



Facoltà di Economia di Forlì
Laurea Magistrale in Economia Sociale



Working Papers

107

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

**Kenneth Harttgen and Matthias
Opfinger,**
ETH Zurich, Leibniz Universität Hannover

Giugno 2012

Info: AICCON - Tel. 0543.62327 - ecofo.aiccon@unibo.it - www.aiccon.it

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

Kenneth Harttgen, Matthias Opfinger*

ETH Zurich, Leibniz Universität Hannover

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: Religious Diversity, National Identity, Common Values

*We thank 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*.

Large empires have disintegrated into smaller units over the last centuries or in some cases even only in the last decades. New entities have been formed along specific lines which are today known as nation states. But how did these states evolve? What do the inhabitants have in common? The analysis of nations is popular in political and social sciences. It is investigated by, e.g. Anderson (2006), Alesina and Spolaore (2005), Bloom (1990), Gillis (1996), Miller (2000), Triandafyllidou (1998), Wodak et al. (1998), or concerning the role for specific nations or Europe by, e.g. Checkel (1999), Maier (1997), Noiriel (1996), or Smith (1992). Smith (1991) devotes a book to the question of national identity. As the reader proceeds throughout the book he comes across interesting hypotheses on the determinants of national identity. However, these can unfortunately not be tested empirically due to a missing measure of national identity. The present paper contributes to this literature by making national identity numerically measurable and comparable across countries. In a later stage of this study we use our newly constructed index to evaluate the impact of ethnic and religious 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.

In the following years several studies have been conducted which further investigate the role identity might play for economic outcomes. Bisin et al. (2010), for example, 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 racialist 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 to integrate into the white, presumably rich, network. Battu et al. (2007) come to a very similar conclusion when they investigate 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.

In this study we try to shed light on the relationship between national identity and ethnic and religious heterogeneity. Earlier studies followed similar approaches, however, they lack a clear concept of national identity, a problem which we try to solve with our index of national identity. 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 (2012) 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) all of which should be more easily obtainable in more homogeneous societies, as this facilitates the finding of compromises on the policies or legal codes. 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 measures of ethnic and religious diversity as well as polarization in order to 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. Ashraf and Galor (2007) claim that it is not cultural traits but the variation in cultural assimilation that affects economic development. In another paper, Ashraf and Galor (2011) suggest that genetic distance, as a consequence of human migration out of Africa, affects economic development.

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 ethnic and religious heterogeneity have 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 explain 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 between our index of national identity and social heterogeneity are presented in Section 3. Section 4 shows the empirical results, followed by some robustness tests in Section 5. The results are discussed in Section 6. Section 7 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. Table 3.1 suggests that using only one indicator might generate biased results. The proportion of the people who declare to be very proud of their country, for example, appears rather small. It might be that in an interview the respondents might hesitate to express overtly nationalistic attitudes. By combining a set of indicators we might be better able to actually capture the underlying factor of national identity. Second, within a unidimensional index, we are directly able to analyze the determinants of national identity.

To derive the national identity index, we apply a principal component analysis. The main idea of this approach is to construct an aggregated unidimensional 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 unidimensional index is widely used in the economic and social literature. We closely follow the approach of Filmer and Pritchett (2001) and Sahn and Stifel (2000, 2003) to construct an asset index of material welfare based on the possession of housing durables. The authors 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 Gundlach (2012) use an index approach to derive a measure of religiosity to analyze the religious transition over time.

Principal component analysis is an aggregation technique to identify from a set of variables those linear combinations that best captures 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 an aggregated index:

$$NI_j = b_1a_{j1} + b_2a_{j2} + \dots + b_ka_{jk} \tag{1}$$

where NI_j is the national identity index, the a_j 's refer to the respective variable of the person j 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 unidimensional index and that are to be estimated. In our model this means that the k^{th} identity variable, identified by a_{jk} is a linear function of a common factor, which in our case is "national identity".

