

The role of ethnic diversity and education in determining national identity and political behaviour

Paolo Masella

London School of Economics and Political Science

Thesis submitted in the fulfilment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in Economics

December 2007

Except where otherwise stated, this thesis is the result of my own work

UMI Number: U501509

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U501509

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

THESES
F
8796



1130966

Abstract

The process of individual identity formation is still an enigma, as is the capacity of public bodies to intervene in it. This thesis is the first to take a step in this direction.

Using individual data from the World Value Survey, the second chapter presents several findings on the relationship between national sentiment and ethnic diversity. Contrary to conventional wisdom, we find no evidence of a lower intensity of national sentiment in more ethnically fragmented countries or in minority groups. National feelings in a minority can be higher or lower than in a majority, depending on the degree of ethnic diversity of a country. On the one hand, in countries with high ethnic polarization, minorities have weaker national sentiments than majorities; on the other hand, in countries with low ethnic polarization, the reverse is true. We then develop a model of national identity formation that is consistent with the facts presented in the empirical section.

As a second step, using survey data from Catalonia, we estimate the impact on identity of the 1983 educational reform, by which the education system became bilingual, and Catalan, together with Spanish, was taught in schools. Using within and between cohort variation in exposure to the Catalan language at school, the results show that individuals who have experienced greater exposure to teaching in Catalan are more likely to say that they feel more Catalan than Spanish.

In the subsequent chapter, we study the effect of the Catalan reform on political behaviour. We find that the change in the educational system stimulated turnout and changed the political choices of the agents.

Acknowledgements

First of all I would like to thank my supervisor, Prof. Maitreesh Ghatak, for his academic and technical advice but also for his encouragement.

The help of my coauthors/friends, Oriol and Irma, has been very important too.

During my Ph.D. I had the chance to present my work in numerous conferences and seminars; I would like to thank everyone who gave me comments and suggestions on my work. I would also like to thank my colleagues at LSE, for their very valuable comments and suggestions.

Finally (but maybe more importantly), I would like to thank my family and all my friends.

Contents

Abstract	2
Acknowledgements	3
List of Tables	8
List of Figures	11
1 Introduction	13
2 National Identity and Ethnic Diversity: Theory and Cross-Country Evidence	17
2.1 Introduction	17
2.2 Empirical evidence	21
2.2.1 Data	21
2.2.2 Empirical strategy and results	24
2.2.3 Measurement issues and other concerns	26
2.2.4 Summary of the empirical findings	27
2.3 The model	27
2.3.1 Preferences	28
2.3.2 Political arena and voting equilibrium	30

2.3.3	Identity choice	33
2.3.4	National sentiment and ethnic diversity	35
2.4	Discussion: a model with n groups	37
2.5	Conclusions	39
2.6	Appendix A	40
2.6.1	Proof of Proposition 2.1	40
2.6.2	Proof of Lemma 2.1	42
2.6.3	Proof of Lemma 2.2	43
2.6.4	Proof of Lemma 2.3	43
2.6.5	Proof of Proposition 2.2	43
2.6.6	Proof of Proposition 2.3	45
2.6.7	Proof of Lemma 2.4	45
2.6.8	Proof of Proposition 2.5	46
2.7	Appendix B	47
2.7.1	Definition of variables	47
2.7.2	Countries and ethnic groups	48
2.8	Figures and Tables	50
3	Identity Formation and Language Policies: the Catalan Case	64
3.1	Introduction	64
3.2	Background, data and empirical strategy	66
3.2.1	Background	66
3.2.2	Data and descriptive statistics	67
3.2.3	Empirical strategy	70

3.3	Results	71
3.3.1	Difference-in-Difference Estimates	71
3.3.2	Effect for each level of treatment	72
3.3.3	Channels	73
3.3.4	Migration patterns	74
3.3.5	Restricting the sample	75
3.3.6	Exogenous Trends	76
3.4	An Alternative Identification Strategy	78
3.5	Heterogeneous effects	80
3.6	Summary and conclusions	81
3.7	Appendix	81
3.7.1	Definition of the variables	81
3.8	Figures and Tables	82
4	Education and Political behaviour: Evidence from the Catalan Linguistic Reform	104
4.1	Introduction	104
4.2	Background	106
4.2.1	Institutional Setting	106
4.2.2	Data Description	106
4.3	Empirical strategy	108
4.4	Results	109
4.4.1	Difference-in-Difference Estimates	109
4.5	Heterogenous effects	111
4.6	Robustness	112

4.6.1	Exogenous Trends	112
4.6.2	Endogeneity of years of education	114
4.6.3	Restricting the sample	115
4.6.4	Migration patterns	116
4.7	Conclusions	117
4.8	Appendix	117
4.8.1	Definition of the variables	117
4.9	Figures and Tables	118
5	Conclusion	134
	Bibliography	135

List of Tables

Table 2.1	National identity by country54
Table 2.2	Descriptive Statistics.55
Table 2.3	Sources of national identification: individual characteristics56
Table 2.4	Sources of national identification: country characteristics (1)57
Table 2.5	Sources of national identification: country characteristics (2)58
Table 2.6	Sources of national identification: country characteristics (3)59
Table 2.7	Sources of national identification: country characteristics (4)60
Table 2.8	Sources of national identification: size of the group -only majority groups-	61
Table 2.9	Sources of national identification: size of the group -only minorities- (1)	62
Table 2.10	Sources of national identification: size of the group -only minorities- (2)	63
Table 3.1	Average years of education by type of family84
Table 3.2	Catalan at school by type of family85
Table 3.3	Descriptive Statistics.86
Table 3.4	Reform and Identity91
Table 3.5	Effects for each level of treatment93
Table 3.6	Channels	...95
Table 3.7	Other robustness checks (1)96
Table 3.8	Other robustness checks (2)97
Table 3.9	Controlling for exogenous trends99
Table 3.10	First stage	...100
Table 3.11	Controlling for endogeneity of years of education	...101
Table 3.12	The role of individual origin	...103
Table 4.1	Descriptive Statistics	..123
Table 4.2	Reform and Turnout	..124
Table 4.3	Reform and Ethnic Voting (1)	..125
Table 4.4	Reform and Ethnic Voting (2)	..126
Table 4.5	Reform and voting patterns	..127

Table 4.6	The role of individual origin	..128
Table 4.7	Robustness checks (1)	..130
Table 4.8	Robustness checks (2)	..131
Table 4.9	Robustness checks (3)	..132

List of Figures

Figure 2.1	National Identity and Ethnic Polarization51
Figure 2.2	National Identity and Ethnic Fractionalization51
Figure 2.3	Ethnic Polarization vs Ethnic Fractionalization52
Figure 2.4	Median Voter53
Figure 2.5	Intensity as a function of size of group A53
Figure 3.1	Treatment by Cohort and Education83
Figure 3.2	Catalan at school83
Figure 3.3	Identity of Catalans with Secondary Education by Year of Birth87
Figure 3.4	Identity of Catalans without Secondary Education by Year of Birth87
Figure 3.5	The effect of Education on Identity88
Figure 3.6	Fraction of respondents who feel only Catalan by level of education and year of birth89
Figure 3.7	Fraction of respondents who feel more Catalan than Spanish by level of education and year of birth89
Figure 3.8	Fraction of respondents who feel as Catalan as Spanish by level of education and year of birth89
Figure 3.9	Fraction of respondents who feel more Spanish than Catalan by level of education and year of birth89
Figure 3.10	Fraction of respondents who feel only Spanish by level of education and year of birth89
Figure 3.11	The effect of Education on feeling only Catalan by year of birth90
Figure 3.12	The effect of Education on feeling more Catalan than Spanish by year of birth90
Figure 3.13	The effect of Education on feeling as Catalan as Spanish by year of birth90
Figure 3.14	The effect of Education on feeling more Spanish than Catalan by year of birth90
Figure 3.15	The effect of Education on feeling only Spanish90

	by year of birth	
Figure 3.16	Change in probabilities (specification 3.1)	...92
Figure 3.17	Effect for each level of treatment (specification 3.2)	...94
Figure 3.18	Migration	...98
Figure 3.19	Level of education by year of birth	...102
Figure 4.1	Participation rate by year of birth and level of education	..119
Figure 4.2	Effect of education on participation rate by year of birth	..120
Figure 4.3	Ethnic voting by year of birth and level of education	..121
Figure 4.4	Effect of education on ethnic voting by year of birth	..122
Figure 4.5	Probability of voting by intensity and origin	..129
Figure 4.6	Probability of ethnic voting (among voters)	..129
	by intensity and origin	
Figure 4.7	Probability of ethnic voting (among all)	..130
	by intensity and origin	

Chapter 1

Introduction

Identity, beliefs and social norms are at the centre of many social and economic phenomena. The outcome of an organization is likely to be affected by the degree of identification of its employees with their tasks within the organization. The identification of the teachers with the mission of the school can improve the quality of education within that school and, at the same time, the effort of the students is likely to be function of the gap between the students' self image and the person that the school intends them to be. The choice of ethnic minorities to adopt oppositional identities may affect the level of cooperation within a society and generate social and economic conflicts. Ethnic riots are recurrent events in many countries and ethnic conflicts are a widespread phenomenon across the world. The effectiveness of policies and their political viability may depend on their impact on individual identity.

Only recently has the economic literature come to focus on individual identity. In an influential series of papers, Akerlof and Kranton (2000, 2002, 2005) and Akerlof (2007) incorporated identity, a person's sense of self, into an economic model of behaviour. They proposed a utility function that depends on the individual's assigned social category, on the match between the prescriptions for that category and the individual's given characteristics, and on his and others' actions. They then presented several applications of their theory in the fields of the Economics of Education, the Economics of Organizations and Macroeconomics. Previous literature on economics mentioned and studied the concept of identity (Sen (1985), Folbre (1994), Kevane (1994) and Landa (1994)), but never considered it as an argument in the utility function. A few theoretical studies followed the work of Akerlof and Kranton; among others, one by Benabou and Tirole (2007) which endogenized identity payoffs and categorical prescriptions. However, empirical research into both the determinants of individual identity and how identity affects individual behaviour remains limited.

This thesis focuses on two particular types of identity, national and ethnic identity, and contains some theoretical work (Chapter Two) and some empirical work (Chapters Two, Three and Four). The concepts of nation, nationalism and national sentiment have captured the attention and fed the creativity of many political scientists, politicians, philosophers and writers, including Armstrong, Breully, Fishman, Hroch, Smith, Stalin and Tilly. Benedict Anderson (1983) provided the following suggestive definition: "*nation is an imagined political community and imagined as both inherently limited and sovereign*". More recently Gellner (1983) wrote: "*nationalism is a theory of political legitimacy, which requires that ethnic boundaries should not cut across political boundaries*". To the best of our knowledge, no definitions, nor answers, are widely accepted. However, the panoply of books, articles and reviews that flourished in the last century testify to the importance of this topic.

Moreover, in the economic literature the study of how ethnic diversity can affect economic outcomes has attracted a lot of attention¹. A vast amount of economic literature has investigated the relationship between ethnic diversity and economic development. Mauro (1995) claims that ethno-linguistic diversity has a direct negative effect on the level of investment. Easterly and Levine (1997) find that a high level of ethnic fragmentation has a negative impact on economic growth. La Porta et al. (1999) show that ethnic heterogeneity contributes to the explanation of the quality of government. More recently, Montalvo and Reynal (2005) suggest that ethnic (and religious) polarization is one of the factors explaining economic development through its impact on the rate of investment, the proportion of government consumption over GDP and the probability of civil wars.

However, there is still a very small body of literature that explores the sources of ethnic identification. Using the Afrobarometer surveys in nine Sub-Saharan African democracies, Miguel and Posner (2006) estimate the likelihood that an individual identifies himself in ethnic terms rather than in class or religious ones. They found, surprisingly, that the salience of ethnicity is negatively related to ethnic diversity, and claim that exposure to competition for jobs and political power are factors that predispose individuals to identify themselves in ethnic terms. Caselli and Coleman (2002) present a model of ethnic conflict in which ethnic identity is endogenous.

Chapters Two and Three analyze the determinants of national and ethnic identity. The second chapter examines the relationship between ethnic diversity and national sentiment. Using individual data from the World Value Survey, the second chapter presents several findings on the relationship between national sentiment and ethnic diversity. Contrary to conventional wisdom, we find no evidence of a lower intensity

¹Alesina and La Ferrara (2005) extensively studied and surveyed the relationship between ethnic diversity and economic performance.

of national sentiment in more ethnically fragmented countries or in minority groups. National feelings in a minority can be higher or lower than in a majority, depending on the degree of ethnic diversity of a country. On the one hand, in countries with high ethnic polarization, minorities have weaker national sentiments than majorities; on the other hand, in countries with low ethnic polarization, the reverse is true. We then develop a model of national identity formation that is consistent with the facts presented in the empirical section.

Limited research has examined the extent to which identity can be shifted by cultural policies and regulation. As the UN report 2004 pointed out, language regulation does exert an influence on the social environment, which encourages identity formation (and, since core individual identity is typically formed in the early ages of life, language at school turns out to be particularly important). Chapter Three takes a step forward in this direction by analyzing the effect of a language policy on the process of identity formation. Using survey data from Catalonia, it estimates the impact on identity of the 1983 educational reform whereby the education system became bilingual, and Catalan, together with Spanish, was taught in schools. Using within and between cohort variation in exposure to the Catalan language at school, the results show that individuals who have experienced greater exposure to teaching in Catalan are more likely to say that they feel more Catalan than Spanish. Interestingly, this effect appears to be present also among individuals whose parents do not have Catalan origins.

In ethnically diverse countries, politicians and political scientists have focused on very different policies as mechanisms for minority integration and conflict reduction (see Horowitz 1985 for a review of this topic). Nation building policies (for example, the implementation of a unique national language across the entire territory of a state) as well as multicultural policies that explicitly recognize cultural differences (regional autonomy, devolution of powers and the use of multiple languages in school and in other contexts) have often been proposed as sources of conflict management. Miguel (2003) shows that nation-building has allowed diverse communities in Tanzania to achieve better economic outcomes than diverse communities in the Kenyan region. The linguistic law we study can be interpreted as an example of a multicultural policy in Spain, where individuals living in Catalonia are the relevant minority, and, at the same time, as an example of a nation building policy in Catalonia, where individuals living in Catalonia who were born elsewhere or whose parents do not have Catalan origins are the relevant minority. As a result, nation building policies and multicultural policies deliver very different outcomes in terms of the effects on individual sentiments: multicultural policies tend to favour the development of regional identities while nation building policies tend to promote the expansion of a common national feeling.

Chapter Four focuses on the relationship between schooling and political behaviour.

We explore how the linguistic reform that took place in Catalonia affected turnout and party choice. Using the survey data described in the third chapter, we find that individuals who have experienced a greater exposure to teaching in Catalan are more likely to declare that they voted in the 1998 regional election and to have chosen a party with a Catalanist (i.e Catalan regionalist) platform. The introduction of a bilingual education system increased the salience of the ethnic issue in the Catalan society and helped the consolidation of a political system organized along ethnic lines. Identity is a likely channel through which the linguistic reform affected political choices.

Chapter 2

National Identity and Ethnic Diversity: Theory and Cross-Country Evidence

2.1 Introduction

Ethnic diversity has a negative impact on economic development and political stability. A high level of ethnic fractionalization implies a lower level of investment and worse institutional quality; a high level of ethnic polarization increases the probability of civil conflict. "Nation building" (policies that promote attachment to nation over ethnic and regional identities) has been often proposed as a mechanism for integration and conflict reduction. This chapter studies the determinants of national sentiment and, in particular, examines the relationship between ethnic diversity and national sentiment.

A vast amount of economic literature has investigated the relationship between ethnic diversity and economic development. Mauro (1995) claims that ethno-linguistic diversity has a direct negative effect on the level of investment. Easterly and Levine (1997) find that a high level of ethnic fragmentation has a negative impact on economic growth. La Porta et al. (1999) show that ethnic heterogeneity contributes to the explanation of the quality of government. More recently, Montalvo and Reynal (2005) suggest that ethnic (and religious) polarization is one of the factors explaining economic development through its impact on the rate of investment, the proportion of government consumption over GDP and the probability of civil wars.

However, there is still only a very small body of literature which explores the sources

of ethnic identification. Using the Afrobarometer surveys in nine Sub-Saharan African democracies, Miguel and Posner (2006) estimate the likelihood that an individual identifies himself in ethnic terms rather than in terms of class or religion. They found, surprisingly, that the salience of ethnicity is negatively related to ethnic diversity and claim that exposure to competition for jobs and political power are factors that predispose individuals to identify themselves in ethnic terms. Caselli and Coleman (2002) present a model of ethnic conflict in which ethnic identity is endogenous.

"Nation building" (policies that promote attachment to nation over ethnic and regional identities) has been proposed as a mechanism for integration and conflict reduction by a large number of scholars. They¹ have mostly focused on policies (particularly educational policies) that are able to change individual preferences and provoke a massive shift of loyalties from the ascriptive group to the State. Miguel (2003) documents how nation building reforms were implemented in the newly independent East African nations in the 1960s and 1970s. He focuses on the economic development of two countries: Tanzania and Kenya. Despite their similar colonial institutional legacy, ethnic make up and geographical conditions, leaders in Kenya and Tanzania have adopted very different policies, especially with regards to ethnic groups, over a wide range of dimensions. Nyerere (Tanzania) followed a "Pan-Africanist" nation building policy, along with a centralized economic policy. Miguel shows that nation-building has allowed diverse communities in Tanzania to achieve better economic outcomes than diverse communities in the Kenyan regions.

The aim of this work is to investigate the main determinants of national sentiment and, more importantly, the relationship between ethnic diversity and the intensity of national feeling. Figures 2.1 and 2.2 cast some doubts on the conventional views. They plot ethnic diversity (in Figure 2.1 we use the index of ethnic polarization, in Figure 2.2 that of ethnic fractionalization²) against the level of national identification of each country³. The two graphs suggest the absence of a negative relationship between ethnic heterogeneity and national sentiment. The chapter has an empirical side that draws on survey data from the last three waves of the World Value Survey and a theoretical side, that relies on a recent contribution by Bisin et al. (2006) and that is consistent with the facts presented in the empirical part.

We use data from the World Value survey, a multi-country survey project that employs

¹Among others, Deutch and Foltz (1963), Coleman and Rosberg (1964), Jacob and Toscano (1964) and Tilly (1975).

²The ethnic fractionalization index can be interpreted as measuring the probability that two randomly selected individuals in a country belong to different ethnic groups. On the other hand, the purpose of the ethnic polarization index is to capture how far the distribution of the ethnic groups is from a bipolar distribution, which represents the highest level of polarization.

³The country level of national sentiments is measured as the proportion of individuals who choose to identify with their nation rather than their ethnic group.

standardized questionnaires to investigate citizens' attitudes in a vast set of countries. We estimate the likelihood that an individual identifies himself in national terms rather than in ethnic ones. Contrary to conventional wisdom, we find no evidence of lower intensity of national sentiment in more ethnically fragmented countries or in minority groups. However, we find that national feelings in a minority are higher (lower) than in a majority depending on the ethnic diversity of a country. On the one hand, in countries that are more polarized, minorities have a lower level of national sentiment than majorities, on the other hand, in less polarized countries, the reverse is true. The empirical section provides two further findings. Larger groups tend to have weaker national feelings and individuals with higher income are less inclined to feel associated with their ethnic group.

As illustrative examples, let us now consider the cases of Macedonia, China and Brazil. China is the country with the lowest level of ethnic polarization in the sample, while Brazil has the highest. In China there is a very large ethnic group, the Chinese Han (around 92% of the population) and several other smaller groups (Zhuang Chinese, Hui Chinese, Uyghur Chinese, Miao Chinese and Manchu Chinese). The proportion of Chinese Han in the sample who identify themselves with China rather than with the Han nationality is around 1%, while the proportion of individuals belonging to other ethnic groups who identify themselves with China rather than with their respective nationalities is far higher (more than 50%). On the contrary, if we consider Brazil we have a very different picture. 55% of the Brazilian population is white. There is a consistent portion of mulatto (above 30%) and quite a few black individuals (above 5%). On the one hand, half of the white people consider themselves just Brazilian rather than white Brazilian; on the other hand, among the individuals belonging to the ethnic minorities, only 37% consider themselves as just Brazilian. Hence, in a country with low ethnic polarization such as China, national sentiments are weaker in the majority than in the minorities, while in a country that is very polarized, like Brazil, national sentiments are weaker among minorities than within the largest group.

