0.980	1081

Source: World Value Survey; calculations by the authors.

Figure 1: Density of National Identity Index



Source: World Value Survey; calculations by the authors.