For the estimation of the weights in equation (3.1) we use principal component analysis, which, has been used often in the recent empirical literature.¹ In particular, we rely on the first principal component

¹For example, a 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 indexes are used to analyze changes and determinants of poverty (e.g.

as our national identity index.²

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 B.1 in Appendix B shows the results by country for those countries where information on all variables that enter the index are available.

We include 8 dichotomous variables as components for the national identity index which are presented in Table 3.1. These are assumed to capture the national identity of a person. Our choice of variables entering the index depends on two factors: First, we use variables that define national identity capturing available information in the World Values Survey on political interest and participation and on national interest and orientations. The second reason is mainly due to data constraints. This means that we rely on this specific set of variables because it is identified as the set which suffers least from missing information.³ Table 3.1 shows the mean values of the indicators, the standard deviation, the number of observations, 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 3.1. In total, the first component explains 21% of the variance.

Table 1: Summary statistics for single items

Indicators	Index value	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.

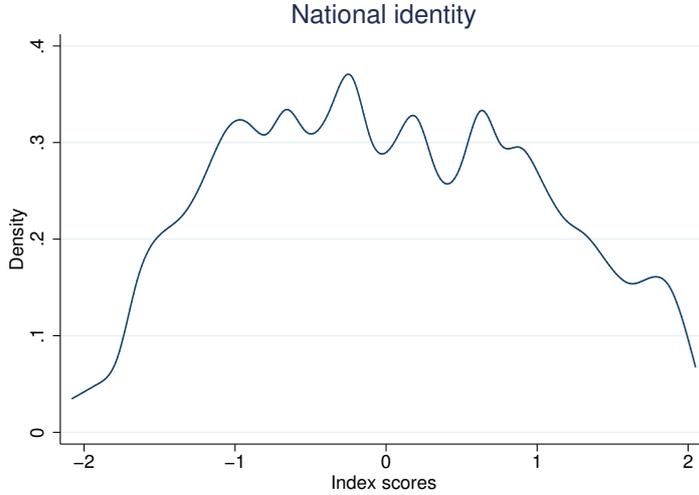
Note: Indicators of national identity and index statistics.

Stifel and Christiaensen 2007).

²An 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 Gundlach (2012). However, the two estimation methods show very similar results. For a systematic overview of different aggregation techniques, see Filmer and Scott (2008).

³We also tried to derive the index based on more variables. But since the 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 B.1.

Figure 1: Density of National Identity Index



Source: World Value Survey; calculations by the authors.

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. The indexes of ethnic and religious diversity measure the probability that two randomly drawn people belong to the same ethnic or religious group. It is calculated as $1 - H$, where H is a Herfindahl-Index which is gained by $\sum_i^N s_i^2$, where s_i is there share of people belonging to each ethnic or religious group i and N is the number of groups. Information on ethnic diversity comes from Alesina et al. (2003). 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). If everybody belongs to the same ethnic or religious group the index takes on the value zero and it equals one if everybody belongs to a different group.

Montalvo and Reynal-Querol (2005) calculate measures for ethnic and religious polarization. They

use these measures to analyze their respective effects on economic development. The indexes of ethnic and religious polarization are gained by $Pol = 1 - \sum_i^N \left(\frac{1/2 - s_i}{1/2} \right)^2 s_i$. 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. The term in parentheses then becomes zero and polarization reaches its maximum. Further increases in diversity reduce the index of polarization. If s_i is very small the term in brackets is close to one. It is then multiplied by the small s_i and summed over all i . Subtracting this term from one yields a small value for the polarization index.

In order to reduce the risk of reverse causality we want to use income from a year before the observation period on national identity begins. The Maddison (2010) online database offers information on income, also for the single former Soviet nations, for the year 1973. We include income to control for the possibility that economic development might have an impact on the formation of a national identity. Furthermore, it serves as a kind of broad catch-all control variable so that it possibly reduces the negative consequences of omitted variables. 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 and thereby test the arguments brought forward by Smith (1991). 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 rescale these two measures so that higher values signify more democratic institutions. 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.

Geography is also supposed to influence the formation of a 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).