The Republic of Macedonia is a country with a level of ethnic diversity that is close to the sample average. Macedonians are obviously the largest group, while Albanians are the largest minority (more than 20%) and there is quite a substantial fraction of Turks (4%) and Serbs (2%). The fraction of Albanians who identify themselves with their country is around 5%; the fraction of Turks and Serbs is around 40% and 60%, respectively. The larger the size of an ethnic minority, the lower seems to be the average intensity of its national sentiment.

We then develop a theoretical model that is consistent with the facts presented in the empirical section. In each ethnic group there is an accepted social norm; deviating from that social norm involves some costs. We assume that these costs are higher the

more widespread the norm in the population. In our model, in order to overcome the psychological shocks due to the distance from their own ethnic group, each individual decides to ask asylum in that "*imagined community*" we call nation⁴ and then chooses to develop national feelings. In other words, national sentiments represent a psychological need to abandon the ethnic convention. In the model, the ethnic composition of the population has two effects on the intensity of national sentiment: a direct effect and an indirect effect, through the political choices of the politicians in power.

We believe that the political oppression of a minority ultimately affects its feelings towards the nation. We decided to model political oppression in the simplest possible way: allowing inter-ethnic redistribution. The degree of integration of individuals from ethnic minorities is going to be weakened by a set of economic policies that constantly discriminate against them. We present a simplified political setting where, in ethnically polarized societies (with a large minority), the (small) ethnic majority in power chooses to implement a strong policy of inter-ethnic redistribution because the benefits of the tax revenue are going to be divided among a small fraction of people. This is consistent with the findings that ethnic conflicts are more likely when society consists of two groups that are fairly similar in size and also with a recent work by Posner (2004). He finds that cultural cleavages become more salient when the size of the ethnic minorities is sufficiently large.

On the other hand, size also has a direct effect. The cost function we assume implies a direct and negative effect of size on the intensity of national feelings. Individuals tend to deviate more from norms set by smaller ethnic groups. This is in harmony with the assumptions used in the construction of the ethnic polarization index (and more in general with all the classes of polarization indices) where identification with one's own (ethnic) group is supposed to be a positive function of its size⁵.

We then compare the level of national sentiment within the largest ethnic group with the level of national sentiment within the ethnic minority. The main results arise from the interaction of the political side with the "direct" *size effect* mentioned previously. In a two-group framework, when ethnic diversity is high, the "direct" *size effect* plays no role since the sizes of the minority and majority are similar. There is, however, a large inter-ethnic redistribution and, as a result, on average, the intensity of national sentiment is higher in individuals from the majority than in individuals from the minority. On the other hand, when there is a very large majority group and ethnic heterogeneity is low, there is a large difference between the size of the majority group and that of the minority. The *size effect* is the prevailing one and national feelings could be even higher

⁴ "*nation is an imagined political community and imagined as both inherently limited and sovereign*" Anderson (1983).

⁵ See Esteban and Ray (1994) and Montalvo and Reynal (2005).

within the minority than within the majority group.

This chapter is closely related to several strands in the literature. On the theoretical side there is an increasing amount of literature on endogenous identity; let us mention, among others, Akerlof and Kranton (2000), Bisin and Verdier (2000) and Bisin et al (2006). The model in the theoretical section builds on these contributions and, in particular, on that of Bisin et al (2006). According to their contribution, the social norm is the behaviour of the majority. Minorities have to bear some costs because of the lack of assimilation. A strong ethnic identity reduces those costs. This chapter focuses more on national identity formation and considers the identity choice of all the ethnic groups, while Bisin and al. (2006) focus only on the ethnic assimilation of minorities. Moreover, with respect to their framework, we introduce a political economy aspect and we make exercises of comparative statics also at a more aggregate level (country level and ethnic group level). These crucial ingredients allow national sentiments in a minority to be stronger (weaker) than in a majority depending on the ethnic diversity of a country.

On the empirical side the study by Miguel and Posner (2005) is the one that is most closely related to this chapter. They study the determinants of ethnic identity and how the salience of ethnicity is related to ethnic diversity. This chapter focuses more on national identity and moreover attempts to study how the intensity of national sentiments may differ depending on the characteristics of the ethnic group to which the individual belongs.

The rest of the chapter is organized as follows. Section 2.2 describes the data and presents some empirical evidence. Section 2.3 analyzes the model. Section 2.4 discusses the 2 group assumption of the model. Section 2.5 provides some concluding remarks. All the proofs are in the Appendix.

2.2 Empirical evidence

2.2.1 Data

The data in this study comes from the World Values Surveys (WVS) waves two (1990-1993), three (1995-1997) and four (1998-2000). This is a multi-country survey project that employs standardized questionnaires to investigate citizens' attitudes in a vast set of countries. In particular, one crucial question for our purposes is the following (here, we provide the US example):

"Which of the following best describes you? 1 above all, I am an Hispanic American 2

above all, I am a Black American 3 above all, I am a White American 4 above all, I am an Asian American 5 I am an American first and a member of some ethnic group second"

We build a variable, "*national identity*", which is equal to 1 if individuals answer "*I am an American first and a member of some ethnic group second*", 0 otherwise. It is constructed from the responses to the previous question by 32644 individuals in 25 separate World Value Survey rounds conducted in 22 countries: the United States, Canada, Spain, China, South Africa, Brazil, Byelorussia, Latvia, Uruguay, Armenia, Azerbaijan, Macedonia, Georgia, Albania, Bosnia, Indonesia, Israel, Moldova, Pakistan, Singapore, Zimbabwe and Jordan.

The World Value Survey was conducted via face to face interviews during which interviewers coded the respondents' ethnic characteristics by their observations. We create a variable that we call "*minority*"; this variable is equal to 0 if the ethnic group to which the individual belongs is the largest in the country, 1 otherwise. We also include controls for individual income (each respondent has been asked to choose among 10 income categories, net of transfers and taxes), individual education (we have "*primary*", a dummy variable equal to one if the agent has, at most, primary education) and individual occupation. During the survey, individuals were asked what was their main occupation. They could choose between 13 options. The omitted category is "agricultural workers".

We also include as explanatory variables characteristics of the ethnic groups and country variables. The dataset by Fearon (2003) provides the size of the ethnic groups in the sample. As a proxy of the average income of the group, we use sample averages. As a measure of the ethnic diversity of a country we use two different indices: the index of ethnic fractionalization and that of ethnic polarization. The former is quite common in the literature, while the index of ethnic polarization has been used only more recently.

A fractionalization index is defined as

$$Frac = 1 - \sum_{i=1}^N q_i^2$$

This indicator can be interpreted as measuring the probability that two randomly selected individuals in a country will belong to different ethnic groups. Therefore this index increases when the number of groups increases.

We also use the index **RQ** proposed by Reynal-Queirol

$$RQ = 1 - \sum_i^N \left(\frac{0.5 - q_i}{0.5} \right)^2 \pi_i = 4 \sum_{i \neq j}^N q_i^2 q_j$$

where q_i is the size of the ethnic group i . The purpose of this index is to capture how far the distribution of the ethnic groups is from a bipolar distribution, which represents the highest level of polarization. This index fulfills the two basic properties that a proper index of polarization needs to satisfy:

- 1) If there are 3 groups of sizes, p , q , and r , and $p > q \geq r$, then if we merge the two smallest groups into a new group, the new distribution is more polarized than the original one.
- 2) If there are 3 groups of sizes, p , q , and p , then if we shift population mass from the q group equally to the other two groups, the polarization index increases.

The index of ethnic polarization and that of ethnic fractionalization are perfectly correlated when the number of groups is two, but differ when the number of groups increases. Montalvo and Reynal (2005) showed that, for low levels of fractionalization, the relationship between the two indices is positive and close to linear. For medium ranges of fractionalization, however, it is zero, while for high levels of fractionalization it becomes negative. Figure 2.3 shows the correlation between the index of ethnic polarization and the one of ethnic fractionalization using only the countries included in the sample.

We should mention that we use two weight variables provided by the survey. One is a national weight, which reflects a country's national distribution of key variables such as the urban-rural divide, education, demographics and economic activity, while the second variable assigns the same weight to every country in order to avoid a large sample bias in the pooled country study.

Table 2.1 reports the share of respondents who identify themselves with their country for each of the twenty-five surveys in our sample. Uruguay has the highest share of respondents who identify themselves with their country (about 79 per cent) and Albania the lowest (about 1.5 per cent). Table 2.2 gives descriptive statistics of the main variables used in the analysis.

2.2.2 Empirical strategy and results

2.2.2.1 Individual sources of national identification

As a very first step, we begin by regressing the variable "national identity" on a set of individual characteristics plus country-year dummies. Standard errors are always clustered at the country-time level. Column 1 of Table 2.3 reports our minimal specification. Respondents with higher income are more likely to choose nation over ethnic group. Being part of a minority group does not affect individual identity. In Column 2 we introduce the variable "primary". Its coefficient is negative and significant. This suggests that low educated people (with at most primary education) tend to be less likely to be nationally identified than ethnically identified. The coefficient, however, does not appear very large. An individual with more than a primary degree is (almost) three per cent more likely to identify himself with the nation than an individual with, at most, a primary degree. In Column 3, we add variables that control for individual occupation. The omitted category is "agricultural workers". It turns out that individual identity varies quite strongly with occupation. Individuals belonging to categories such as "professionals" (teachers, lawyer...), "armed forces", "farmer" and to some extent "foreman" are more likely to be identified with the nation than "agricultural workers". In Column 4, we include all the possible individual characteristics. The coefficient of "minority" is never significant, while the coefficient of income is always positive and significant.

2.2.2.2 National identity and ethnic diversity

As a further step, we check if there is any relationship between identification and ethnic diversity. We regress identity on the usual set of individual controls and on ethnic diversity (in Table 2.4, we use the index of ethnic polarization, in Table 2.5 that of ethnic fractionalization). We find no evidence of any negative relationship between ethnic diversity and national identity. The tables show weak evidence in favour of an inverted U shaped relationship between ethnic diversity and national identity. The size of the sample, however, is limited and the countries included in our study may not be a good substitute for a world-wide sample.

We turn now to the main specification. In the previous section, we found that the coefficient of the variable minority is not significant. Respondents from minority groups do not tend to be more or less identified with their nation than respondents from majority groups. In this section, we study whether the feelings of respondents from minority groups depend somehow on the degree of ethnic diversity of the country in

which they live. We introduce an interaction term between the variable "minority" and a country variable that measures ethnic diversity.

The specification of the pooled cross-country analysis is given by:

$$\text{national identity}_{i,c,t} = \alpha_{c,t} + \text{income}_{i,c,t}\beta + \text{minority}_{i,c,t}\gamma + \text{minority}_{i,c,t} * \text{ethnic diversity}_c\delta + \varepsilon_{i,c,t}$$

There are country-time fixed effects, so it cannot be the case that omitted country characteristics, correlated with ethnic polarization, affect individual identity. However, there can be omitted country-variables having differential impacts on individual identity and correlated with right hand side variables (ethnic diversity in this case).

In Table 2.6, we use ethnic polarization as a measure of ethnic diversity, in Table 2.7, ethnic fractionalization.

Let us describe briefly Table 2.6. The variable minority has a positive coefficient while the sign of the coefficient of the interaction term is negative. Both are significant. Let us now discuss the magnitude of the main result. In countries with average ethnic polarization, there is no difference between the identity choice of the minority and that of the majority group. However if we increase the ethnic polarization index by one standard deviation, we find that individuals from ethnic minority groups are 8 per cent less likely to have national identity than those from the majority group, while if we decrease the ethnic polarization index by one standard deviation, we find that individuals from ethnic minority groups are 8 per cent more likely to have national identity than those from the majority group.

In Table 2.7, we perform the same exercise using ethnic fractionalization as a measure of ethnic diversity. This exercise is less successful. The signs of the coefficients are the same as in the previous Table, but they are not significant.

2.2.2.3 Size of the ethnic group

In a second set of regressions we study the relationship between national identity and the size of the ethnic group to which the individual belongs. We divide the sample in two subsamples: a subsample with all the individuals belonging to a minority group and a subsample with all the individuals belonging to the majority group.

We regress identity on a set of individual characteristics and on the size of the group. In several regressions, we also include the sample mean income for each ethnic group.

Obviously, we can include country fixed effects only when we use the subsample that includes agents from minority groups. The results are shown in Tables 2.8, 2.9 and 2.10.

The coefficient of the variable measuring the size of the ethnic groups to which the respondent belongs is always negative and significant. As in the previous subsection, no causality can be claimed; omitted group characteristics (and also country characteristics when country fixed effects are not included), correlated with the size of the group, may affect individual identity.

2.2.3 Measurement issues and other concerns

Several limitations of the analysis are worthy of mention. Firstly, the respondents' answers are context specific. We try to control for this by using what we know about the context in which the individuals were being surveyed (particularly where and when the survey was conducted). However, clearly, there are aspects for which we cannot control, such as the proximity of ethnic festivals.

Secondly, there can be measurement errors, which are particularly important in attitude surveys. Respondents in societies in which there are social norms that prevent any open talk about ethnicity may be less likely to declare ethnic identification. There would be an upward bias in measured national identity in that society. However the survey was conducted by private enumerators who were not affiliated to any political party or government. So, it is likely that the survey has not been perceived as related to any national institution. More importantly, the question analyzed was only one of 220 questions asked in the standard World Values Survey, the only one which clearly mentions ethnicity. It is fair to claim that ethnicity is only a marginal issue with which the survey deals. In addition, country fixed effects can partially address this problem.

Thirdly, it is not clear if the countries included in our study may be a good substitute for a world-wide sample. Our findings will have to be interpreted with the caveat that they may not be representative of the whole world.

We find that income is positively correlated with the likelihood that an individual identifies himself in national terms rather than ethnic ones. The size of the group is negatively correlated. Omitted variables correlated with size or income may affect national identity. On the other hand, some causality may be claimed in the case of the main specification when we use country fixed effects. However, there can be omitted country-variables having differential impacts on individual identity and correlated with right hand side variables (ethnic diversity in this case).

2.2.4 Summary of the empirical findings

- Respondents with higher income are more likely to choose nation over ethnic group
- There is no evidence of any negative relationship between ethnic diversity and national identity
- On the one hand, in countries with low ethnic diversity, individuals from minorities have stronger national sentiments than individuals from majorities; on the other hand, in countries with high ethnic diversity, on average, individuals from minorities have weaker national sentiments than individuals from majorities. This result, however, is robust only if we use ethnic polarization as an index of ethnic diversity.
- The likelihood that an individual identifies himself in national terms rather than in ethnic ones is negatively related with the size of the group the individual belongs to

2.3 The model

Consider a society populated by a continuum of citizens of mass 1. Citizens belong to one of two ethnic groups, A or B. The size of group A is q_A . Each individual i has an initial exogenous income stream y_i . Income distribution $F(y)$ is uniform on $[y - \varphi, y + \varphi]$ and, for simplicity, we assume that this is independent of ethnicity.

Each agent decides the intensity of his national identity. We suppose identity to be a personal choice; it is not inherited through the family. The proposed framework builds on that provided recently by Bisin et al (2006)⁶.

In each ethnic group, there is an accepted social norm, deviating from which involves some costs. Individuals develop strong national feelings as a mechanism for reducing such psychological costs. The intensity of national sentiment depends on the individual's disposable income and on the size of the ethnic group to which he belongs. As a result the ethnic composition of the population has two effects on the individual's choice: a direct one, through the size of the group, and an indirect one, through the political outcome and its impact on the disposable income of each agent.

The disposable income is determined after the voting stage. We use a standard Downsian model of electoral competition. The policy space consists of transfers to the whole

⁶In Bisin et al (2006), the social norm is the behaviour of the majority. Minorities have to bear some costs because of the lack of assimilation. A strong ethnic identity reduces those costs.

population or only targeted towards a specific ethnic group. Transfers are financed by a standard proportional income tax. The voting equilibrium depends on the degree of ethnic diversity of the society.

The sequence of events is as follows: first there is a voting stage -politicians are elected; once policies are implemented, each individual decides the intensity of his national identity.

The section is organized as follows: as a first step we describe the individual preferences, then introduce the policy space and study the voting equilibrium (and the individual disposable income). Finally we consider the identity choice and study how it depends on the ethnic composition of the population.

2.3.1 Preferences

Preferences are summarized by the utility function $u(x, y_{i,z}^d)$ where $y_{i,z}^d$ is the disposable income of an individual i belonging to group z , while x is an abstract argument that collects the whole set of choices of the individual i, z . As usual, the utility function is a positive function of disposable income. We postulate that there exists a social norm within each ethnic group that prescribes a behaviour $x = x_0$. Ethnic norms often prescribe individual behaviour in a vast array of activities. Members of an ethnic group might be expected to respect a specific dress code (the Islamic veil, for instance) or a particular diet (the Jewish dietary laws are one of the many examples), as discussed by Epstein (2006). Norms can involve the use of the local language at home (Lazear, 1999), influence the choice of the partner or interfere with the gender division of labour and more in general with many gender relationships within the family (Amott and Matthei, 1996). Individuals are likely to be asked to invest their time in ritual and festivals (Kuran, 1998) and, in the extreme case, to participate in episodes of violence, like during the Rwandan genocide (as documented by Bhavnani, 2006).

According to Kuran (1998), individuals follow ethnic norms in order to retain social acceptance. Norms might be sustained by economic sanctions, reputational concerns or simply psychological need to accommodate social pressure. In our model deviating from such norms entails some psychological costs. This approach is consistent with a broad class of conformity models (see, among others, Akerlof, 1980; Jones, 1984; Akerlof, 1997; Ballester, Calvo Armengol, and Zenou, 2005; Battu, Mwale, and Zenou, 2005; Bernheim, 1994; Glaeser and Scheinkman, 2001; Kandel and Lazear, 1992; Fershtman and Weiss, 1998) where failing to conform to the own group norms involves some costs. In Akerlof (1980) persons who disobey the code of honour of a community sacrifice "reputation", in Kandel and Lazear (1992) sentiments like shame and

guilt can affect individuals who, in large organizations, work less than the group norm. Berhheim (1997) and Fershstmam and Weiss (1998) suppose that, in addition to standard preferences, individual also cares about esteem and social status, while Akerlof (1997) adopts a utility function that explicitly includes a conformist component: the agents lose utility from failing to conform to others.

We denote such psychological costs with $\frac{\lambda(\alpha, q_z)(x-x_0)^2}{2}$ where q_z is the size of group z . They are assumed to be higher the more widespread the norm in the population. Similar assumptions are adopted by Bisin et al. (2006) and Akerlof (1980)⁷. Kandel and Lazear (1992) discuss the relationship between firm size and peer pressure; they suggest that the level of monitoring can increase with the size of the firm and also that, if more workers observe an individual, the sanctions imposed might be greater. α is the intensity of national feeling and, as mentioned earlier, it is supposed to alleviate such psychological costs⁸.

$$\begin{aligned}\lambda_{q_z} &= \frac{\partial \lambda(\alpha, q_z)}{\partial q_z} > 0 \\ \lambda_\alpha &= \frac{\partial \lambda(\alpha, q_z)}{\partial \alpha} < 0\end{aligned}$$

Building a national identity, however, involves some costs too. We adopt standard convex costs $\frac{\alpha^2}{2}$.