It is reasonable to assume that openness might affect national identity as it can be seen as a proxy for tolerance towards others or the unknown. We use the trade share and constructed trade share from 1985 (Frankel and Romer, 1999) to control for it. Smith (1991) argues further 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 variable estimations we rely on data on the disease environment from Fincher and Thornhill (2008) and on climatic conditions which is taken from Sachs (2001). Fincher and Thornhill (2008) provide data on diseases and pathogen exposure in each country. From Sachs (2001) we take the data on the percentage of the population in each country living in temperate climatic areas.

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 heterogeneity, 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. This a societal phenomenon and we want to compare different societies and measure the impact of social heterogeneity on identity formation at the national level.

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:

$$NI_i = \alpha + \beta \cdot ethdiv_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i \quad (2)$$

$$NI_i = \alpha + \beta \cdot ethpol_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i \quad (3)$$

$$NI_i = \alpha + \beta \cdot reldiv_i + \gamma \cdot y_i + \delta X_i + \varepsilon_i \quad (4)$$

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

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 first stage F-statistic is 4.7 on average. The null hypothesis of underidentification can be rejected at the one percent level. The Cragg-Donald Wald statistic that is used to test for weak instruments is on average 11.1. The comparison with the Stock and Yogo (2005) critical values reveal that the disease environment is not a weak instrument. 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. Cervellati et al. (2011) find that the disease environment influences the possibility of civil conflicts. Following the argument of Montalvo and Reynal-Querol (2005a) and Fincher and Thornhill (2008), the disease environment leads to ethnic diversity which then results in a higher probability for civil conflict.

Ethnic polarization can be instrumented by the percentage of the population living in temperate climatic 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, as a consequence of modern migration, which leads to low levels of ethnic polarization. The first stage statistics imply that the instrument is valid, as the F-statistic is on average 7.1, the underidentification hypothesis can be rejected at the one percent level and the Cragg-Donald statistic is on average 18.4. 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 (F-statistic over 20 and Cragg-Donald statistic over 50). 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. But religious diversity in 1900 must also not have influenced the formation of nations. In Europe, the present nations have not been formed along religious borders. In Germany, for example, there are Catholics as well as Protestants and if diversity in 1900 would have played an important role in the formation of nations, there should have been a Protestant and a Catholic Germany. Although there certainly exist exceptions, such as former Yugoslavia, this should nevertheless be true for most countries. Furthermore, countries that were under colonial rule have not been shaped by religious borders but by the wishes of the ruling colonizers. As a consequence, religious diversity from the year 1900 appears to be 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 (F-statistic over 15 and Cragg-Donald statistic over 20). We can assume that the disease environment affects national identity only through religious and ethnic heterogeneity so that the exclusion restriction should not be violated.

3.3 Summary Statistics

Table 3.2 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.⁴ 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

⁴The 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.

Table 2: Summary Statistics on the country level

<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

country in our sample is South Korea with an index value of 0.002. Nigeria is the most ethnically diverse 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. Religious polarization in the Dominican Republic reaches an index value of 0.999.

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 % of the population completed secondary education. The highest value is observed in Armenia with 56.47 %. 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 % of the soil is fertile whereas the value for Malta is 100 %. 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 1,000 inhabitants. The highest value in this category is reported in Ireland with 25.95 kilometers per 1,000 inhabitants. In Brazil there are only 0.01 kilometers 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.

Our main goal is to study the relationship between national identity and ethnic and religious heterogeneity. Correlations of these raw data propose that ethnic diversity and ethnic polarization are negatively related to national identity (-0.27 and -0.34) whereas religious diversity and polarization are slightly positively related to national identity (0.19 and 0.12). In order to find out if this pattern holds when we introduce more explanatory variables we rely on regression analysis. The results will be presented in the next section.