As a result, the entire problem can be summarized in the following way: each individual has to choose a set of actions x and the intensity of national identity α such that $U_{i,z}$ is maximized, where

$$U_{i,z} = u(x, y_{i,z}^d) - \frac{\lambda(\alpha, q_z)(x-x_0)^2}{2} - \frac{\alpha^2}{2}$$

⁷In Akerlof (1980), the individual reputation depends on his obedience of the code of behaviour of the community. The larger the number of believers in that code, the larger is the individual loss of reputation from disobeying it. However, the setting is dynamic and the fraction of believers in the next generation is a negative function of the fraction of individuals who, in the current generation, disobey the code.

⁸As in Bisin et al (2006), we make further assumptions in order to make the maximization problem well defined and the exercise interesting enough. We assume $u_{xx} < 0$ and without any loss of generality $x_i(y_{i,z}^d) = \arg \max u(x, y_{i,z}^d) > x_0$ for all $y_{i,z}^d$. Moreover we assume that if $\alpha \rightarrow 0$, then $\lambda \rightarrow \infty$, while to guarantee the concavity of the problem $\lambda_{\alpha\alpha} > 0$.

2.3.2 Political arena and voting equilibrium

We first describe the policy space. There are two types of transfer: transfers (Tr) to the whole population and transfers towards a particular ethnic group z (Tr_z) that are thus enjoyed only by a fraction q_z of the population.⁹ Ethnicity allows the policy maker to target transfers towards a specific fraction of the population.¹⁰ Transfers are financed by proportional income taxation.¹¹ Taxation, however, is assumed to be distortionary in the sense that it wastes resources of $C(\tau) = \frac{\gamma\tau^2}{2}$. The cost function is increasing and convex in τ .¹² As a consequence the policy maker faces the following budget constraint:

$$\tau y = Tr + Tr_{AQA} + Tr_{BQB} + \frac{\gamma\tau^2}{2}$$

where y is the total (and average) income in the country.

We use a standard model of Downsian electoral competition. There is a 2 party electoral competition, political candidates are opportunistic and, more precisely, their only motivation is to hold office. Moreover, we assume that candidates can commit to a well defined policy ahead of the elections. So, as a result, the timing is as follows: the two candidates announce a policy platform, so as to maximize the probability of victory, then voters select their preferred policy; eventually the policy announced by the winning candidate gets implemented. In this political setting, the policy implemented coincides with the Condorcet winner (if any). Despite the multidimensional policy space and thanks to the assumptions about the income distribution, it is possible to show the existence of a policy π^M capable of beating any alternative policy π in a pairwise voting.

The median voter belongs to the largest group (we consider group A as being the largest) and he is the individual, with income y^M , such that the size of the group of agents, belonging to group A, with lower income stream than him, is one half: $F(y^M, z = A) = \frac{1}{2}$ (as illustrated by Figure 2.4).

⁹Alternatively we can think in terms of public goods. So we could have had an environment that is more similar to the one provided by Fernandez and Levy (2005). The policy maker can provide two types of public good: one is a standard public good from which the whole society derives utility, the other is a "targeted" public good. The fact that those goods can be targeted may result from preference differences (a public good can be useful only for one particular ethnic group) or from geographical differences across groups (locally provided public good, if an ethnic good is concentrated in a particular area of the country).

¹⁰It is close to the Caselli and Coleman's (2002) view. In their work, the role of ethnicity was to enforce coalition membership.

¹¹Taxation cannot be ethnically targeted. This is crucial to guarantee the existence of a Condorcet winner.

¹²As usual the assumption of distortionary tax allows us to get an interior solution.

As a result, the higher is the degree of ethnic heterogeneity the richer is the median voter. Fernandez and Levy (2005) and Roemer (1998) obtained similar results using different political settings. Fernandez and Levy (2005) show that, under certain conditions, when ethnic diversity increases the set of equilibrium policies involves lower pro-poor policies; Roemer (1998) shows that the existence of the ethnic issue (if sufficiently salient) can work against the interests of the poor. This is consistent with empirical findings in the literature. Easterly and Levine (1997) and Alesina, Glaeser and Sacerdote (2001) provide cross country evidence of a strong negative correlation between ethnic (and racial) fragmentation and public goods provision (mainly education and infrastructure). Alesina, Baqir and Easterly (2000) finds a similar relationship across US States.

Proposition 2.1 *There exists a unique voting equilibrium π^M . The tax rate is $\tau^M = \frac{(1-q_A)(y-\varphi)}{\gamma}$ and all the tax revenue is devoted to transfers to group A. $Tr_A^M = \left(\tau^M y - \frac{\gamma(\tau^M)^2}{2} \right) \frac{1}{q_A}$*

Proof: In Appendix

The median voter is an individual who belongs to the largest ethnic group with an initial income stream higher than the average income stream, $y^M > y$. As a direct consequence, in a world with proportional income taxation, he does not finance transfers (Tr) to the whole population with a positive amount of tax rate. On the other hand he accepts that $\tau^M > 0$ in order to finance a positive amount of transfers directed only towards individuals from his ethnic group (Tr_A)¹³. Then the total tax revenue, $\tau^M y$, net of the distortions $C(\tau^M)$, is redistributed among a fraction of agents q_A . The median voter chooses the tax rate that maximizes his disposable income, $(1 - \tau^M) y^M + \frac{\tau^M y - C(\tau^M)}{q_A}$. Since the benefits of the tax revenue are divided among fewer people, the tax rate is higher the lower the size of the majority group.

The model seems to be consistent with casual and formal evidence provided by previous literature. A governance outcome that is commonly associated with severe divided societies is patronage. Patronage refers to the situation of extended benefits to members of an ethnic group (the one in power) while discriminating members of other ethnic groups. Usual examples (see Padro, 2006) of groups advantaged by political leaders are Northern groups in both Nigeria and Uganda or Tutsis in Burundi. More formal evidence is provided by Alesina, Baqir and Easterly (1998): they show that public employment is significantly higher in US cities where ethnic fragmentation is higher.

¹³There is no other policy able to beat π^M in a pairwise voting. π^M uses only transfers towards group A. This allows politicians to channel a high amount of resources towards a specific group of the population without using a high level of taxation and so without wasting a huge amount of resources (taxes are distortionary).

They interpret public employment as an example of subsidy to ethnically defined interest groups. An other outcome of the model is the positive relationship between tax rate and level of ethnic diversity. Alesina, Baqir and Easterly (1999) finds that per capita local taxes are positively associated with ethnic fractionalization in US cities (but not in US counties and in US metro areas). However, using a different specification, Berry (2002) shows that per capita property taxes are positively correlated with ethnic fractionalization in US counties as well. Moreover, La Porta et al (1998) find that more ethnically heterogenous countries have an higher top marginal tax rate.

From Proposition 2.1 we easily derive that:

$$y_{i,A}^d = \left(1 - \frac{(1 - q_A)(y - \varphi)}{\gamma}\right) y_{i,A} + \frac{\frac{(1 - q_A)(y - \varphi)}{\gamma} y - \frac{\gamma \left[\frac{(1 - q_A)(y - \varphi)}{\gamma}\right]^2}{2}}{q_A}$$

$$y_{i,B}^d = \left(1 - \frac{(1 - q)(y - \varphi)}{\gamma}\right) y_{i,B}$$

Lemma 2.1 $y_{i,B}^d$ is increasing in q_A for every i, B ; while $y_{i,A}^d$ is increasing (decreasing) in q_A if $y_{i,A}$ is higher (lower) than a given threshold

Proof: In Appendix

The intuition is straightforward: the higher the size of the largest group, the lower the tax rate and the lower the transfers to group A. So, the higher the size of A (and the lower the size of their own group), the higher the disposable income of each agent from group B. An increase in q_A , however, has 2 opposite effects on the disposable income of an agent from group A: on the one hand, it decreases the tax rate, on the other hand it decreases the transfers too. Since taxes are proportional to income, it is easy to conclude that the first effect dominates the second only if the initial income stream is sufficiently high.

In this simple political setting, inter-ethnic redistribution is the only form of redistribution that can arise as equilibrium outcome. The existence of a Condorcet winner, however, is guaranteed by the assumptions on the income distribution: $F(y)$ is assumed to be uniform and independent of ethnicity. Once we remove one of these 2 assumptions, π^M might be beaten by a policy that allows for both inter-ethnic redistribution and general redistribution ($TR > 0$) and that is supported by individuals with a low level of income from both the ethnic groups. Although the structure of the political side is very simple and does not allow for standard redistribution in equilibrium, it

delivers results consistent with more complex theoretical frameworks which, however, can be less easily integrated with the other part of our model, the identity decision. Fernandez and Levy (2005) study a framework with a similar policy space but with a different structure of political competition. They study the equilibrium of a game in which political parties endogenously arise. Political parties are restricted to offering platforms that belong to the Pareto set of their members. In equilibrium, parties of different kind can arise, they can represent economic interests but also ethnic interests. They find, consistently with the results delivered by our political setting, that when ethnic diversity increases the set of equilibrium policies involves lower pro-poor policies and larger ethnic transfers. However, in Fernandez and Levy (2005), under a large set of parameters, there are multiple equilibria and an equilibrium with maximum general redistributive taxation coexists with equilibria in which the party in power implements a mix of general redistribution and inter-ethnic redistribution. Esteban and Ray (2007) provided an other setting that allows for both class and ethnic coalitions to emerge. They study a non democratic setting in which peace and conflict are the two possible outcomes. They find that, under broad parametric conditions, class conflict is never the equilibrium outcome and ethnic conflict or peace are the only possible equilibria. In their framework, ethnic salience is the product of economic inequality: while class coalitions display strong within group economic homogeneity, in ethnic coalitions there is strong economic heterogeneity which allows important synergies. Rich components of the ethnic coalitions provide economic resources, poor components conflict labor.

2.3.3 Identity choice

As previously stated, the individual problem is:

$$\max_{\alpha; x} u(x, y_{i,z}^d) - \frac{\lambda(\alpha, q_z)(x - x_0)^2}{2} - \frac{\alpha^2}{2}$$

It is straightforward to see that, as a result of the maximization problem, we get $\alpha_{i,z}^*(q_z; y_i)$; feelings are a function of the size of the ethnic group to which the individual belongs and of his initial income stream. We study the signs of $\frac{\partial \alpha_{i,z}^*(q_z; y_i)}{\partial q_z}$ and $\frac{\partial \alpha_{i,z}^*(q_z; y_i)}{\partial y_i}$ and we get results analogous to Bisin et al. (2006).

Lemma 2.2 $\frac{\partial \alpha_{i,z}^*}{\partial y_{i,z}}$ and $\frac{\partial \alpha_{i,z}^*}{\partial y_{i,z}^d}$ are positive if $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$

Proof: In Appendix

If $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ individuals from group z with a higher disposable income (and so also with higher initial income stream) tend to get higher utility from actions that are more distant from the social norm of their group and, as a result, they choose a higher intensity of national identity. The lower the income, the higher is the utility gained from respecting dress and language codes, gender roles imposed by the group or from giving up work time to participate to (not necessarily violent) ethnic activities. In other words, wearing specific clothes, not working on the Sabbath, praying five times a day, doing most of the housework (gender division of labour) or being part of a violent faction in a conflict environment, all have an opportunity cost. The condition on the cross derivative states that those opportunity costs are increasing with income.

Casual evidence seems to be consistent with Lemma 2.2. According to Jeffres (1983), income is negatively correlated with ethnic behaviour. There is also some evidence that the use of regional languages (Grin, 1999 and Padilla, 2004) and endogamy (McCaa, 1993 and Wildsmith et al., 2007) are associated with lower socio-economic status. Moreover, very poor individuals in Sierra Leone (Humphreys and Weinstein, 2004) and landless individuals in Rwanda (Verwimp, 2005) were more likely to participate in episodes of violence.

We are now looking for the sign of $\frac{\partial \alpha_{i,z}^*(q_z; y_{i,z})}{\partial q_z}$. We need to take into account several effects.

- a direct effect of q_z on the costs of deviating from the social norm established within each group. We assumed that $\frac{\partial \lambda(\alpha, q_z)}{\partial q_z} > 0$; so the higher the q_z , the higher are the costs that the agent has to bear, the smaller the deviation from the ethnic norm undertaken by him, then the lower the intensity of national sentiment he needs to alleviate such costs. As a result, there is a negative direct effect.
- there is also an indirect effect: the effect of a change in q_z on $\frac{\partial \lambda(\alpha, q_z)}{\partial \alpha}$, which is the marginal effect of α on the unit cost of deviating from the social norm. Consider a positive sign, $\lambda_{\alpha q} > 0$. This implies that the larger the size of the ethnic group, the smaller, in absolute value, the marginal effect of α on the unit cost. $\lambda_{\alpha q} > 0$ implies a negative indirect effect.
- moreover, we need to consider the effect of an increase in q_z on the disposable income of each individual. The size of the largest group affects the tax rate and so the disposable income of the population. In a 2 group model, with population normalized to 1, an increase in q_A coincides with a decrease in q_B and viceversa. Consider the case $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$. From Lemma 2.2, the higher the disposable income, the stronger the intensity of the national feeling. From Lemma 2.1, an increase in q_A leads to an increase in the disposable income of the (very) rich

people in A and to a decrease of the disposable income of the poor people in A. As a result, we can conclude that there is a negative effect for the poor agents from A and a positive one for the rich ones. On the other hand, an increase in q_B (and consequently a decrease in q_A) leads to a decrease in the disposable income of every agent from B. Then, we can conclude that there is a negative effect for all the agent from B. Consider now the case $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} < 0$. Following similar reasoning, we can conclude that there is a positive effect for all the agents from B and the poor agents from A, while there is a negative effect for the rich agents in A.

As a result, when $\lambda_{\alpha q}$ is positive the first and second effects we mentioned are negative, while the third effect is positive for a fraction of the population and negative for the remaining part. If $\lambda_{\alpha q}$ is sufficiently large, then the sum of all the three effects will be negative for every agent in the society.¹⁴

Lemma 2.3 *If $\lambda_{\alpha q} > 0$ is positive and sufficiently large, then $\frac{\partial \alpha_{i,z}^*}{\partial q_z} < 0$*

Proof: In Appendix

2.3.4 National sentiment and ethnic diversity

In a 2 ethnic group environment, the lower the size of the largest ethnic group, the higher the degree of ethnic diversity.¹⁵ As an initial exercise, we compare the average intensity of the national feeling of the majority (group A) with the average intensity of national feeling of the minority (group B). We call α^A the average intensity of national feelings of group A ($\alpha^A = \int_{y-\varphi}^{y+\varphi} \alpha_{i,A}^*(y_i; q_A) dF(y_i|z = A)$) and α^B the average intensity of national feeling of group B ($\alpha^B = \int_{y-\varphi}^{y+\varphi} \alpha_{i,B}^*(y_i; q_B) dF(y_i|z = B)$).

Proposition 2.2 *Let $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ and $\lambda_{\alpha q} > \bar{\lambda}$, then $\alpha^A > \alpha^B$ if $q_A < k$; $\alpha^A < \alpha^B$ otherwise*

¹⁴We could get similar results even if $\frac{\partial \lambda(\alpha, q_z)}{\partial q_z} > 0$. In this case the first effect would be positive. However, if the second effect is sufficiently large, then it can dominate the other 2 effects.

¹⁵The index of ethnic fractionalization is the most common index of ethnic diversity. More recently a new index was introduced in the literature: the index of ethnic polarization. The purpose of the ethnic polarization index is to capture how far the distribution of the ethnic groups is from a bipolar distribution, which represents the highest level of polarization. In a 2 group context, however, it turns out that they both provide the same ranking. The higher the size of the majority, the lower they are.

Proof: In Appendix¹⁶

In order to study the relationship between ethnic diversity and the difference between α^A and α^B , we need to take into account 2 conflicting effects: a direct effect (the *size effect*) and an indirect effect (the *political effect*).

On the one hand, from Lemma 2.3, we find that when $\lambda_{\alpha q} > \bar{\lambda}$, the intensity of national identity chosen by each agent is a negative function of the size of the group to which he belongs. Larger groups (group A in our case) protect the (ethnic) identity of their components more efficiently. This effect, on average, clearly makes national feelings weaker in a majority group than in a minority group. Hence, the *size effect* has a negative impact on the difference between α^A and α^B .

On the other hand, an indirect effect is also at work. Resources are redistributed from the minority to the majority. This affects individual identity through changes in disposable income. When $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$, since intensity of national feeling is positively related to disposable income (Lemma 2.2), political outcomes have a positive effect on the national sentiments of group A (majority) and a negative one on the sentiments of group B (minority). Thus, the *political effect* positively influences $(\alpha^A - \alpha^B)$.

The balance of the two forces depends on the level of ethnic diversity. Suppose we live in a very heterogeneous society. This means that there are 2 groups of very similar size. The *size effect* plays no role. The majority group is relatively small and the benefits from the tax revenues are divided among a small fraction of people, so the tax rate is very high and inter-ethnic redistribution assumes large proportions. The *political effect* dominates the *size effect* and the average intensity of the national feeling of the majority group is stronger than the average intensity of national feeling of the minority. On the contrary, in homogeneous societies with a very large ethnic majority and a very small ethnic minority, the *size effect* is the prominent one and $(\alpha^A - \alpha^B)$ has a negative sign. The interplay between the two effects and the role of ethnic diversity is illustrated in Figure 2.5.¹⁷

Now, we turn to the relationship between ethnic diversity and the average intensity of national feeling in a country, which we call α^T ; we want to study the sign of $\frac{\partial \alpha^T}{\partial q_A}$.

Proposition 2.3 *Let $\lambda_{\alpha q} > \bar{\lambda}$, then $\frac{\partial \alpha^T}{\partial q_A} < 0$ if $q_A > \bar{q}_A$*

¹⁶The Appendix shows also how results change when we remove the conditions $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ and $\lambda_{\alpha q} > \bar{\lambda}$.

¹⁷We arbitrarily depicted linear intensity functions for both group A and group B.

Proof: In Appendix¹⁸

We can see α^T as a weighted average of α^A and α^B , where the weights are, of course, the sizes of the 2 groups.

When q_A increases (and, of course, at the same time q_B decreases), both the *size effect* and the *political effect* (lower inter ethnic redistribution) have a negative impact on α^A and a positive one on α^B . As a consequence, there are two conflicting effects on α^T : $q_A\alpha^A$ decreases while $q_B\alpha^B$ increases.

The balance of these 2 forces depends on the shape of the intensity function (namely, on α_{qq} and α_{qy} that, in turn, depends on several characteristics of the utility function and cost function). The effect on group A, however, is directed towards a larger share of the population (the weight of group A is larger than group B and becomes larger when q_A increases). This explains the previous proposition.

Let us summarize the results obtained in the following corollary.

Corollary 2.1 *There is no negative relationship between ethnic diversity and average national identity. Once we distinguish between minority and majority groups, however, it is possible to claim that: on the one hand, in countries with low ethnic diversity, on average, individuals from minorities have stronger national sentiments than individuals from majorities; on the other hand, in countries with high ethnic diversity, on average, individuals from minorities have weaker national sentiments than individuals from majorities.*

2.4 Discussion: a model with n groups

The empirical section of the chapter produced, among other considerations, two interesting results:

- within each country, agents in larger ethnic minorities have a lower intensity of national identity.
- the results are robust only if we use ethnic polarization as an index of ethnic diversity.

¹⁸Also in this case, the Appendix shows how the results change when we relax the conditions $\frac{\partial u(x, y, z)}{\partial x \partial y} > 0$ and $\lambda_{\alpha q} > \bar{\lambda}$.

Both of the results mentioned above need the theoretical support of a model with n groups.

The first change we need to consider is related to the policy space. In an n group setting there are $n+1$ types of transfer: transfers to all the population and transfers towards each ethnic group z , with $z = 1..n$. It turns out that, as long as the size of the largest ethnic group (group A) is greater than a half ($q_A > \frac{1}{2}$), Proposition 2.1 and Lemma 2.1 still holds. There is still a Condorcet winner, the identity of the median voter is still the same and, moreover, the lower the size of the largest ethnic group, the higher the redistribution *between* the ethnic groups (from the ethnic minorities to group A). In the following part of this section, we keep assuming $q_A > \frac{1}{2}$.