4 Regression Results

In this section, we present the results of various regressions with which we want to analyze the effect of ethnic and religious 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 3.3 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 does income have 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 relation to national identity at the 5% 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.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Ethnic Diversity	-0.32* (-1.80)	-0.22 (-1.26)	-0.17 (-0.95)	-0.23 (-1.27)	-0.33 (-1.60)	-0.35* (-1.89)	-0.34* (-1.92)	-0.25 (-1.25)	-0.32* (-1.73)	-0.28 (-1.53)
Log of Income '73	-0.03 (-0.54)	-0.14** (-2.44)	-0.08 (-1.50)	-0.07 (-1.27)	-0.05 (-0.83)	-0.03 (-0.63)	0.01 (0.11)	0.00 (0.00)	-0.03 (-0.50)	-0.03 (-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 (0.75)	1.25** (2.56)	0.60 (1.29)	0.55 (1.16)	0.49 (0.95)	0.38 (0.82)	0.02 (0.05)	0.05 (0.08)	0.33 (0.71)	0.43 (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

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

Geographical factors that might influence the formation of a national identity are introduced in columns 6 through 11 which reveals that only population size has a positively significant impact at the 10% level. Thus, countries with a larger population reveal higher levels of national identity, the impact is very small as it differs from zero only in the fourth decimal. If the population is larger by 1 million inhabitants national identity is increased by only 0.0006 index points. 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 10% 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 10% 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.

At first sight we do not find evidence for the hypothesis that ethnic diversity is 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 correlation with national identity. The impact of population size is also slightly positive whereas a communist past is negatively related to national identity.

4.1.2 Instrumental Variable Results

Since the OLS regressions might suffer from endogeneity issues we rely on instrumental variable regression in order to gain more robust results on the effects of ethnic diversity on national identity.

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 3.4. The control variables remain in the same order as in Table 3.3.

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 is no important predictor of 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 1% level, the other two variables at the 5% level.

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

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

Population size and roads per inhabitant per square kilometer also stay positive and significant at the 10% level. Compared to Table 3.3 communist past loses statistical significance at the 10% 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 an important impact 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 heterogeneity.

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 3.3 and 3.4.

4.2.1 OLS Results

Table 3.5 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.

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

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

Population size and roads per inhabitant per square kilometer also stay positive and significant at the 10% level. Compared to Table 3.3 communist past loses statistical significance at the 10% 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.

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 2% 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 through 11 we add again the geographical variables and 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 3.6.

Table 6: Instrumental Variable regression results, 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

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

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, it no longer reveals significant effects. It fails to reach significance at the 10% level in all regressions.

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 1% level. In columns 3 and 4 the indexes of political rights and civil liberties are significant at the 5 and 10% 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 correlates with 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.

The OLS results suggest that ethnic polarization might have an important effect on the formation

of a national identity. The instrumental variable regressions reveal that the results were probably due to endogeneity, which might have been caused by omitted variables, and that ethnic polarization is not significantly related to 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 3.7.

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 10% level in only 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 1% 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 5% level.

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

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

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, which might arise due to omitted variables. If a variable is omitted that affects religious diversity as well as national identity, this might cause bias in the OLS estimates. We use rates of religious diversity in 1900 as instrument for religious diversity today. The results of these estimations are shown in Table 3.8.

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 all of the 19 estimations. It is significant at the 1% level in 17 regressions and at the 5% level in two more.

The coefficient is larger compared to the OLS results as 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 relation to the level of national identity. All three variables are significant at the 1% 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 reduce 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 3.7 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.

Table 8: Instrumental Variable regression results, dependent variable: national identity

	(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

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

The findings of this subsection, especially the instrumental variable results, support the idea that social heterogeneity is related to 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 reveal a robust positive relationship to national identity. A communist past appears to have a detrimental effect on our measure of national identity.

4.4 Religious Polarization

We use religious polarization as a last possible measure of social heterogeneity. 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

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

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

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 3.9.

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.

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 1% level, while the index of civil liberties is significant at the 5% 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 results of the instrumental variable regressions are presented in Table 3.10.

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.

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

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

Once more, we find democratic institutions to have a significant effect on national identity. The coefficient is positive and significant at the 1% level for the polity score and political rights and at the 5% level for civil liberties. The instrumental variable estimations reveal significance of some of the geographical variables. In column 7 population size is positively correlated with national identity which is significant at the 5% level. Also, the number of neighboring countries and the percentage of land that is fertile are significant at the 10% level. Both variables relate negatively to the level of national identity.