Lemma 2.4 *If $q_A > \frac{1}{2}$ there exists a unique voting equilibrium π^M . The tax rate is $\tau^M = \frac{(1-q_A)(y-\varphi)}{\gamma}$ and all the tax revenue is devoted to transfers to group A. $Tr_A^M = \left(\tau^M y - \frac{\gamma(\tau^M)^2}{2} \right)$*

Proof: In Appendix

The intensity of the national feelings of each individual i, z depends on the size of the ethnic group to which he belongs (q_z) and on his disposable income that, in turn, is a function of his initial income stream and of the size of the largest ethnic group (q_A), $\alpha_{i,z}^*(q_z; y_i, q_A)$. Lemma 2.2 and Lemma 2.3 still hold, then:

Proposition 2.4 *If $\lambda_{\alpha q}$ is positive and sufficiently large and $q_A > \frac{1}{2}$, then if $q_z < q_h$, with $z, h \neq A$, $\alpha_{i,z}^*(q_z; y_i, q_A) > \alpha_{i,h}^*(q_h; y_i, q_A)$*

Proof: It comes directly from Lemma 2.3

The proposition simply states that within each country, the agents in larger ethnic minorities have a lower intensity of national identity.

Let us define the average intensity of national sentiment within the minority (in a society in which the minority consists of several ethnic groups): $\alpha^m = \frac{\sum_{z \neq A} \alpha_{i,z}^*(q_z; y_i, q_A)}{1-q_A}$. We are interested in the sign of $\alpha^A - \alpha^m$.

From the previous discussion, it is straightforward to claim that: when $\lambda_{\alpha q} > \bar{\lambda}$ and $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$

- if $q_A > k$ then $\alpha^A < \alpha^m$ (if q_A is sufficiently large, the "size" effect dominates the "political" effect)

- if the ethnic composition of the population tends towards a perfect bipolar distribution (the maximum level of ethnic polarization), then $\alpha^A > \alpha^m$ (the "political" effect dominates the "size" effect).

Finally, let us consider the two basic properties of a polarization index:

Proposition 2.5 (i) *Let society consists of 3 groups of sizes, q_A, q_B and q_C , and $q_A > q_B \geq q_C$. Then if we merge the two smallest groups into a new group $\alpha^A - \alpha^m$ increases* (ii) *Let society consists of 3 groups of sizes, q_B, q_A , and q_B , and $q_A > q_B$. Then, if we shift mass (2Δ) from group A equally to the other two groups then $\alpha^A - \alpha^m$ increases.*

Proof: In Appendix

Both of the properties imply an increase in the polarization index; however, only the second one implies an increase in the fractionalization index¹⁹.

Since the "size effect", when we merge the two smallest group α^m decreases, while α^A is constant. This explains the first part of the proposition. Also the intuition of the second part of the proposition is quite straightforward. When we shift mass from group A equally to the other two groups, the size of the largest group decreases and then the between groups income redistribution increases. The size of each of the two minorities increases. Both the effects point in the same direction. National feelings among group A increase, while national feelings among each of the two minorities decrease.

These considerations suggest that, in countries with low ethnic polarization (q_A is sufficiently large), the average intensity of the national identity within the minority is higher than the average intensity of the national identity within the majority, while in countries with high ethnic polarization (close to a perfect bipolar distribution) the average intensity of the national identity within the majority is higher than the average intensity of the national identity within the minority. On the other hand, this is not necessarily true if we replace the ethnic polarization index with the ethnic fractionalization one.

2.5 Conclusions

This chapter investigates the relationship between ethnic diversity and individual identity. Using data from the World Values Surveys, we first provide some empirical facts,

¹⁹This is because we need to impose $q_A - 2\Delta > \frac{1}{2}$ in order to have a Condorcet winner.

then we develop a model that is consistent with the data. There is no evidence of a lower intensity of national sentiment in more ethnically fragmented countries or in minority groups. National feelings in a minority can be higher or lower than in a majority, depending on the degree of ethnic diversity of a country. On the one hand, in countries with high ethnic polarization, minorities have weaker national sentiments than majorities; on the other hand, in countries with low ethnic polarization, the reverse is true.

We then provide a model that can help explaining such empirical findings. The model builds on a previous work by Bisin et al. (2006). In each ethnic group there is an accepted social norm and deviating from that social norm involves some costs. Individuals develop national feelings in order to reduce these psychological costs. National sentiments are affected by political decisions (inter-ethnic redistribution) and by the ethnic composition of the population.

2.6 Appendix A

2.6.1 Proof of Proposition 2.1

Let us first prove that the policy preferred by the median voter is the one stated in the previous proposition and then let us prove that that policy is the winner of Condorcet.

- The median voter is the citizen belonging to group A such that $F(y^M, z = A) = \frac{1}{2}$ and so, since $F(y)$ is a uniform distribution and is independent of ethnicity, the income stream of the median voter is $y^M = y + \frac{\varphi}{q_A} - \varphi$. Clearly the best policy for an agent from group A is to devote all the tax revenue to transfers targeted only towards agents belonging to group A, so $Tr_B = Tr = 0$. As a result he will choose τ^M that maximizes his disposable income which is $y^{M,d} = (1 - \tau) \left(y + \frac{\varphi}{q_A} - \varphi \right) + \frac{\tau y - \frac{\tau^2}{2}}{q_A}$. It is straightforward to see that $\tau^M = \frac{(1-q_A)(y-\varphi)}{\gamma}$.
- Let us call π^M the policy preferred by the agent with income y^M .

$$\pi^M = \left(\tau^M = \frac{(1-q_A)(y-\varphi)}{\gamma}; Tr_A^M = \frac{\frac{(1-q_A)(y-\varphi)}{\gamma} y - \frac{[(1-q_A)(y-\varphi)]^2}{2\gamma}}{q_A}; Tr_B^M = 0; Tr^M = 0 \right)$$

Now, let us prove that the policy π^M , chosen by the individual with income stream y^M , is the winner of Condorcet. In a pairwise competition π^M must defeat all of the possible other policies.

- A general policy π^K ($\tau^K \leq \tau^M; Tr_A^K \geq 0; Tr_B^K \geq 0; Tr^K \geq 0$) would be defeated by π^M . All the agents belonging to A with $y_{i,A} \leq y^M$ would vote for π^M .
- It is easy to check that any other policy with $Tr_B = Tr = 0$ and $Tr_A > Tr_A^M$ would be also defeated by π^M . All the agents belonging to B and all the agents belonging to A with $y_{i,A} \geq y^M$ would vote for π^M .
- Let us now take a general policy π^K ($\tau^K > \tau^M; Tr_A^K \geq 0; Tr_B^K \geq 0; Tr^K \geq 0$). π^K generates $y_{i,A}^{d(K)}$ and $y_{i,B}^{d(K)}$.
- Let us now consider the set of policies π^H ($\tau^H > \tau^M; Tr_A^H \geq 0; Tr_B^H \geq 0; Tr^H = 0$). It is easy to see that it is always possible to find a policy π^H ($\tau^H = \tau^K > \tau^M; Tr_A^H \geq 0; Tr_B^H \geq 0; Tr^H = 0$) such that $y_{i,A}^{d(H)} = y_{i,A}^{d(K)}$ and $y_{i,B}^{d(H)} = y_{i,B}^{d(K)}$. We then check if there is any policy π^H that can defeat π^M .

It turns out that there is not. In order to prove this, we use the following procedure:

i) for each agent i,A with income $y_{i,A}$, we find the set of policies $\pi^{i,A}(\tau; Tr_A, Tr_B)$ such that $y_{i,A}^d(\pi^{i,A}) = y_{i,A}^d(\pi^M)$.

Such a set of policies can be summarized by the following equation:

$$Tr_A = (\tau - \tau^M) y_{i,A} + Tr_A^M$$

ii) if an agent i,A with income $y_{i,A}$ is indifferent between $\pi^{i,A}$ and π^M , then all the agents with lower income (and belonging to A) prefer $\pi^{i,A}$, while all the agents with higher income (and belonging to A) prefer π^M

It is straightforward since $\tau > \tau^M$.

iii) for each $\pi_{i,A}$ we find the size q'_A of agents from group A that would favor $\pi_{i,A}$ over π^M .

It turns out $q'_A(y_{i,A}) = \frac{q_A(y_{i,A} - y + \varphi)}{2\varphi}$.

iv) we then find the correspondent size of B that $\pi_{i,A}$ needs in order to defeat π^M .

$$q'_B(y_{i,A}) = \frac{1}{2} - q'_A(y_{i,A}).$$

v) if an agent i,B with income $y_{i,B}$ prefers a policy $\pi_{i,A}$ to π^M , all the agents with lower income (and belonging to B) do likewise.

As a result, in order $\pi_{i,A}$ to defeat π^M , it has to be true that $y'_{i,B}(\pi_{i,A}) > y'_{i,B}(\pi^M)$, where $y'_{i,B}$ is the disposable income of the agent from group B with income stream $y'_{i,B}$ such that $F(y_i, z = B) = q'_B$. Of course it is possible to write $y'_{i,B}$ as a function of q'_A and then as a function of $y_{i,A}$.

vi) we then find the set of policies such that $y'_{i,B}(\pi_{i,A}) > y'_{i,B}(\pi^M)$

$$Tr_A < \frac{\tau y}{q_A} - \frac{\gamma \tau^2}{2q_A} - \frac{(\tau - \tau^M)(1 - q_A)}{q_A} y'_{i,B}(y_{i,A})$$

vii) end of the proof.

The set of points such that:

$$\begin{aligned} Tr_A &< \frac{\tau y}{q_A} - \frac{\gamma \tau^2}{2q_A} - \frac{(\tau - \tau^M)(1 - q_A)}{q_A} y'_{i,B}(y_{i,A}) \\ Tr_A &= (\tau - \tau^M) y_{i,A} + Tr_A^M \\ \tau &\geq \tau^M. \end{aligned}$$

is an empty set for each possible level of $y_{i,A}$.

Intuition: There is no other policy able to beat π^M in a pairwise voting. π^M uses only transfers towards group A . This allows politicians to channel a high amount of resources towards a specific group of the population without using a high level of taxation and so without wasting a huge amount of resources (taxes are distortionary).

2.6.2 Proof of Lemma 2.1

$$\frac{\partial y'_{i,B}}{\partial q_A} = \frac{1}{\gamma} y_{i,B} (y - \varphi) > 0$$

$$\frac{\partial y'_{i,A}}{\partial q_A} = \frac{1}{\gamma} y_{i,A} (y - \varphi) + \frac{1}{q_A} \left(-\frac{y}{\gamma} (y - \varphi) - \frac{1}{2\gamma} (y - \varphi)^2 (2q_A - 2) \right) -$$

$$-\frac{1}{q_A^2} \left(\frac{y}{\gamma} (y - \varphi) (1 - q_A) - \frac{1}{2\gamma} (y - \varphi)^2 (1 - q_A)^2 \right) > 0$$

$$\text{if } y_{i,A} > \gamma \frac{\frac{1}{q_A} \left(-\frac{y}{\gamma} (y-\varphi) - \frac{1}{2\gamma} (y-\varphi)^2 (2q_A-2) \right) - \frac{1}{q_A} \left(\frac{y}{\gamma} (y-\varphi) (1-q_A) - \frac{1}{2\gamma} (y-\varphi)^2 (1-q_A)^2 \right)}{(y-\varphi)}$$

2.6.3 Proof of Lemma 2.2

As in Bisin et al (2006), we need to make a further assumption in order to have the Hessian matrix negative semidefinite, that is

$$D = -(u_{xx} - \lambda) \left[1 + \lambda_{\alpha\alpha} \frac{(x-x_0)^2}{2} \right] - \lambda_{\alpha}^2 (x-x_0)^2 > 0 \quad \forall i$$

Now we apply the implicit function theorem and we get:

$$\begin{aligned} \frac{\partial x_{i,z}^*}{\partial y_{i,z}} &= \frac{\lambda_{\alpha\alpha} \frac{(x-x_0)^2}{2} \frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} \frac{\partial y_{i,z}^d}{\partial y_{i,z}}}{D} \\ \frac{\partial \alpha_{i,z}^*}{\partial y_{i,z}} &= \frac{-\lambda_{\alpha} (x-x_0) \frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} \frac{\partial y_{i,z}^d}{\partial y_{i,z}}}{D} \end{aligned}$$

where $\frac{\partial y_{i,z}^d}{\partial y_{i,z}} > 0$ for every i, z

2.6.4 Proof of Lemma 2.3

We apply the implicit function theorem again and we get

$$\begin{aligned} \frac{\partial x_{i,z}^*}{\partial q_z} &= \frac{-\lambda_q (x-x_0) \left[1 + \lambda_{\alpha\alpha} \frac{(x-x_0)^2}{2} \right] + \lambda_{\alpha} \lambda_{\alpha q} \frac{(x-x_0)^3}{2} + \left[1 + \lambda_{\alpha\alpha} \frac{(x-x_0)^2}{2} \right] \frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} \frac{\partial y_{i,z}^d}{\partial q_z}}{D} \\ \frac{\partial \alpha_{i,z}^*}{\partial q_z} &= \frac{(x-x_0)^2 \lambda_{\alpha} \lambda_q + (u_{xx} - \lambda) \lambda_{\alpha q} \frac{(x-x_0)^2}{2} - \lambda_{\alpha} (x-x_0) \frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} \frac{\partial y_{i,z}^d}{\partial q_z}}{D} \end{aligned}$$

2.6.5 Proof of Proposition 2.2

$\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ and $\lambda_{\alpha q} > \bar{\lambda}$. We know from Lemma 2.2 and Lemma 2.3 that $\frac{\partial \alpha_{i,z}^*}{\partial y_{i,z}^d} > 0$ $\forall i, z$ and $\frac{\partial \alpha_{i,z}^*}{\partial q_z} < 0 \forall z$. We start by considering individuals with the same initial income

stream y^* , but who belong to different ethnic groups. As emerged from the previous discussion, we know $\alpha_{i,z}^* \left(y_{i,z}^d (y^*; q_A); q_z \right)$. We compare $\alpha_{i,A}^* \left(y_{i,A}^d (y^*; q_A); q_A \right)$ and $\alpha_{i,B}^* \left(y_{i,B}^d (y^*; q_A); q_B \right)$ and we study 2 cases:

1) $(q_A - q_B) \rightarrow 0$

$y_{i,A}^d (y^*) - y_{i,B}^d (y^*) > 0$, and so since $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$, by Lemma 2.2, we can claim that

$$\alpha_{i,A}^* \left(y_{i,A}^d (y^*; q_A); q_A \right) - \alpha_{i,B}^* \left(y_{i,B}^d (y^*; q_A); q_B \right) > 0$$

2) $q_A \rightarrow 1; q_B \rightarrow 0$

From Proposition 2.1 and Lemma 2.1, $\tau \rightarrow 0$ and then, $\left[y_{i,A}^d (y^*) - y_{i,B}^d (y^*) \right] \rightarrow 0$,

$$\alpha_{i,A}^* \left(y_{i,A}^d (y^*; q_A); q_A \right) - \alpha_{i,B}^* \left(y_{i,B}^d (y^*; q_A); q_B \right) < 0$$

Since we assumed that $F(y)$ is uniform and independent of ethnicity, we can then conclude that $\alpha^A - \alpha^B > 0$ when $q_A \rightarrow \frac{1}{2}$ and $\alpha^A - \alpha^B < 0$ when $q_A \rightarrow 1$. Since $\alpha^A - \alpha^B$ is a decreasing function of q_A , it is possible to claim that there exists a value q_A^* such that $\alpha^A > \alpha^B$ if $q_A < q_A^*$; $\alpha^A < \alpha^B$ otherwise. This proves Proposition 2.2.

Let us now investigate how results change when we remove the conditions $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ and $\lambda_{\alpha q} > \bar{\lambda}$, that is when we do not restrict ourselves to having a negative "size effect" on national identity and a positive correlation between national sentiments and disposable income. As is predictable, the interaction between *size effect* and *political effect* delivers very different results, depending on the signs of the two cross derivatives:

Proposition 2.2 bis:

- if $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ and $\lambda_{\alpha q} > \bar{\lambda}$, then $\alpha^A > \alpha^B$ if $q_A < k$; $\alpha^A < \alpha^B$ otherwise
- if $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} > 0$ and $\lambda_{\alpha q} < \underline{\lambda}$, then $\alpha^A > \alpha^B$
- if $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} < 0$ and $\lambda_{\alpha q} > \bar{\lambda}$, then $\alpha^A < \alpha^B$
- if $\frac{\partial u(x, y_{i,z}^d)}{\partial x \partial y_{i,z}^d} < 0$ and $\lambda_{\alpha q} < \underline{\lambda}$, then $\alpha^A < \alpha^B$ if $q_A < k$; $\alpha^A > \alpha^B$

2.6.6 Proof of Proposition 2.3

We define the average intensity of national feelings in a country α^T , and how it depends on ethnic diversity.

$$\begin{aligned}\alpha^T &= \int_{y-\varphi}^{y+\varphi} \alpha_{i,z}^* (y_i; q_z) dF(y_i) = \int_{y-\varphi}^{y+\varphi} \alpha_{i,A}^* (y_i; q_A) dF(y_i, z = A) + \\ &+ \int_{y-\varphi}^{y+\varphi} \alpha_{i,B}^* (y_i; q_B) dF(y_i, z = B) = \\ &= q_A \int_{y-\varphi}^{y+\varphi} \alpha_{i,A}^* (y_i; q_A) dF(y_i|z = A) + (1 - q_A) \int_{y-\varphi}^{y+\varphi} \alpha_{i,B}^* (y_i; q_B) dF(y_i|z = B) \\ \frac{\partial \alpha^T}{\partial q_A} &= \alpha^A - \alpha^B + q_A \int_{y-\varphi}^{y+\varphi} \frac{\partial \alpha_{i,A}^*}{\partial q_A} dF(y_i|z = A) - (1 - q_A) \int_{y-\varphi}^{y+\varphi} \frac{\partial \alpha_{i,B}^*}{\partial q_B} dF(y_i|z = B)\end{aligned}$$

Since $\lambda_{\alpha q} > \bar{\lambda}$, when $q_A \rightarrow 1$, by Proposition 2.2 bis, $\alpha^A - \alpha^B < 0$.

$q_A \int_{y-\varphi}^{y+\varphi} \frac{\partial \alpha_{i,A}^*}{\partial q_A} dF(y_i|z = A) < 0$ and the last term goes to 0 (provided that $\lambda(\cdot)$ and $u(\cdot)$ are such that $\frac{\partial \alpha_{i,z}^*}{\partial q_z}$ is bounded below). This proves Proposition 2.3.

Let us now investigate how the results change when we remove the condition $\lambda_{\alpha q} > \bar{\lambda}$.

Proposition 2.3 bis:

- If $\lambda_{\alpha q} > \bar{\lambda}$, then $\frac{\partial \alpha^T}{\partial q_A} < 0$ if $q_A > \bar{q}_A$
- If $\lambda_{\alpha q} < \underline{\lambda}$, then $\frac{\partial \alpha^T}{\partial q_A} > 0$ if $q_A > \underline{q}_A$

2.6.7 Proof of Lemma 2.4

The structure of the proof is very similar to the case with 2 groups. We need to check if there is any policy π^H that can defeat π^M where π^H is such that

$$\pi^H (\tau^H > \tau^M; Tr_j^H \geq 0 \forall j; Tr^H = 0)$$

i) for each agent i, A with income $y_{i,A}$, we find the set of policies $\pi^{i,A} (\tau; Tr_A, Tr_j^H)$ such that $y_{i,A}^d (\pi^{i,A}) = y_{i,A}^d (\pi^M)$

ii) if an agent i, A with income $y_{i,A}$ is indifferent between $\pi^{i,A}$ and π^M , then all the agents with lower income (and belonging to A) prefer $\pi^{i,A}$, while all the agents with higher income (and belonging to A) prefer π^M . As a second step we find the size q'_A of agents from group A that would favor $\pi_{i,A}$ over π^M .

iii) then we find the amount of transfers $Tr_j^* \geq 0$ needed for a generic amount $q_j' \geq 0$ from $n-2$ groups to support $\pi_{i,A}$ and finally we check if the remaining part of the budget is enough to convince $\frac{1}{2} - q_A' - \sum_{j \neq A, N} q_j'$ individuals from the remaining group (group N) to support $\pi_{i,A}$.