In column 13 we find, as already mentioned before, that a communist past decreases national identity,

which is significant at the 5% 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 is significantly correlated with 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 Section 6. 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 Robustness

In this section we show the results of some simple robustness tests. First, we will present regressions in which we included more than only one additional explanatory variable. Since some of our control variables measure basically the same we do not include all variables. The polity score and the index of political rights are very similar so that including both in the same regression might introduce multicollinearity problems. Second, we present the results of stepwise regressions in order to find out which explanatory variables are indeed important for the formation of a national identity.

Table B.2 shows the results of the regressions with multiple explanatory variables. The variables that are included in the regressions are listed on the left side. This exercise supports the main result concerning the relationship between national identity and ethnic and religious heterogeneity. Of all the variables we use as proxy variables for social heterogeneity only religious diversity has a significant impact on the formation of a national identity.

We also find phone lines to be significant in most estimations. This confirms the insight that nonphysical mobility seems to be an important determinant of national identity. Interestingly, income enters negatively

and significantly in all estimations. Apparently higher income has a detrimental effect on national identity. In contrast, democratic institutions, as proxied by the polity score, fails to reach statistical significance in these regressions. Since we also included colonial and communist past as explanatory variable, this result might possibly be due to correlation of the explanatory variables.

Table B.3 in Appendix B shows the results of stepwise regressions. The explanatory variables included at the beginning are the same we used for Table B.2. All variables of which the p-value lies above 0.15 are dropped from the model. We can see that our main result is again confirmed. Religious diversity has a positive impact on national identity. We also find that ethnic polarization has a negative effect as we expected. But our results section shows that this finding disappears once one relies on instrumental variable estimation which is not possible to apply in stepwise regressions.

Again, phone lines have a positive effect in all estimations and income enters negatively and significantly. Democratic institutions once more appear not to have a significant impact on national identity in this setting. When we use ethnic polarization as measure for social heterogeneity openness enters negatively, which is surprising and secondary education has a positive effect. All the other control variables are dropped.

6 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 from which we extract questions that refer to national identity. Due to data scarcity on several variables we are restricted to those eight items presented in Table 3.1 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 an average of 1,500 respondents per country.

In the empirical section, we investigate the relationship between national identity and ethnic and religious heterogeneity. In addition, we try to reveal other important determinants of 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. He concludes that people identify with their religious group if they cannot identify at the national level when the society is ethnically too fragmented.

We find that religious diversity has a strong positive relationship to national identity, whereas ethnic diversity does not enter significantly. 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 people choose to identify with their religious group. By sharing the same belief people send out signals that they also share a set of common values. Forming a common identity builds on this set of common values. Only if the society is religiously highly fragmented people do not identify with their religion. They start to doubt that their neighbors share the same set of common 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 common religious norms when people adhere to different denominations they choose to identify on another level with broader common values. 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. These persons will identify with their religious group as long as they adhere to the same denomination. They share a set of common values, which is based on their religious beliefs. These two persons 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 set of common 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 (2012) does not find a significant effect of ethnic diversity either. However, he does not offer other explanatory factors for national identity. We propose that religious diversity can explain that social heterogeneity does indeed affect the formation of a national identity. 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 determinants of national identity, we reveal further interesting insights. In our baseline regressions the most robust and probably most important finding is that democratic institutions have a positive impact on national identity. Democracy does not appear as significant explanatory variable in the robustness section. As this might be due to multicollinearity we still think that democratic institutions should be an important correlate of national identity. However, the way of causation is less clear. It could be that democracy offers people freedom they need in order to find their favored identity. A limitation to the liberal rights of the population reduces people's closeness to their country. If they are granted political rights and civil liberties the population feels comfortable in its country and can identify with the values that are established by the society. On the other hand, national identity could increase democracy. Only if people trust in the same norms based on their national identity they might be willing to engage in collective actions which make democracy work properly.

We also find support for the proposition that mobility throughout the country has an important influence on national identity. We estimate the separate impacts of physical and non-physical mobility and find that both are significantly related to 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 common 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. Better educated societies might possibly find more compromises on which similar values to 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 people's bonds towards their nation.

Surprisingly, we did not find a significant impact of income on national identity in many cases when we entered the explanatory variables separately. However, income is significantly and negatively related to national identity when we control for non-physical mobility and in the robustness section. These partially ambiguous results hint in a direction that the concept of national identity might be above the level of material wealth. People can identify with the rest of the society due to shared values and beliefs, independent of their economic situation. However, the negative effect might also be a sign for the decreasing importance of social networks to identify with in the most developed countries. Paldam and Gundlach (2012) show that religiosity loses importance as countries develop. Potentially the ties to the nation also vanish as societies become richer. The importance for networks with which people identify might decrease with growing prosperity, a fact which might be termed "rising individualism". This could explain why trade openness did not show significant results. If this is actually true has to be clarified by future research, as has to be the role that income inequality or income growth might play.

7 Conclusion

The present paper is a first attempt to make the concept of national identity measurable numerically. 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. Income might be negatively related to national identity which might possibly be a result of rising individualism. The other variables showed no clear pattern of significant effects.

We argue that people identify with the group that shares the narrowest set of common 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.

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 set of common values and norms which can be found in a national identity.

References

- Ainsworth M, Filmer D (2006) Children's Schooling: AIDS, Orphanhood, Poverty, and Gender. *World Development* 34(6):1099-1128.
- Akerlof GA, Kranton RE (2000) Economics and Identity. *Quarterly Journal of Economics* 115(3):715-753
- Alesina A, Baqir R, Easterly W (1999) Public Goods and Ethnic Divisions. *Quarterly Journal of Economics* 114(4):1243-1284
- Alesina A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R (2003) Fractionalization. *Journal of Economic Growth* 8(2):155-194
- Alesina, Alberto, Enrico Spolaore (2005). *The Size of Nations*. Cambridge, MA: MIT Press.
- Anderson, Benedict (2006). *Imagined Communities* (2nd ed.). London: Verso.
- Austen-Smith D, Fryer RG (2005) An Economic Analysis of 'Acting White'. *Quarterly Journal of Economics* 120 (2):551-583
- Barrett DB, Kurian GT, Johnson TM (2001) *World Christian Encyclopedia*. Oxford University Press, Oxford
- Barro RJ, Lee JW (2010) A New Data Set of Educational Attainment in the World, 1950-2010. NBER Working Paper, 15902
- Battu H, Mwale M, Zenou Y (2007) Oppositional identities and the labor market. *Journal of Population Economics* 20(3):643-667
- Bisin A, Patacchini E, Verdier T, Zenou Y (2010) Bend It Like Beckham: Ethnic Identity and Integration. NBER Working Paper, 16465
- Bloom, William (1990). Personal Identity, National Identity and International Relations. *British Journal of Political Science* 20, 1-10.
- Bodenhorn H, Ruebeck CS (2003) The Economics of Identity and the Endogeneity of Race. NBER Working Paper, 9962
- Bollen KA, Glanville JL, Stecklov G (2002) Economic Status Proxies in Studies of Fertility In Developing Countries: Does the Measure matter?. *Population Studies* 56(1):81-96
- Bond R (2006) Belonging and Becoming: National Identity and Exclusion. *Sociology* 40(4):609-626
- Bruce S (2000) The Supply-Side Model of Religion: The Nordic and Baltic States. *Journal for the Scientific Study of Religion* 39(1):32-46
- Calhoun C (1993) Nationalism and Ethnicity. *Annual Review of Sociology* 19(1):211-239
- Checkel, Jeffrey T. (1999). Norms, Institutions, and National Identity in Contemporary Europe. *International Studies Quarterly* 43(1), 84-114.