In order to convince q_j' people from group j , we need to provide Tr_j^* such that $Tr_j^* = (\tau - \tau^M) y_{i,j}$ where $y_{i,j}$ is such that $q_j' = \frac{q_j(y_{i,j} - y + \varphi)}{2\varphi}$.

It turns out that in order to convince $\frac{1}{2} - q_A' - \sum_{j \neq A, N} q_j'$ from group N to support $\pi_{i,A}$, it must be true that

$$Tr_A \leq \frac{\tau y}{q_A} - \frac{\gamma \tau^2}{2q_A} + (\tau - \tau^M)(y_{i,A} - y) - \frac{(\tau - \tau^M) q_N y}{q_A}$$

As a final step, we need to check if the following system has any solution

$$\begin{aligned} Tr_A &\leq \frac{\tau y}{q_A} - \frac{\gamma \tau^2}{2q_A} + (\tau - \tau^M)(y_{i,A} - y) - \frac{(\tau - \tau^M) q_N y}{q_A} \\ Tr_A &= (\tau - \tau^M) y_{i,A} + Tr_A^M \\ \tau &\geq \tau^M \\ Tr_A &\leq \frac{\tau y}{q_A} - \frac{\gamma \tau^2}{2q_A} \end{aligned}$$

It turns out that the system does not have any solution for $\forall y_{i,j}$ with $j \neq N$

2.6.8 Proof of Proposition 2.5

(i) It comes directly from Lemma 2.3 (ii) We shift mass from group A equally to the other two groups. So the size of group A becomes $q_A - 2\Delta$, while the size of each of the other 2 groups becomes $q_B + \Delta$. Of course, in order to have a Condorcet winner, it is crucial that $q_A - 2\Delta > \frac{1}{2}$. From Lemma 2.3 we know that, when $\lambda_{\alpha q} > \bar{\lambda}$, $\alpha_{i,A}^*$ is a decreasing function of q_A ; as a result, when the size of group A decreases by 2Δ , α_A increases. On the other hand, $\alpha_{i,z}^*$ depends on the size of the ethnic group to which he belongs (q_z) and on his disposable income, that is a function of his initial income stream and of the size of the largest ethnic group (q_A), $\alpha_{i,z}^*(q_z; y_i, q_A)$. From Lemma 2.3 $\frac{\partial \alpha_{i,z}^*}{\partial q_z} < 0$ and from Lemma 2.1 $\frac{\partial \alpha_{i,z}^*}{\partial q_A} > 0$. As a result, when we shift mass Δ from group A equally to the other two groups, α_m decreases.

2.7 Appendix B

2.7.1 Definition of variables

-*National identity*: dummy equal to 1, if the respondent says "*I am an American first and a member of some ethnic group second*" (US example). Source: WVS

-*Minority*: dummy equal to 0, if the ethnic group to which the individual belongs is the largest in the country; otherwise it is equal to 1. Each interviewer has been asked to code the ethnic group of each individual in the sample. Source: WVS

-*Age*: age of the respondent. Source: WVS

-*Female*: dummy equal to 1, if the respondent is female. Source: WVS

-*Primary*: dummy equal to 1, if the respondent has, at most, primary school education. Source: WVS

-*Married*: dummy equal to 1, if respondent is married. Source: WVS

-*Income*: Proxy for individual income stream (not disposable). There are 10 income categories (net of transfers and taxes); they have been coded by deciles for each country, 1=lowest decile; 10=highest decile. Source: WVS

-*Employer>10*: dummy equal to 1, if the respondent is employer/manager of an establishment with 10 or more employees. Source: WVS

-*Employer<10*: dummy equal to 1, if the respondent is employer/manager of an establishment with less than 10 employees. Source: WVS

-*Professional*: dummy equal to 1, if the respondent is a professional worker, lawyer, accountant, teacher, etc Source: WVS

-*Supervisor*: dummy equal to 1, if the respondent is a supervisory, non-manual office worker Source: WVS

-*Office*: dummy equal to 1, if the respondent is a non-supervisory, non-manual office worker. Source: WVS

-*Foreman*: dummy equal to 1, if the respondent is a foreman and supervisor Source: WVS

-*Skilled*: dummy equal to 1, if the respondent is a skilled manual worker Source: WVS

-*Semi-skilled*: dummy equal to 1, if the respondent is a semi-skilled manual worker
Source: WVS

-*Unskilled*: dummy equal to 1, if the respondent is an unskilled manual worker Source:
WVS

-*Farmer*: dummy equal to 1, if the respondent has his/her farm Source: WVS

-*Armed forces*: dummy equal to 1, if the respondent is a member of armed forces.
Source: WVS

-*Never job*: dummy equal to 1, if the respondent never had a job Source: WVS

-*Size*: size of ethnic group to which the individual belongs. Source: Fearon's dataset

-*Etfra*: country index of ethnic fractionalization. Source: Fearon's dataset

-*Etpol*: country index of ethnic polarization. Source: Fearon's dataset

2.7.2 Countries and ethnic groups

2.7.2.1 Criteria of inclusion

We include in the sample only countries whose ethnic groups coded by the interviewers match the ones included by Fearon in his dataset (and whenever there is not a perfect match between WVS and Fearon's source, we also include those countries whose ethnic groups as coded by the interviewers match the ones listed by the CIA in "The World Factbook", as in the case of South Africa and Zimbabwe²⁰).

2.7.2.2 Countries selected

For each ethnic group we report the average values of the variable "national identity".

US (wave IV): white (.29); black (.18); Asian (.09); Hispanic (.23); other (.4)

US (wave III): white (.33); black (.17); Asian (.35); other (.18)

US (wave II): white (.44); black (.23); Asian (.41); Hispanic (.23); other (.49)

Canada (wave IV): english canadian (.67); french canadian (.15); black (.51); south Asian (.53); east Asian (.5); other (.54)

²⁰Results do not change if we remove South Africa and Zimbabwe from the sample.

Spain (wave IV): Castellano (.33); Catalan (.34); Basque (.16); Galician (.07); other (.32)

Spain (wave III): Castellano (.4); Catalan (.24); Basque (.19); Galician (.09); other (.25)

China (wave II): Han (.06); other (.58)

South Africa (wave IV): black (.12); white (.21); coloured (.43); Indian (.37)

Brazil (wave III): white (.49); mulatto (.55); black (.37); other (.15)

Bielorussia (wave III): Bielorussian (.28); Russian (.5); Polish (.53); Ukranian (.75); other (.95)

Latvia (wave III): Latvian (.27); Russian (.54); Byelorussian (.6); Ukrainian (.71); Polish (.68)

Uruguay (waveIV): white (.79); black (.81); south Asian (.68); other (.85)

Armenia (wave III): Armenian (.04); Russian (.33); Kurd (.2)

Azerbaijan (wave III): Azerbaijani (.03); Russian (.13); Tataria (.26); Lezgin (.14); other (.19)

Macedonia (wave IV): Macedonian (.17); Albanian (.05); Turkish (.4); Serb (.6); other (.52)

Albania (wave IV): Albanian (.01); Greek (.5); other (.33)

Bosnia (waveIV): Bosniak (.53); Serbs (.08); Croats (.4); other (.96)

Indonesia (wave IV): Javanese (.54); Malay (.54); Chinese (1); Sundanese (.4); Sumatranese (.5)

Israele (wave IV): Jewish (.51); Arabic (0)

Moldova (wave III): Moldovian (.11); Russian (.34); Bulgarian (.3); Ukrainian (.67); Gaugas (.16)

Pakistan (wave IV): Punjabi (.83); Pathan (.7); Baluchi (.54); Sindhi (.6); Urdu speaking (.85)

Singapore (wave IV): Malay (.11); south Asian (.3); Chinese (.53); Arabic (.08)

Zimbabwe (wave IV): black (.07); white (.07); other (0)

Jordania (wave IV): Jordanian (.2); Palestinian (.07); other (.1)

Georgia (wave III): Georgian (.17); other (.15)

2.8 Figures and Tables

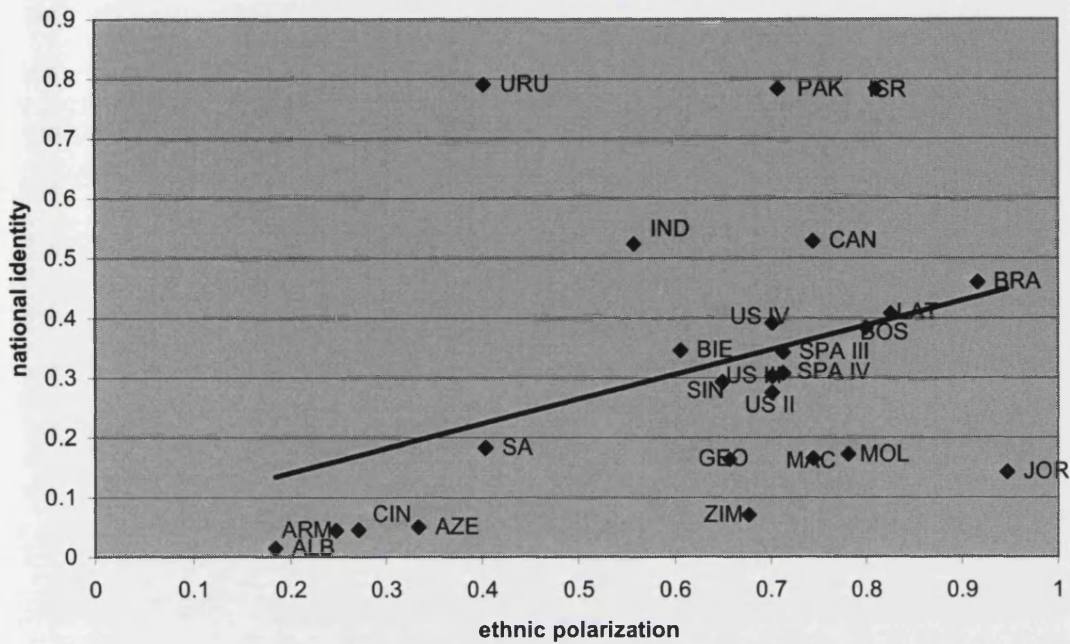


Figure 2.1: National Identity and Ethnic Polarization

Note: The country level of National Identity is measured as the fraction of individuals in each country who choose to identify with the nation rather than with the ethnic group.

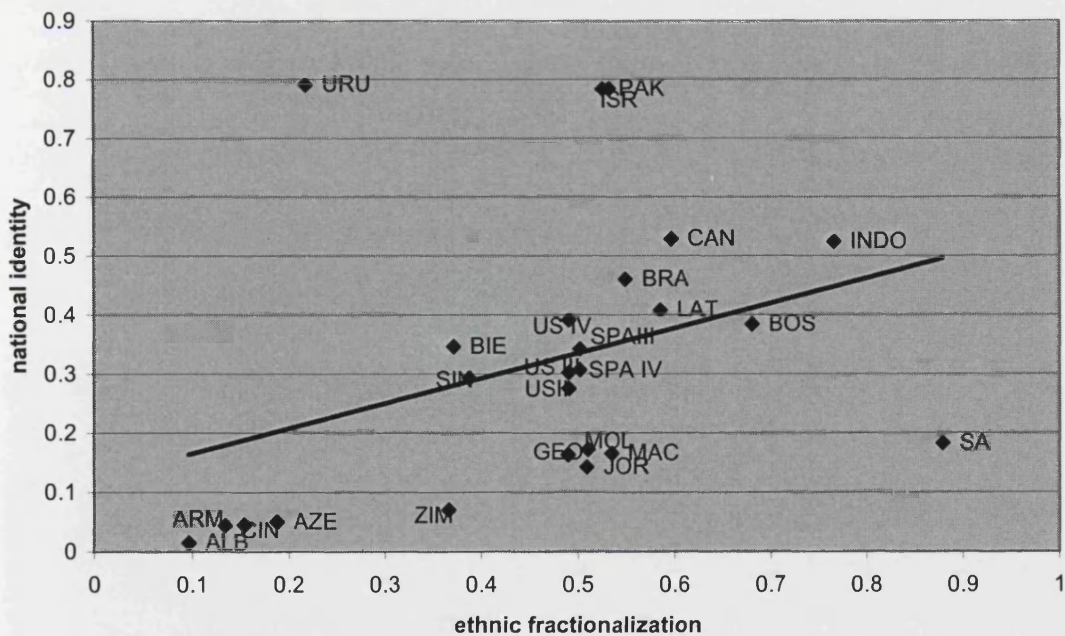


Figure 2.2: National Identity and Ethnic Fractionalization

Note: The country level of National Identity is measured as the fraction of individuals in each country who choose to identify with the nation rather than with the ethnic group.

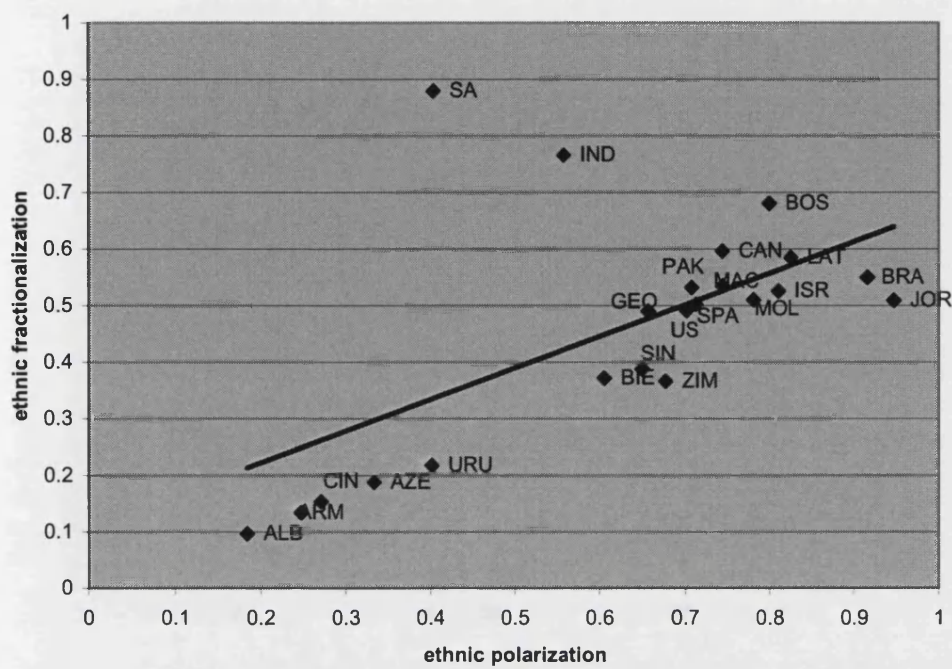


Figure 2.3: Ethnic polarization vs Ethnic Fractionalization

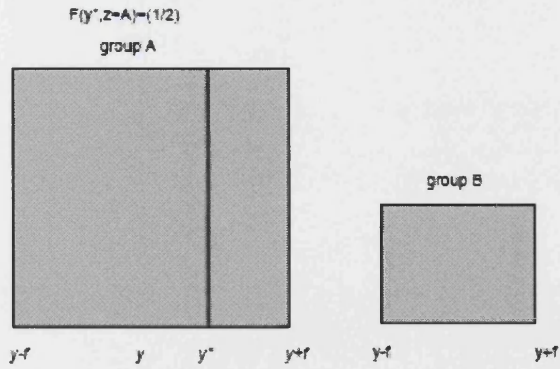


Figure 2.4: Median Voter

Note: Figure 2.3 depicts both the (uniform) income distribution of the majority group (A) and of the minority group (B). The median voter belongs to group A and has a higher initial income stream than the average

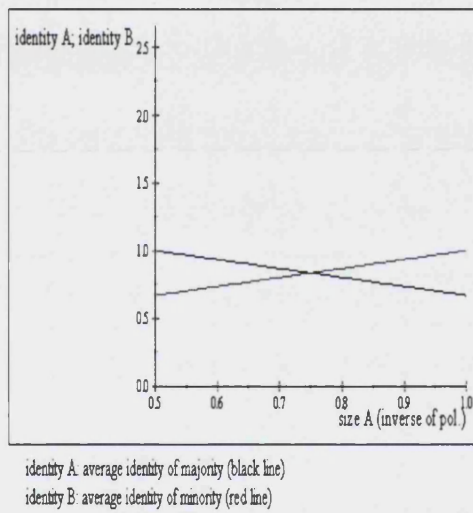


Figure 2.5: Intensity as a function of the size of group A

Note: Figure 2.4 depicts the average level of national identity of each ethnic group as a function of its size. We arbitrarily depict linear intensity function.

Table 2.1: National identity by country

National Identity		National Identity	
Albania (w. IV)	0.015	Macedonia (w. IV)	0.1655
Armenia (w. III)	0.0444	Moldova (w. III)	0.1726
Azerbaijan (w.III)	0.0501	Pakistan (w. IV)	0.7835
Belorussia (w. III)	0.3459	Singapore (w. IV)	0.2929
Bosnia (w.IV)	0.3837	South Africa (w. IV)	0.1828
Brazil (w. III)	0.4599	Spain (w. IV)	0.3075
Canada (w. IV)	0.5288	Spain (w. III)	0.3424
Cina (w. II)	0.0454	Uruguay (w.IV)	0.7906
Georgia (w.III)	0.1633	US (w. IV)	0.3912
Indonesia (w. IV)	0.524	US (w. III)	0.3025
Israele (w.IV)	0.7835	US (w. II)	0.2752
Jordania (w.IV)	0.1427	Zimbabwe (w. IV)	0.0703
Latvia (w. III)	0.4077		

Note: The country level of National Identity is measured as the fraction of individuals in each country who choose to identify with the nation rather than with the ethnic group.

Table 2.2: Descriptive Statistics

	Mean	St. dev.	obs
Age	42.41	40.76	37377
Female	0.5255	0.4993	37377
Married	0.5805	0.3934	37377
Income	4.34	2.34	37377
manager>10	0.0302	0.1713	35319
manager<10	0.0432	0.2033	35319
professional	0.1378	0.3447	35319
supervisor	0.0611	0.2395	35319
nonmanual	0.0957	0.2942	35319
foreman	0.0347	0.183	35319
skilled	0.1485	0.3536	35319
semiskilled	0.0811	0.273	35319
unskilled	0.09	0.2862	35319
farmer	.0239	0.1529	35319
armed forces	0.0452	0.2079	35319
never worked	0.19	0.3923	35319
primary	0.2283	0.4197	34533
av. Group income	3.7825	1.5861	95
minority	0.2525	0.2502	37377
etfra	0.4667	0.1885	25
size	0.1825	0.2669	95
etpol	0.6446	0.2119	25
national identity	0.675	0.1671	37377

Table 2.3 Sources of national identification: individual characteristics

	national identity			
Age	.0001 (0.95)	0.0004 (1.33)	.0001 (0.96)	0.0003 (1.25)
Female	-0.0005 (-0.06)	0.0028 (0.32)	-0.0013 (-0.17)	0.0018 (0.88)
married	0.0133 (1.54)	0.0095 (1.09)	0.0085 (0.96)	0.0053 (0.6)
income	0.0077** (2.07)	0.0066* (1.86)	0.0065** (2.11)	0.0064** (2.08)
minority	-0.0387 (-0.81)	-0.0477 (-0.98)	-0.0307 (-0.66)	-0.0411 (-0.87)
primary(at most)		-0.027** (-2.33)		-0.0222 (-1.64)
man>10			0.0182 (0.87)	0.0046 (0.19)
man<10			0.0456* (1.77)	0.0366 (1.41)
professional			0.074*** (3.25)	0.0598*** (2.83)
supervisor			0.0199 (1.09)	0.0083 (0.43)
nonmanual			.0119 (0.7)	0.0033 (0.18)
foreman			0.05*** (3.10)	0.0427** (2.26)
skilled			0.0075 (0.5)	0.0066 (0.41)
semiskilled			-0.0024 (-0.2)	0.0003 (0.03)
unskilled			0.0043 (0.23)	0.0074 (0.39)
farmer			0.0432** (2.57)	0.0465** (2.72)
armforc			0.0522** (2.33)	0.0442* (1.82)
neverworked			0.0058 (0.34)	0.0036 (0.21)
country-time f.e.	yes	yes	yes	yes
observations	32644	30038	31000	28514
country-time obs	25	23	24	22
R squared	0.2152	0.2141	0.2289	0.2278

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. The variable *primary* is a Dummy equal to 1 if the respondent has a low level of education (with at most primary education). The variable *minority* is a Dummy equal to 1 if the individual belongs to a minority group. All specifications include country-time fixed effects. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.4 Sources of national identification: country characteristics (1)

National identity								
income	0.0218*** (4.5)	0.0173*** (3.33)	0.0226*** (4.03)	0.0186*** (3.12)	0.0213*** (4.56)	0.0174*** (3.72)	0.0226*** (4.24)	0.0188*** (3.44)
minority	-0.0223 (-0.45)	-0.0269 (-0.47)	-0.0333 (-0.65)	-0.0394 (-0.75)	-0.0105 (-0.22)	-0.0164 (-0.74)	-0.0221 (-0.45)	-0.00297 (-0.6)
etpol	0.3768** (2.08)	1.9612*** (3.72)	0.3011 (1.6)	1.9657*** (3.61)	0.346* (1.91)	1.9807*** (3.57)	0.3028 (1.39)	1.9706*** (3.51)
etpol squared		-1.3971** (-2.81)		-1.4151** (-2.8)		-1.4409*** (-2.79)		-1.4532*** (-2.83)
country*time F.E.	no	no	no	no	no	no	no	no
country*time obs	25	25	23	23	24	24	22	22
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all
R squared	0.0412	0.0693	0.0343	0.0639	0.0477	0.0786	0.0427	0.0752

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.5 Sources of national identification: country characteristics (2)

	National identity							
income	0.0183*** (3.37)	0.0191*** (3.93)	0.0195*** (2.94)	0.0206*** (3.49)	0.0179*** (3.33)	0.0192*** (4.02)	0.0196*** (3.11)	0.0211*** (3.8)
minority	-0.0272 (-0.51)	-0.036 (-0.72)	-0.0373 (-0.67)	-0.0472 (-0.92)	-0.0147 (-0.28)	-0.0245 (-0.51)	-0.0258 (-0.49)	-0.0363 (-0.74)
etfra	0.3999 (2.56)	1.3939** (2.6)	0.3521 (1.23)	1.3585** (2.17)	0.3703 (1.49)	1.3129** (2.43)	0.3189 (1.15)	1.2529* (1.94)
etfra squared		-1.127** (-2.15)		-1.1151* (-1.94)		-1.0641* (-2)		-1.0313* (-1.75)
country*time F.E.	no	no	no	no	no	no	no	no
country*time obs	25	25	23	23	24	24	22	22
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all
R squared	0.0385	0.0524	0.0316	0.0451	0.0465	0.0592	0.0418	0.0537

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.6 Sources of national identification: country characteristics (3)

National identity								
income	0.0077** (2.07)	0.0068* (1.93)	0.0066* (1.86)	0.0059* (1.72)	0.0065** (2.11)	0.0058* (1.99)	0.0064** (2.08)	0.0058* (1.99)
minority	-0.0387 (-0.81)	0.3539** (2.58)	-0.0477 (-0.98)	0.2581** (2.11)	-0.0307 (-0.66)	0.3348** (2.34)	-0.0411 (-0.87)	0.2358* (2)
etpol*min		-0.5647** (-2.51)		-0.435** (-2.08)		-0.479** (-2.48)		-0.396* (-1.94)
country*time F.E.	yes	yes	yes	yes	yes	yes	yes	yes
country*time obs	25	25	23	23	24	24	22	22
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all
R squared	0.2152	0.2213	0.2141	0.2175	0.2289	0.2345	0.2278	0.2307

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. All the specifications include country time fixed effects. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.7 Sources of national identification: country characteristics (4)

	National identity							
income	0.0077** (2.07)	0.0086** (2.31)	0.0066* (1.86)	0.0071* (1.98)	0.0065** (2.11)	0.0074** (2.41)	0.0064** (2.08)	0.0069** (2.23)
minority	-0.0387 (-0.81)	0.1662 (0.98)	-0.0477 (-0.98)	0.0728 (0.46)	-0.0307 (-0.66)	0.1724 (1.08)	-0.0411 (-0.87)	0.0751 (1.76)
etfra*min		-0.386 (-1.14)		-0.2235 (-0.71)		-0.3813 (-1.18)		-0.2147 (-0.72)
country*time F.E.	yes	yes	yes	yes	yes	yes	yes	yes
country*time obs	25	25	23	23	24	24	22	22
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all
R squared	0.2152	0.2177	0.2141	0.2149	0.2289	0.2315	0.2278	0.2286

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. All the specifications include country time fixed effects. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.8: Sources of national identification: size of the group -only majority groups-

		National identity							
av.group inc.		0.054		0.0478		0.0666		0.0591	
		(1.26)		(1.18)		(1.43)		(1.37)	
size	-0.7095**	-0.6749**	-0.6472**	-0.6299*	-0.6595**	-0.6042*	0.5823*	-0.5515*	
	(-2.52)	(-2.26)	(-2.13)	(-2.01)	(-2.38)	(-1.99)	(-1.91)	(-1.72)	
country*time F.E.	no	no	no	no	no	no	no	no	no
country*time obs	25	25	23	23	24	24	22	22	
total obs.	23523	23523	21236	21236	22215	22215	20030	20030	
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all	
R squared	0.0726	0.0839	0.0616	0.0707	0.0795	0.0961	0.0721	0.0853	

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. The variable *size* measures the size of the ethnic group which the respondent belongs to. The variable *average group income* measures the sample average of the income of the components of the ethnic group which the respondent belongs to. All the specifications include only respondents belonging to the largest ethnic group (the majority group) of the country where they reside. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.9 Sources of national identification: size of the group -only minorities- (1)

		National identity							
av.group inc.		0.0623**		0.0646**		0.0626**		0.065***	
		(2.68)		(2.78)		(2.76)		(2.86)	
size	-0.6862***	-0.6204***	-0.68**	-0.6046***	-0.6501***	-0.59389**	-0.6408***	-0.5762***	
	(-4.19)	(-4.17)	(-4.06)	(-4.09)	(-3.54)	(3.59)	(-3.47)	(-3.54)	
country*time F.E.	no	no	no	no	no	no	no	no	no
country*time obs	25	25	23	23	24	24	22	22	
total obs.	8968	8968	8649	8649	8641	8641	8340	8340	
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all	
R squared	0.0576	0.0738	0.0617	0.0789	0.058	0.0735	0.0616	0.0783	

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. The variable *size* measures the size of the ethnic group which the respondent belongs to. The variable *average group income* measures the sample average of the income of the components of the ethnic group which the respondent belongs to. All the specifications include only respondents belonging to the minority groups. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.10 Sources of national identification: size of the group -only minorities- (2)

		National identity							
av.group inc.		0.0426		0.0465		0.0421		0.0469	
		(0.96)		(1.02)		(0.92)		(1.01)	
size	-0.792*	-0.7611*	-0.7924*	-0.7599*	-0.8125*	-0.7806*	-0.8107*	-0.7761*	
	(-1.85)	(-1.94)	(-1.85)	(-1.94)	(-1.89)	(-1.97)	(-1.88)	(-1.97)	
country*time F.E.	yes	yes	yes	yes	yes	yes	yes	yes	yes
country*time obs	25	25	23	23	24	24	22	22	
total obs.	8968	8968	8649	8649	8641	8641	8340	8340	
control	standard	standard	+ edu	+ edu	+occup	+occup	all	all	
R squared	0.1724	0.174	0.1753	0.1773	0.174	0.1756	0.1762	0.1782	

Note: The dependent variable is a Dummy variable equal to 1 if the respondent identifies himself with the nation rather than his ethnic group. The variable *size* measures the size of the ethnic group which the respondent belongs to. The variable *average group income* measures the sample average of the income of the components of the ethnic group which the respondent belongs to. All the specifications include only respondents belonging to the minority groups and country time fixed effects. The specifications reported in Columns 1 and 2 control for individual income. The specifications reported in Columns 3 and 4 control for individual income and the level of education of the respondent. The specifications reported in Columns 5 and 6 control for individual income and the occupation of the respondent. The specifications reported in Columns 7 and 8 control for individual income, the level of education and the occupation of the respondent. t-statistics are reported between parenthesis. Standard errors are clustered at the country-time level. *Significant at 10%, ** significant at 5%, *** significant at 1%.

Chapter 3

Identity Formation and Language Policies: the Catalan Case

(joint with Oriol Aspachs-Bracons and Irma Clots-Figueras)

3.1 Introduction

Several social scientists (Anderson 1983, Bates 1983, Horowitz 1985) have argued that the boundaries of ethnic groups are not exogenous and change over time. They are likely to be affected by social and economic conditions and be dependent on a number of economic and policy choices.

Thus, ethnic diversity can be influenced by precise political strategies aimed at promoting national or ethnic identities among the population. In fact, in ethnically diverse countries, politicians and political scientists have focused on very different policies as mechanisms of minority integration and conflict reduction. Nation building policies (for example the implementation of a unique national language across the entire territory of a State) as well as multicultural policies that explicitly recognize cultural differences (regional autonomy, the devolution of powers, the use of multiple languages in schools and in other contexts) have often been proposed as sources of conflict management.

Limited research, however, has examined the extent to which identity can be shifted by cultural policies and regulation. As the UN report 2004 pointed out, language regulation does exert an influence on the social environment that encourages identity formation. Since the core individual identity is typically formed in the early stages of life, language in schools turns out to be particularly important.

This chapter takes a step in this direction by analyzing the effect of a particular educational policy on the process of identity formation. In 1983, the Catalan education system became bilingual, and Catalan, together with Spanish, was taught in schools. Using survey data from Catalonia and exploiting within and between cohort variation in the exposure to the Catalan language in schools, the results show that individuals who have experienced greater exposure to teaching in Catalan are more likely to say that they feel more Catalan than Spanish. Interestingly, this effect appears to be positive also among individuals educated in Catalonia after the reform but whose parents do not have Catalan origins.

This reform can be interpreted as an example of a multicultural policy in Spain, where individuals living in Catalonia are the relevant minority and, at the same time, as an example of a nation building policy in Catalonia, where individuals living in Catalonia who were born elsewhere or whose parents do not have Catalan origins are the relevant minority. As a result, nation building policies and multicultural policies deliver very different outcomes in terms of the effects on individual sentiments: multicultural policies tend to favour the development of regional identities, while nation building policies tend to promote the expansion of a common national feeling.

Our empirical argument relies on within and between cohort variation in exposure to Catalan language at school. Between cohorts, the 1983 educational reform affected younger cohorts more than the older generations. Within a given cohort, the exposure to Catalan varies according to the number of years they remain in education (some students leave school earlier than others).

A number of factors leads us to believe that the link between the linguistic reform and individual identity is causal. Firstly, a number of robustness checks allow us to rule out the existence of education specific cohorts' trends for Catalan feelings. Secondly, we can exclude the idea that the reform changed the composition of the Catalan population, as comparisons between migration patterns in Catalonia and other Spanish regions show that Catalonia does not present anomalous patterns. Moreover, since individual years of schooling are typically an individual and parental choice, we may be concerned that they are affected by the reform. Both another identification strategy that relies on the number of years for which students were taught in Catalan within compulsory education and comparisons with the patterns of the education achievements in other Spanish regions suggest that this is not the case.

This chapter is related to a growing amount of economic literature on the identity formation process. In an influential series of papers, Akerlof and Kranton (2000, 2002, 2005) and Akerlof (2007) incorporated identity, a person's sense of self, into an economic model of behaviour. They proposed a utility function that depends on the individual's

assigned social category, on the match between the prescriptions for that category and the individual's given characteristics, and on his and others' actions. They then presented several applications of their theory in the fields of the Economics of Education, the Economics of Organizations and Macroeconomics. Previous literature in economics discussed and studied the concept of identity (Sen (1985), Folbre (1994), Kevane (1994) and Landa (1994)), but never considered it as an argument in the utility function. Several theoretical studies followed the work of Akerlof and Kranton, including that by Benabou and Tirole (2007) which endogenized identity payoffs and categorical prescriptions. However, the number of empirical works which study both the determinants of individual identity and how identity affects individual behaviour remains limited.

Our work is also connected to the literature on endogenous preferences. This literature (see Bowles 1998 for a review of this topic) studies how political and economic institutions shape preferences through their effects on social norms, cultural transmission or through other channels. If identity is part of the utility function, this study isolates a particular institutional arrangement (bilingual education) that is able to influence individual preferences.

The structure of the chapter is as follows: in Section 3.2 we provide a brief description of the reform and more generally of the institutional setting in Catalonia after 1983, then we discuss the data and the empirical strategy employed. Section 3.3 presents the basic empirical evidence and provides several robustness checks. Section 3.4 shows the results under a different identification strategy, while Section 3.5 discusses the role of individual origin. Section 3.6 provides the conclusion.

3.2 Background, data and empirical strategy

3.2.1 Background

The Catalan language had developed by the 9th century from vulgar Latin on both sides of the eastern part of the Pyrenees. Its territorial expansion went hand in hand with the expansion of the Catalano-Aragonese Crown, which was established in the "País Valencià", the Balearic Islands, the south of France and the town of Alghero (in Sardinia, Italy) over many centuries.

These are the regions where Catalan is in use nowadays with over 9 million speakers, making it the most important regional language of the European Union. Catalonia is the region that hosts most of them, with over 6 million speakers. However, to achieve this, the Catalan Government has had to protect the Catalan language, and sometimes

even promote its usage, because it was banned during Franco's dictatorship (1940-1975).

The democratization process that took place in Spain during the late seventies led to the approval of a Constitution that, within a decentralized government structure, recognized the existence of a set of autonomous communities and allowed them to have their own regional parliaments. One of the most important laws approved by the new Catalan Government was the "General Directory of Linguistic Policy (1983)" which allowed the educational system (which, prior to that date, was all in Spanish as an inheritance from Franco's period) to become bilingual. The law states that:

Art. 2: "The Certificate of Basic Educational Attainment will not be achieved without proving the proficiency of the official languages of Catalonia (Catalan and Spanish)".

Art. 6: "The Catalan language, as it is the own language of Catalonia, it is also the language used in the education system".

Art. 7.3: "The language used in the education system will never separate the alumni due to linguistic differences".

Since the Catalan Education system employed Spanish for a long time period, the transition to a bilingual system was smooth. As stated in the "Order of Application of the General Directory of Linguistic Policy (1983)", for the first 4 years of primary education the presence of Catalan had to be smoothly increased, and in no case could it impede the normal learning process of students already enrolled. For the subsequent 4 years of primary education, the presence of the Catalan language in the education system was initially introduced in social and natural science courses, and was increased in the following years. For secondary education, the Order again makes it clear that, in order to minimize its effect on the normal learning process, the increased presence of Catalan language in each school had to be done in accordance with the students' prior knowledge of Catalan.

3.2.2 Data and descriptive statistics

The empirical analysis uses representative data about the institutional attachment of the residents of Catalonia to the Spanish State that was published by the "Centro de Investigaciones Sociologicas" in 2002. This survey is performed relatively often, but the one done in 2002 is of special interest, as individuals were asked not only about their identity, but also their own origin and that of their parents. This additional piece of information turns out to be important in analyzing the effects that a change in the school language has on the identity of individuals. The reason lies in the important

inflows of immigrants from Spain that Catalonia experienced during the 60s and 70s, a fact that has to be taken into account, since identity is likely to be affected by origin.

To identify the individuals' attachment to Spain or Catalonia, we rely on the following question: "With which of the following sentences do you identify the most? (i) I feel only Spanish, (ii) I feel more Spanish than Catalan, (iii) I feel as Spanish as Catalan, (iv) I feel more Catalan than Spanish, and (v) I feel only Catalan". That is, the closer this number is to 5, the more Catalan and less Spanish you feel, and vice versa.

In Spain, primary education lasts for 8 years and starts at the age of 6, while secondary education lasts for 4 years. Therefore, the cohorts that may have been treated are those born between 1966 and 1983, while the control cohorts are those born between 1948 and 1965. The treated cohorts are those exposed, or potentially exposed to more years of teaching in Catalan, while the control cohorts are those who have not been exposed to it. Thus, an individual from the 1966 cohort who completed his/her secondary education received one year of treatment, while an individual from the same cohort, but with only primary education, had no treatment at all. Similarly, an individual from the 1970 cohort who completed his/her secondary education received 5 years of treatment, while an individual from the same cohort but with only primary education received only 1 year of treatment.

Figure 3.1 describes the relationship between cohorts, years of schooling and years of treatment. There are 4 major educational categories: (i) individuals who did not complete their primary education, who we assume received 4 years of schooling; (ii) individuals who completed only primary education and so received 8 years of education; (iii) individuals who started but did not complete their secondary education or who received some professional training, who we assume received 10 years of education; and (iv) individuals who completed secondary education or a higher level of professional training, who had 12 years of education.¹

Figure 3.2 plots the average time for which each cohort has been exposed to Catalan in schools. The average presence of Catalan in schools is obtained from the questionnaire, where individuals are asked to classify between 1 and 5 how much of their education was in Catalan and how much of it was in Spanish, 1 being only Spanish and 5 only Catalan. This graph makes it clear that there has been a huge increase over time in the use of Catalan in schools. However, the increase started slightly after the 1960s', and hence, prior the 1966, the year that we define as the first cohort to be potentially treated. This may show (as the anecdotal evidence we collected confirms) that some schools started teaching some subjects in Catalan earlier than this, at their own discretion.

¹We do not consider university education since the law we study did not affect it.

What makes the change in the linguistic policy of Catalan schools a unique case study is the following: although Spanish immigrants saw that the language of instruction had been changed to a foreign one, they reached levels of education in Catalan very similar to those received by natives, both in terms of the average number of years of education and the number of hours of education in Catalan (see Tables 3.1 and 3.2). The main reasons for this are that primary and secondary education were publicly provided, schooling was mandatory until the age of 14 for most cohorts, and the reform was uniformly implemented across the Catalan territory.

Our identification strategy relies on the variation in Catalan instruction across cohorts and years of education. An initial, simple analysis of the data shows interesting patterns in the variation of the Catalan sentiments between the treated and the control cohorts, and within each cohort (high vs. low educated). If there is a link between language policy and peoples' identity, the reform will boost the Catalan sentiment for those cohorts who were educated in Catalan. Figures 3.3 to 3.15 show that this is indeed the case. Figures 3.3 and 3.4 plot, respectively, the identity of high (with secondary education) and low (without secondary education) educated Catalans against year of birth. The vertical lines divide the control and treated cohorts. As seen from Figure 3.3, among the control cohorts, the younger and more educated cohorts are more likely to feel "as Catalan as Spanish" and less likely to feel "only Catalan": Catalan sentiments decrease with the year of birth. The reform, however, seems to have reversed this trend in Catalan sentiments. A similar process took place among respondents without secondary education. Figure 3.4 shows that, among the control cohorts, the younger ones tend to feel "only Spanish". The linguistic policy seems to have altered this tendency and this is, in part, absorbed by the increase in the fraction of low educated Catalans feeling "more Spanish than Catalan". Moreover, it is also possible to notice that prior to the reform, younger cohorts are more likely to feel "as Catalan as Spanish" and less likely to feel "more Catalan than Spanish". After the reform these trends change their sign, preventing the widening of the gap between the two.