- CIA (2009) The World Factbook 2009. Central Intelligence Agency, Washington, DC
- Collier P, Hoeffler A (1998). On economic causes of civil war. *Oxford Economic Papers* 50(4):563-573
- Constant AF, Gataulina L, Zimmermann KF (2009) Ethnosizing immigrants. *Journal of Economic Behavior and Organization* 69(3):274-287
- Constant AF, Zimmermann KF (2008) Measuring Ethnic Identity and its Impact on Economic Behavior. *Journal of the European Economic Association* 6(2-3):424-433
- Darity WA, Mason PL, Stewart JB (2006) The economics of identity: The origin and persistence of racial identity norms. *Journal of Economic Behavior and Organization* 60(3):283-305
- Easterly W, Levine R (1997) Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics* 112(4):1203-1250
- Esteban JM, Ray D (1994) On the Measurement of Polarization. *Econometrica* 62(4):819-851
- Fearon JD, Laitin DD (2003) Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97(1):75-90
- Filmer D, Pritchett LH (2001) Estimating Wealth Effects without Expenditure Data - or Tears: An Application to Educational Enrollments in States of India. *Demography* 38(1):115-132
- Filmer D, Scott K (2008) Assessing Asset Indices. World Bank Policy Research Working Paper No. 4605
- Fincher CL, Thornhill R (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 JA, Romer D (1999) Does Trade Cause Growth? *American Economic Review* 89(3):379-399
- Freedom House (2010). *Freedom in the World*.
- Gillis, John R. (Ed.) (1994). *Commemorations: the politics of national identity. Princeton: Princeton University Press.*
- Gurr TR, Marshall MG, Jaggers K (2011) Polity IV Project: Political Regime Characteristics and Transitions, 1800-2010.
- Jones FL (1997) Ethnic diversity and national identity. *Journal of Sociology* 33(3):285-305
- La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny R (1999) The Quality of Government. *The Journal of Law, Economics and Organization* 15(1):222-279
- Maddison A (2010) *Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD*
- Maier, Charles S. (1997). *The unmasterable past: history, holocaust and German national identity. Cambridge, MA: Harvard University Press.*
- Masella P (2011) National Identity and Ethnic Diversity. *Journal of Population Economics*, forthcoming
- Miles WFS, Rochefort DA (1991) Nationalism Versus Ethnic Identity in Sub-Saharan Africa. *American*

Political Science Review 85(2):393-403

Miller, David (2000). Citizenship and national identity. *Malden, MA: Polity Press.*

Montalvo JG, Reynal-Querol M (2005) Ethnic diversity and economic development. *Journal of Development Economics* 76(2):293-323

Montalvo JG, Reynal-Querol M (2005a). Ethnic Polarization, Potential Conflict, and Civil Wars. *American Economic Review* 95(3):796-816

Noiriel, Gérard (1996). The French melting pot: Immigration, citizenship, and national identity. *Minneapolis: University of Minnesota Press.*

Nunn N, Puga D (2009) Ruggedness: The Blessing of Bad Geography in Africa. NBER Working Paper, 14918

Opfinger M (2011) Religious Market Theory vs. Secularization: The Role of Religious Diversity Revisited. University of Hannover Discussion Paper, 475

Paldam M, Gundlach E (2009) The Religious Transition: A long-run perspective. Kiel Working Paper, 1576

Sachs JD (2001) Tropical Underdevelopment. NBER Working Paper, 8119

Sahn DE, Stifel D (2001) Poverty Comparisons over Time and Across Countries in Africa. *World Development* 28(12):2123-2155

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

Sastry N (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 AD (1991) National Identity. University of Nevada Press, Reno

Smith, Anthony D. (1992). National identity and the idea of European unity. *International Affairs* 68(1), 55-76.

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

Stock JH, Yogo M (2002) Testing for Weak Instruments in Linear IV Regression. NBER Working Paper, 10284

Triandafyllidou, Anna (1998). National identity and the 'other'. *Ethnic and Racial Studies* 21(4), 593-612.

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

Voas D, Crockett A, Olson DVA (2002) Religious Pluralism and Participation: Why Previous Research Is

Wrong. *American Sociological Review* 67(2):212-230

Wodak, Ruth, Rudolf de Cillia, Martin Reisigl, Karin Leibold (1998). Zur diskursiven Konstruktion nationaler Identität. *Frankfurt/Main: Suhrkamp*.