The linguistic law is also expected to be accompanied by an increase in the within cohort variation, augmenting the Catalan sentiments of those individuals who received higher levels of treatment, that is, the more educated. To check whether this is the case, Figure 3.5 plots, for each type of answer, the difference between the fractions of high and low educated people against year of birth. It is interesting to describe the pattern of the effect of the education reform on the category of "only Catalan". While, prior to the reform, the difference was pretty stable around zero, with the reform, there is a clear change that leads the difference to increase significantly, as the gap between the treatment intensity widens. This is partly reduced when the low educated people begin to be affected by the reform and the fraction of them feeling only Catalan increases.

Figures 3.6 to 3.10 and 3.11 to 3.15 show between and within variation, respectively, in more detail.

Table 3.3 in the appendix provides the full set of descriptive statistics of the main variables used.

3.2.3 Empirical strategy

Teaching in Catalan started at the beginning of the academic year 1983-1984. Cohorts who started their primary education in or after that academic year were completely affected, while cohorts who started their primary education prior to this, but were still in primary or secondary education after 1983-1984 were only partially affected by the reform. The students' exposure to the Catalan language varies according to the number of years they were in education after the reform.

Thus, exposure to the reform varies according to year of birth and the number of years of completed education for each individual. The effect of the Catalan law is identified using a difference-in-differences approach that exploits between cohort variation in Catalan instruction and within cohort variation in years of education. The identification strategy used in this chapter is similar to that used in Angrist and Lavy (1997) and Angrist et al (2006).

The first econometric specification to be tested is:

$$y_{ijm} = \alpha + \beta_k L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (3.1)$$

Where the dependent variable y_{ijm} indicates whether individual i , from cohort j and with m years of schooling answers: (1) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan.

L_{ijm} is the length of exposure to teaching in Catalan. γ_j is the cohort effect, δ_m is a dummy for educational background and X_{ijm} is a vector of the individual-level control variables, including gender, individual's origin, parents' origin and location dummies.²

The identification strategy relies on the fact that there is no other variable that affects how some cohort-years of education groups feel about the Catalan identity issue³. We

²We exclude from the sample all the migrants who had received some education outside Catalonia.

³In another section of the paper we discuss the fact that years of education may not be exogenous to the reform.

run this specification, first as a linear regression and then as an ordered logit, due to the ordinal nature of the dependent variable⁴. Robust standard errors are clustered at the cohort-years of education level, to control for the fact that observations in a given cohort-years of education group may be correlated.

3.3 Results

3.3.1 Difference-in-Difference Estimates

Table 3.4 reports estimates for equation 3.1. In Columns 1 and 2, we show estimates in which we only control for years of education and year of birth fixed effects, for the OLS and the ordered logit specifications, respectively. OLS estimates show a positive and significant coefficient, suggesting that an increase in exposure to the reform increases the probability that an individual feels more Catalan. The ordered logit results go in the same direction and are significant. In Columns 3 and 4, we add province dummies and controls for gender, family and individual origin. The coefficients are now slightly larger than before. Even if the coefficients for an OLS regression in which the dependent variable is a discrete variable with values from 1 to 5 are difficult to interpret, the results suggest that those who received Catalan-intensive instruction tend to have significantly more intense Catalan feelings. The intensity of Catalan sentiment is, as predictable, correlated with individual origin and with the origin of the parents of the interviewee. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia. We find that Catalan identity is strongest among respondents with Catalan origins and with parents with Catalan origins and, weakest among respondents who were not born in Catalonia.

In order to interpret the ordered logit coefficients, in Figure 3.16, we plot how the probability of each answer changes with the intensity of the reform, i.e. the number of years of Catalan instruction. This corresponds to the ordered logit regression in Column 4. This figure helps us understand the size of the effect of the exposure to Catalan instruction. On the one hand, while the probability that an individual with 0 years of treatment says "I am only Catalan" is less than 0.1, the probability that a respondent with full treatment says "I am only Catalan" is slightly less than 0.5. On the other hand, the probability that an individual says "I am as Spanish as Catalan"

⁴Results for ordered probit are very similar.

appears to be strongly decreasing with years of treatment (from more than 0.5 with 0 treatment to just more than 0.1 with 12 years of treatment). The probability of choosing "Only Spanish" and "more Spanish than Catalan" decreases slightly and the probability of choosing "More Catalan than Spanish" shows a small increase with the length of exposure to the reform. Thus, the results suggest that, as a consequence of the reform, individuals are less likely to answer that they feel "As Spanish as Catalan" and more likely to say that they feel "Only Catalan".

3.3.2 Effect for each level of treatment

The marginal effect of the treatment can be very different for each level of treatment. We wish to investigate how the effect increases with years of treatment. As a result, and similarly to Angrist et al. (2006), we decided to introduce a dummy for each possible level of treatment and run the following specification:

$$y_{ijm} = \alpha + \sum_{k=1}^{12} \beta_k I(L_{jkm} = k) + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (3.2)$$

Where $I()$ is the indicator function. This specification is run both as an OLS and as an ordered logit model.

Columns 1 and 3 in Table 3.5 report ordinary least square estimates of Equation (3.2). In Column 1, only dummies for years of education and cohort are included. In Column 3, individual controls and province dummies are added to the regressions. The results in both columns go in the same direction. Identity is affected by the reform; however, treatment effects do not increase monotonically with years of treatment, even though the effect seems to be stronger after 5 years of treatment. Ordered logit estimates for the same two specifications are provided in Columns 2 and 4.

Figure 3.17 confirms the patterns proposed in the previous section. In Figure 3.17, we plot how the effect the reform changes with the intensity of the reform, i.e. the number of years of Catalan instruction. This corresponds to the ordered logit regression in Column 4. The probability of answering "I am only Catalan" increases with the years of treatment, while the probability of answering "I am as Spanish as Catalan" appears to strongly decrease with years of treatment. The effect of the reform on the probability of choosing the other 3 categories (only Spanish/ more Spanish than Catalan/ more Catalan than Spanish) does not seem to be very relevant, as the lines fluctuate around zero. The effect seems to be stronger during the first eight years of treatment than later on.

3.3.3 Channels

We now investigate whether language is one of the channels through which the reform affects individual identity. In order to check the validity of our hypothesis, we take advantage of several questions asked during the survey.

As already mentioned previously, individuals were asked to classify how much of their education was in Catalan and how much in Spanish. We then create an index from 0 to 1, with 0 corresponding to education only in Spanish and 1 to education only in Catalan. In order to have a measure of the length of exposure of each respondent to Catalan teaching, we build a variable "Language exposure" that is the product between the index previously described and the years of education.

In the survey, there are also several questions that provide information about the use of the Catalan language in everyday life: 1) when the interviewee meets friends, 2) when he is at home, 3) when he goes shopping, 4) if he is asked for directions in the street, 5) when he answers the phone, 6) when he interacts with civil servants 7) or when he is with the colleagues at work. We then create an index of the "social use" of Catalan, that goes from 0 to 7, where 7 indicates that the interviewee uses the Catalan language in all of these circumstances. As a first step, in the first and second Columns of Table 3.6 we check whether the two indices are related to identity. As expected, there is a positive and strong relationship between the two: the greater the use of the Catalan language at school, the more intense the Catalan feelings and, at the same time, the greater the use of Catalan in everyday life, the more intense are the Catalan feelings. As a second step, in Columns 7 and 8, we run a 2SLS in which we regress the two indices on our measure of the intensity of the reform and then we regress our measure of identity on the predicted value obtained in the first stage. The coefficients are both positive and significant at the standard level, meaning that language was truly an important channel through which the reform affected individual feelings.

The fact that the 2SLS results are larger than the OLS ones indicates that, if there is an omitted variable that affects an individual identity and the self-reported measure of Catalan at school, this will be negatively correlated with identity and positively correlated with the self-reported measure of Catalan at school (or vice versa). A possible explanation for this is that people who say they feel more Catalan may be more likely to perceive the amount of education they received in Catalan as being less or insufficient, while people who feel more Spanish tend to perceive the amount of education they received in Catalan as excessive. There is no question in the survey that allows us to test this, but, since our self-reported measure of schooling in Catalan is based on recall, this remains a possibility.

3.3.4 Migration patterns

Migration patterns in and out of Catalonia changed the composition of the Catalan population. As table 3.3 shows, the fraction of respondents whose parents were both born in Catalonia is larger among control cohorts than among treated cohorts. In general, the family characteristics of a respondent belonging to the control cohorts are different from the family characteristics of a respondent belonging to the treated (or potentially treated) cohorts. Although in the main specification we control for the type of family of the respondents, migration patterns, affecting the composition of the cohort educational groups, might have biased our results. We might be capturing a change in Catalan society's composition rather than the effect of the reform.

In order to control for this possible source of bias, we add a new set of controls in the regressions. As already mentioned, we classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia. We then calculate, within each cohort-educational group, the fraction of respondents belonging to each of the 4 categories. We adopt the following econometric specification

$$y_{ijm} = \alpha + \beta_k L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + s_{jm} + \varepsilon_{ijm} \quad (3.3)$$

Where the dependent variable y_{ijm} indicates, as usual, whether individual i , from cohort j and with m years of schooling answers (1) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan. s_{jm} is a vector of 4 explanatory variables that control for the composition of each cohort educational group jm . Each variable measures the size, within the cohort educational group jm , of one of the 4 family groups.

It turns out, as columns 1-2 of Table 3.7 show, that β_k , the coefficient of our variable of interest, is still positive and significant and the size of the coefficient is very similar to the size of the coefficient in our baseline specification.

It could also be the case that the reform changed migration patterns into and out of Catalonia. Schooling in Catalan could have implied an additional cost of migration to Catalonia for Spanish speakers. Potential migrants with very intense Spanish sentiments could decide not to migrate to Catalonia and, vice-versa, the additional migration cost could have been lower for open minded migrants.

Since we do not know the year in which the parents of the interviewee migrated, we cannot exclude from the sample individuals who are potentially affected by this problem (with parents who migrated as a consequence of the new law). We then decide to check if our results hold only for individuals born in Catalonia. Since the youngest cohorts were born in 1983, we are then restricting the sample to individuals whose parents were already in Catalonia when the reform was implemented, so the parents' decision to migrate to Catalonia is less likely to be affected by the linguistic reform. Columns 3 and 4 of Table 3.7 show that the results are robust to these further checks.

However, it may still be the case that parents somehow anticipated the reform that was taking place in 1983 and then took the migration decision before 1983. To avoid this anticipation problem, we restrict the sample even further and consider only cohorts born in Catalonia between 1953 and 1978. Columns 5 and 6 of Table 3.7 show that the results are robust to this check.

Table 3.7 should be sufficient to convince us that the results we find are not related to a change in the pattern of migration inflows in Catalonia. However, it can still be the case that patterns of migration outflows are affected by the reform and, at the same time, lie at the root of the results we found in Tables 3.4 and 3.5. Unfortunately, the survey data we have does not allow for any checks that would help us in this direction. Using data from INE, in Figure 3.18 we compare the patterns of migration in Catalonia over the last 40 years with the patterns of migration of another Spanish region that is comparable in terms of income, i.e. Madrid. We plot net migration flows as a fraction of total flow. It turns out there is no substantial difference, either before or after the reform.

3.3.5 Restricting the sample

In our first specification, we were comparing individuals born in 1948 with those born in 1983. While this allows us to use the maximum data available, it may also introduce some noise. In order to mitigate this, we perform the same regression but compare individuals aged 26 (cohort 1975) with those aged 45 (cohort 1956). As the cohorts we are comparing are more similar, the results should be more reliable. Table 3.8 shows that, when we restrict the sample and compare cohorts of similar ages, the results improve. This may be because we are comparing cohorts who are more similar to each other and who have been affected by similar shocks. This allows us to increase the precision of our estimates and to capture better the effects of the reform.

3.3.6 Exogenous Trends

The estimates in Tables 3.4 and 3.5, which point to higher Catalan feelings among those who had been exposed to more intensive Catalan instruction, may be confounded by education-specific cohort trends in Catalan feelings. In particular, these trends may exist if different cohorts have been raised by parents with different values and preferences (the younger parents being more pro-Catalan), and if those parents with stronger Catalan preferences have invested more in their children's education.

In order to check whether this is the case, and similarly to Angrist (2006), we propose the following two exercises. As a first step, we consider only cohorts who are not affected by the reform. These are cohorts born between 1930 and 1965. Then, among those cohorts, we assigned a pseudo-treatment to the younger cohorts (those born between 1948 and 1965 whom we used as controls in the main specification). This pseudo-treatment consists of the intensity of the reform received by individuals with the same number of years of education but born 18 years later. We run the specification proposed by Equation 3.1. We use both an ordinary least squares specification and an ordered logit one. The results are reported in Columns 1 and 2 of Table 3.9. The coefficient of the pseudo treatment variable is negative and not significant. This is reassuring, as it may indicate that our results are not spuriously capturing the effect of a trend.

As a second check, we consider cohorts born between 1921 and 1980 and use the following specification⁵. To correct the difference-in-difference-type estimates for the education-cohort trends, we adopt a triple differences-type strategy. Triple difference estimates are based on the following equation,

$$y_{ijm} = \alpha + \beta L_{ijm} + \theta l_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (3.4)$$

where L_{ijm} is a term that captures real exposure to teaching in Catalan, while l_{ijm} captures spurious effects.

L_{ijm} are the years of real treatment for cohorts 1966-1980; while l_{ijm} equals years of pseudo treatment for cohorts 1936-1950 (where now the pseudo-treatment consists of the intensity of the reform received by individuals with the same number of years of education but born 30 years later) and years of real treatment for cohorts 1966-1980. β are the triple difference estimates, i.e. the treatment effects from the real experiment minus the pseudo treatment effects estimated using cohorts 1921-1950. Columns 3 and 4 report the results for this specification. Triple difference estimates are positive and

⁵Due to lack of data about very old cohorts, we could not use a larger number of cohorts, such as, for example, from 1909 to 1983.

significant, while the coefficient that would be capturing spurious trends is negative and not significant.

Finally we try to establish whether or not Franco's death (1975) affected the identity of younger and more educated cohorts. We perform the following exercise. We create a variable called "Franco Intensity" that measures the number of years for which individuals attended school once Franco had died but before the educational system became bilingual. Then we run Specification 3.1 with our intensity measure and this additional variable. Results are shown in Columns 5 and 6 and show how our measure of intensity remains positive and significant, and that the coefficient is even slightly higher. The coefficient of the Franco intensity variable is, however, positive too and significant in the OLS specification. This suggests that the end of Franco's regime may have had an effect on the feelings of the younger and more educated cohorts, but it may also simply reveal that (as shown in Figure 3.2) some schools started teaching some subjects in Catalan earlier than 1983.

As a further test, we consider an other Spanish region: the Basque Country. In 1983 the education system became bilingual in the Basque region too. However, whereas in Catalonia the reform was compulsory and homogeneously spread across the Catalan population, in the Basque Country parents could choose the language used to educate their children. The Basque and the Catalan languages are also very different: Basque is a non-Indo European language and it presents a completely different structure for native Spanish speakers. Moreover, the salience of the language issue is quite different in the two regions. While language is the main vehicle of identification within the Catalan group, this may not be the case within the Basque group. As a result, we are likely to find that the reform in the Basque Country had no effect on the identity of residents (or at least lower than in Catalonia). However, since the death of Franco is likely to have affected residents of Catalonia and Basque Country in the same way and education-specific cohort trends are likely to follow similar patterns in both the regions, if our results were driven by such factors, the coefficient of the variable of interest should be positive and significant also using the Basque sample. We use representative survey data on the institutional attachment of residents in the Basque Country to the Spanish State. This survey was also published by CIS in 2001 and is very similar to the survey conducted in Catalonia. The only difference is that the Basque survey was conducted one month before the Catalan one. When we apply the same empirical strategy to the Basque sample we find that the coefficient of the variable of interest (years of treatment) is not significantly different from zero (Columns 7-8). Thus, the comparison between Catalonia and Basque Country helps us to exclude the possibility that the findings discussed in the previous sections are driven either by the existence of education-specific cohort trends or by the end of the Franco's dictatorship.

3.4 An Alternative Identification Strategy

The identification strategy previously described relies both on between cohort variation (older cohorts have not been affected by the reform, while the younger ones have been) and on within cohort variation (individuals of the same age may have different levels of education and so may have been exposed to different levels of treatment). However, the number of years of schooling is typically an individual choice (or a parental one) and may be affected by the introduction of the linguistic reform. One may imagine that individuals with very intense Spanish (and anti-Catalan) feelings, for example a migrant from another region of Spain, may have decided to drop out of school because of the new scheme and because they may feel less comfortable about receiving education in the Catalan language. Individuals who have more difficulties with the Catalan language may drop out of school earlier as a consequence of the reform, and these are the ones who are more likely to feel Spanish.

As an initial check to confirm whether or not this is the case, we add as a control into the regression a dummy variable that is equal to 1 if the individual spoke Catalan at home with his parents. An individual who normally speaks Catalan would have experienced less difficulties learning Catalan at school and was less likely to drop out of education after the reform. The results of this specification are shown in Columns 1 and 2 of Table 3.11 and are very similar to those obtained previously. This is reassuring, as the language spoken at home could be the main source of endogeneity.

Secondly, we use a specification that relies mostly on between cohorts variation and we eliminate cohort-fixed effects. The econometric specification to be tested is now

$$y_{im} = \alpha + \beta_k L_{i,m} + \delta_m + X_{i,m}\mu + \varepsilon_{im} \quad (3.5)$$

where the dependent variable $y_{i,m}$ indicates whether individual i with m years of schooling answers: (1) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan. $L_{i,m}$ is the length of exposure to teaching in Catalan, δ_m is a dummy for education background and $X_{i,m}$ is a vector of individual-level control variables, including the individual's gender, origin, parents' origin and location dummies. There are no more cohort dummies, but we also include as controls the age and the squared age of the respondents.

The results of this specification are shown in Columns 3 and 4 of Table 3.11 and are similar to those obtained previously, although the coefficient is slightly smaller.

As a third step, in order to avoid considering any within cohort variation, we instrument $L_{i,m}$ with a measure of exposure to the reform that is not the result of individual choice: the number of years of compulsory education for each individual. In this way we remove any within cohort variation. Each respondent belonging to the same cohort will be subject to the same amount of exposure to compulsory education. This should provide an exogenous measure of years of education, but it should also increase the measurement error and cause attenuation bias in the coefficient of our variable of interest.

The results are reported in Tables 3.10 and 3.11. As usual, we always use both the linear regression model and the ordered logit one. As predictable, the first stage of the IV analysis (Table 3.10) shows a strong correlation between the variable of interest and its instrument. Column 5 reports the second stage of the instrumental variables strategy. Finally, Columns 6 and 7 report the results when we use the reduced form (this allows us to use an ordered logit model too). The results are consistent with the previous findings.

In addition, instead of using years of education as intra-cohort variation in years of treatment, we use variation in the timing of arrival as a determinant of more or less exposure to the electoral reform⁶. In order to perform this exercise, we restrict the sample to individuals who were not born in Catalonia. These individuals arrived at different points in time and at different stages during their education. For each individual we compute the number of years, within compulsory education, for which they have studied in Catalonia after 1983. This is now our intensity variable; we run the following specification:

$$y_{ijm} = \alpha + \beta L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (3.6)$$

Where the dependent variable y_{ijm} indicates whether individual i , from cohort j who arrived in Catalonia when he was m years old answers: (1) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan.

L_{ijm} is the length of exposure to teaching in Catalan. γ_j are the cohort dummies, δ_m are age at arrival dummies and X_{ijm} is a vector of individual-level control variables, including gender, origin and years of education dummies.

The results are shown in Column 8 of Table 3.11 and are very similar to those obtained when the within-cohort variation used was given by differences in the number of years

⁶For this purpose, we include in the sample also migrants who received some education outside Catalonia and who were not included in any of the previous specifications.

of education. Now we do not have enough observations to run an ordered logit but the fact that the OLS results are very similar to those obtained previously is reassuring. This shows that our results are not driven by changes in years of education induced by the reform.

In any case, years of education are unlikely to have changed endogenously as a consequence of the reform. This is supported by the fact that schools provided facilities and extra time for students for whom Catalan was not their mother tongue. As a final check, we plot average years of education by cohort for Catalonia, the Madrid region (a region which have always been at the same level of development as Catalonia) and the rest of Spain. This is done using data from the Spanish Census in 2001. The graph (Figure 3.19) shows that years of education for cohorts affected by the reform (those born after 1965) evolved quite similarly in Catalonia and in Madrid and did not decline for the Catalan cohorts affected by the reform. Years of education increased more for the other communities in Spain, as a result of a process of convergence.

3.5 Heterogeneous effects

The introduction of a Catalan-based schooling system (even if mitigated by the rule that at least two subjects per year must be taught in Spanish) can also be interpreted as an example of a nation building policy within the Catalan region. It is interesting then to check whether the reform has had an impact on the feelings of minorities living in Catalonia. We define as part of a minority group all of the respondents of non-Catalan origin or with parents without Catalan origin (we also distinguish between individuals from families with at least one parent born in Catalonia and those with neither parents born in Catalonia).

Columns 1-2 of Table 3.12 show that the reform had a positive and uniform impact on the whole population; the effect is positive not only on respondents of Catalan origin and whose parents were both born in Catalonia, but also on individuals born in Catalonia and from families with neither parents born in Catalonia or with only one parent of Catalan origin. Moreover, the coefficient of the interaction term between the variable that captures the length of exposure to teaching in Catalan and a dummy variable different from zero if the respondent was born outside Catalonia is only marginally insignificant in Column 2.⁷ Overall, the results convey evidence that the new schooling system successfully increased the level of integration of minorities within Catalonia.

⁷We also use other specifications: without cohort dummies, and with the instrumentation previously explained. All provide similar results.

3.6 Summary and conclusions

Much has been said about the possibility that policies can affect individual identity, sentiments and preferences. However, to date, little case study research has been conducted on this matter. This chapter presents evidence of the positive effect of the educational reform (by which the education system became bilingual, and Catalan - together with Spanish- was taught in schools) on Catalan feelings. This effect extends also to individuals whose parents have no Catalan origins.

The Catalan case allows us also to compare multicultural policies and nation building policies. We interpret the reform as an example of multicultural policies within Spain and of nation building policies within Catalonia. We conclude that, while multicultural policies stimulate regional identities, nation building policies encourage the growth of a shared national sentiment.

Research of this sort together with the study of the effects of such reforms on the political and economic outcomes can be highly relevant in ethnically divided societies where policies (linguistic and not) can be seen as mechanisms of integration and conflict reduction.

3.7 Appendix

3.7.1 Definition of the variables

-Identity: ordered variable which assumes the following values: (1) if the respondent answered "I feel only Spanish" (2) if the respondent answered "I feel more Spanish than Catalan" (3) if the respondent answered "I feel as Spanish as Catalan" (4) if the respondent answered "I feel more Catalan than Spanish" and (5) if the respondent answered "I feel only Catalan". Source: CIS

-Years of education: (i) if the respondent answered that he did not complete primary education, we assume that he received 4 years of schooling; (ii) if the respondent answered that he completed only primary education, we assume that he received 8 years of education; (iii) if the respondent answered that he started but did not complete secondary education or that he received some professional training, we assume that he received 10 years of education; and (iv) if the respondent answered that he completed secondary education or a higher level of professional training, we assume that he received 12 years of education. Source: CIS

-Catalan origin, Catalan family: dummy equal to 1 if the respondent answered that he was born in Catalonia and both his parents were born in Catalonia. Source: CIS

-Catalan origin, mixed family: dummy equal to 1 if the respondent answered that he was born in Catalonia and only one of his parents were born in Catalonia. Source: CIS

-Catalan origin, non Catalan family: dummy equal to 1 if the respondent answered that he was born in Catalonia and neither of his parents were born in Catalonia. Source: CIS

-non Catalan origin: dummy equal to 1 if the respondent answered that he was not born in Catalonia. Source: CIS

-Catalan at school: ordered variable which assumes the following values: (1) if the respondent answered that his education was only in Spanish (2) if the respondent answered that his education was more in Spanish than in Catalan (3) if the respondent answered that his education was half in Spanish half in Catalan (4) if the respondent answered that his education was more in Catalan than in Spanish (5) if the respondent answered that his education was only in Catalan. Source: CIS

-Language exposure: Using the variable "Catalan at school", we create an index that goes from 0 to 1, with 0 corresponding to education only in Spanish and 1 to education only in Catalan. "Language exposure" is the product between this index and years of education.

-Catalan social use: respondents were asked to provide information about the use of Catalan language 1) when they meet friends, 2) when they are at home, 3) when they go shopping, 4) if they are asked for directions in the street, 5) when they answer the phone, 6) when they interact with civic servants 7) or when they are with colleagues at work. We then create an index of the "social use" of Catalan, that goes from 0 to 7, where 7 indicates that the interviewee uses the Catalan language in all of these circumstances.

-Language at home: dummy equal to 1 if the individual spoke Catalan at home with his parents

-Female: dummy equal to 1 if the respondent is female

3.8 Figures and Tables

Treatment by Cohort and Education

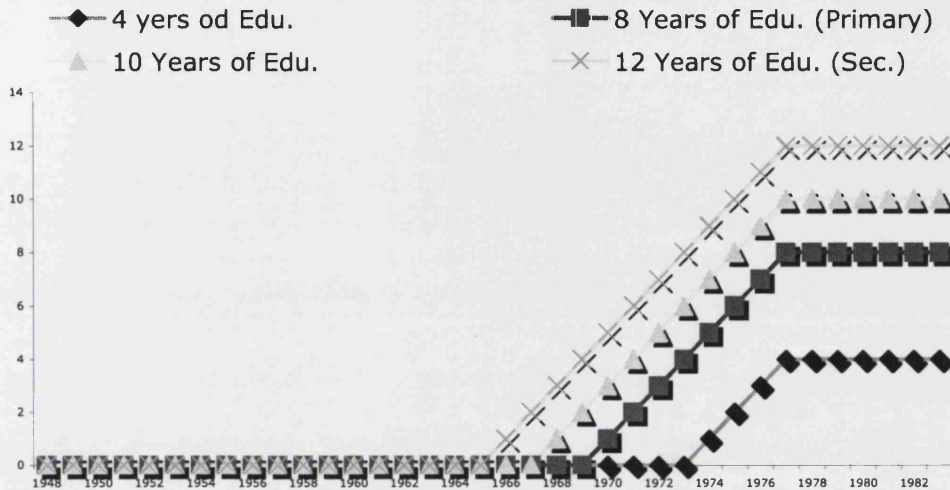


Figure 3.1: Treatment by Cohort and Education

Presence of the Catalan language at school

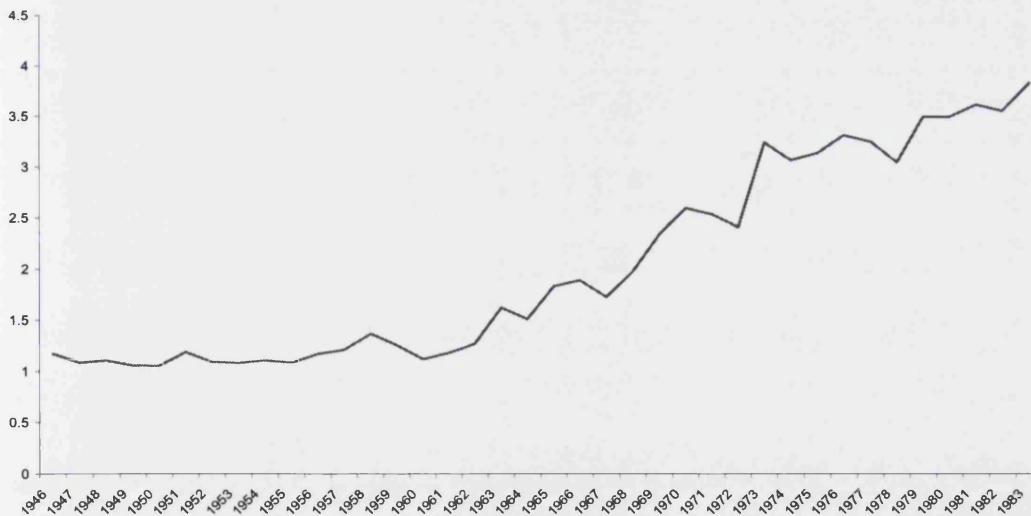


Figure 3.2: Catalan at School

Note: We plot the average time for which each cohort has been exposed to Catalan in schools. The average presence of Catalan in schools is obtained from the questionnaire, where individuals are asked to classify between 1 and 5 how much of their education was in Catalan and how much of it was in Spanish, 1 being only Spanish and 5 only Catalan.

Table 3.1: Average years of education by type of family

		Total Sample (1948-83)	Control Cohorts (1948-65)	Treated Cohorts (1966-83)
Total	Average	10.22	9.81	10.53
	St. Dev.	1.90	1.94	1.82
	Observations	1490.00	626.00	864.00
Catalan Origin	Average	10.30	9.92	10.56
	St. Dev.	1.87	1.89	1.81
	Observations	1354.00	547.00	807.00
Cat. Origin with mixed Families	Average	10.34	9.90	10.55
	St. Dev.	1.81	1.86	1.76
	Observations	246.00	79.00	167.00
Cat. Origin and No Cat. Family	Average	10.04	9.62	10.23
	St. Dev.	1.93	1.89	1.92
	Observations	343.00	104.00	239.00
No Catalan Origin	Average	9.50	9.09	10.07
	St. Dev.	2.10	2.16	1.88
	Observations	136.00	79.00	57.00

Note: We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia.

Table 3.2: Catalan at school by type of family

		Total Sample (1948-83)	Control Cohorts (1948-65)	Treated Cohorts (1966-83)
Total	Average	2.72	1.28	2.99
	St. Dev.	1.45	0.79	1.40
	Observations	1502.00	633.00	869.00
Catalan Origin	Average	2.34	1.30	3.04
	St. Dev.	1.46	0.83	1.38
	Observations	1364.00	552.00	812.00
Cat. Origin with mixed Families	Average	2.51	1.20	3.12
	St. Dev.	1.49	0.72	1.37
	Observations	249.00	79.00	170.00
Cat. Origin and No Cat. Family	Average	2.18	1.29	2.58
	St. Dev.	1.28	0.75	1.28
	Observations	346.00	106.00	240.00
No Catalan Origin	Average	1.62	1.16	2.28
	St. Dev.	1.18	0.51	1.52
	Observations	138.00	81.00	57.00

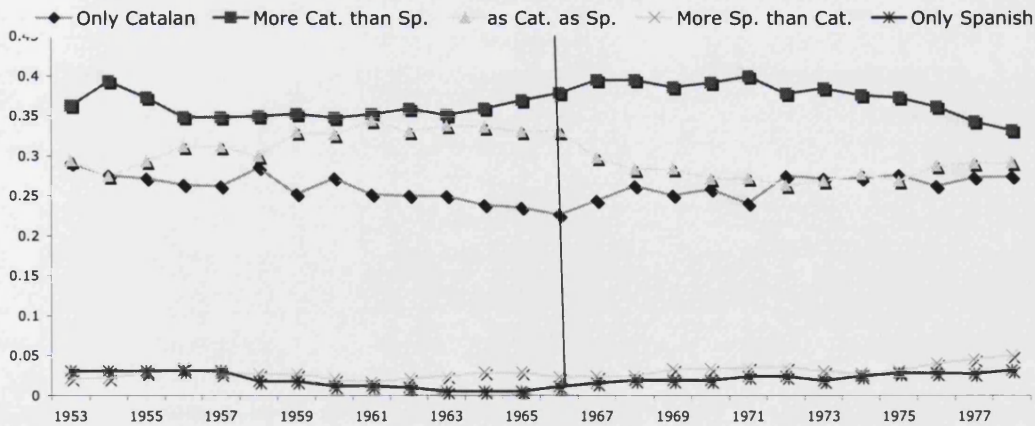
Table (2)

Note: We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia. The variable *Catalan at school* is obtained from the questionnaire, where individuals are asked to classify between 1 and 5 how much of their education was in Catalan and how much of it was in Spanish, 1 being only Spanish and 5 only Catalan.

Table 3.3: Descriptive Statistics

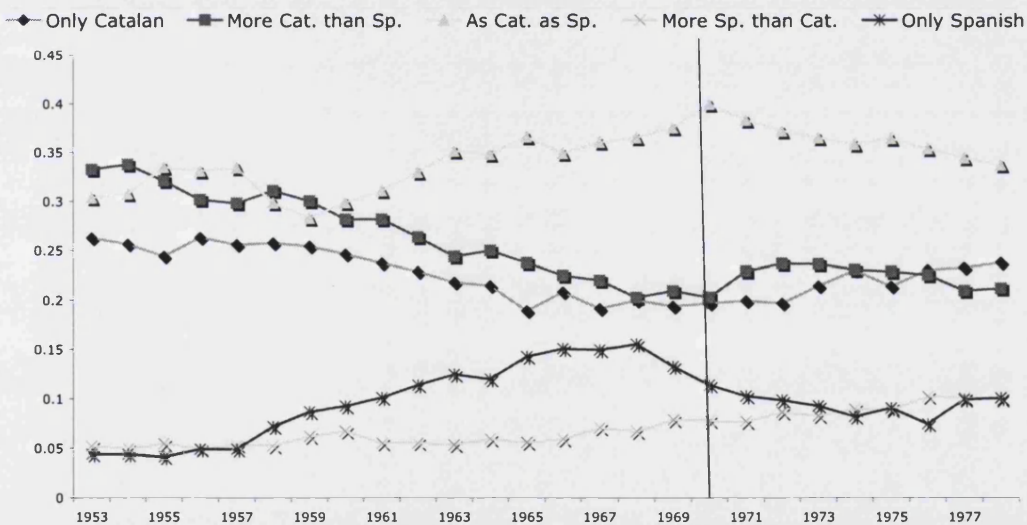
		total sample			with sec. edu			without sec. edu		
		1948-83	1948-65	1966-83	1948-83	1948-65	1966-83	1948-83	1948-65	1966-83
Identity	Average	3.51	3.56	3.47	3.7	3.73	3.69	3.29	3.42	3.16
	St. Dev.	1.04	1.03	1.05	0.91	0.9	0.92	1.12	1.1	1.13
	Observatio	1467	618	849	732	250	482	735	368	367
Years of Education	Average	10.33	9.92	10.61	12	12	12	8.45	8.29	8.61
	St. Dev.	1.88	1.97	1.77	0	0	0	0.97	0.96	0.95
	Observatio	1490	626	864	749	255	494	741	371	370
Cat. Origin, Cat. Family	Average	0.43	0.48	0.40	0.49	0.56	0.45	0.37	0.42	0.32
	St. Dev.	0.49	0.50	0.49	0.5	0.49	0.49	0.48	0.49	0.46
	Observatio	1490	626	864	749	255	494	741	371	370
Cat. Origin, mixed Family	Average	0.17	0.14	0.19	0.17	0.14	0.19	0.16	0.14	0.18
	St. Dev.	0.37	0.35	0.39	0.38	0.35	0.39	0.37	0.35	0.39
	Observatio	1490	626	864	749	255	494	741	371	370
Cat. Origin, no Cat. Family	Average	0.28	0.20	0.33	0.24	0.15	0.28	0.32	0.23	0.39
	St. Dev.	0.44	0.40	0.47	0.42	0.36	0.45	0.46	0.42	0.49
	Observatio	1490	626	864	749	255	494	741	371	370
No Catalan Origin	Average	0.11	0.16	0.07	0.08	0.13	0.06	0.13	0.19	0.08
	St. Dev.	0.31	0.37	0.26	0.28	0.33	0.24	0.34	0.39	0.28
	Observatio	1490	626	864	749	255	494	741	371	370
Catalan at school	Average	2.15	1.23	2.78	2.44	1.34	3.01	1.82	1.15	2.46
	St. Dev.	1.35	0.71	1.33	1.39	0.85	1.27	1.23	0.57	1.35
	Observatio	1487	625	862	747	255	492	740	370	370
Language exposure	Average	3.13	0.63	4.86	4.33	1.03	6.04	1.79	0.32	3.17
	St. Dev.	3.78	1.98	3.77	4.18	2.55	3.81	2.72	1.28	2.98
	Observatio	1487	625	862	747	255	492	740	370	370
Catalan Social use	Average	4.12	4.45	3.91	4.81	5.13	4.66	3.3	3.87	2.81
	St. Dev.	2.70	2.68	2.72	2.34	2.22	2.39	2.86	2.82	2.8
	Observatio	1340	534	806	702	230	472	638	304	334
Language at home	Average	0.44	0.49	0.40	0.5	0.56	0.47	0.37	0.44	0.29
	St. Dev.	0.49	0.50	0.49	0.5	0.49	0.49	0.48	0.49	0.45
	Observatio	1480	622	858	743	252	491	737	370	367
Intensity of treatment	Average	4.02	0.00	6.78	5.37	0	8.12	2.49	0	4.83
	St. Dev.	4.63	0.00	4.18	4.99	0	3.91	3.63	0	3.78
	Observatio	1490	626	864	749	255	494	741	371	370
Intensity of (comp.) treat.	Average	2.68	0.00	4.52	3.02	0	4.57	2.29	0	4.45
	St. Dev.	3.51	0.00	3.53	3.53	0	3.43	3.45	0	3.67
	Observatio	1490	626	864	749	255	494	741	371	370
Female	Average	0.49	0.49	0.49	0.49	0.43	0.51	0.49	0.53	0.45
	St. Dev.	0.50	0.50	0.50	0.5	0.49	0.5	0.5	0.49	0.49
	Observatio	1490	626	864	749	255	494	741	371	370

Figure 3.3: Identity of Catalans with Secondary Education by year of birth



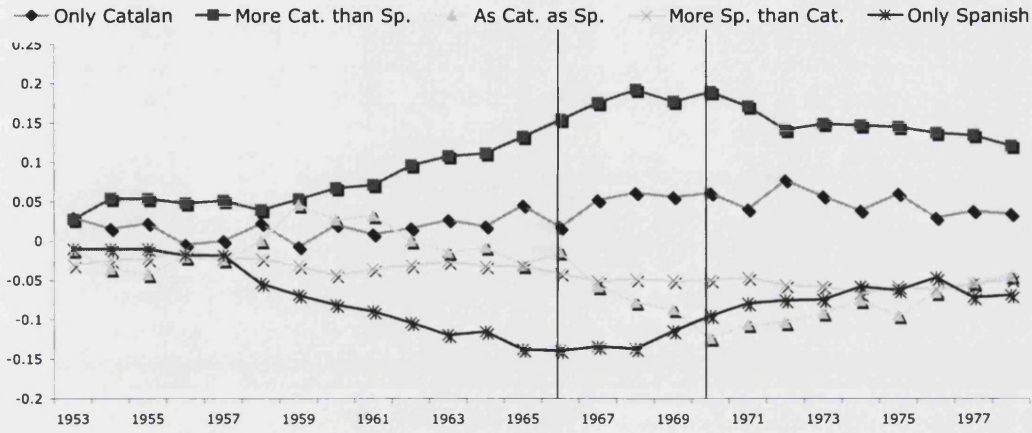
Note: When they are asked to define their identity, respondents can choose among 5 categories: (1) I feel only Spanish, (2) I feel more Spanish than Catalan, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan. The graph plots the fraction of respondents with secondary education who choose each of the 5 categories against year of birth. All series are (5,1,5) moving averages.

Figure 3.4: Identity of Catalans without Secondary Education by year of birth



Note: When they are asked to define their identity, respondents can choose among 5 categories: (1) I feel only Spanish, (2) I feel more Spanish than Catalan, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan. The graph plots the fraction of respondents without secondary education who choose each of the 5 categories against year of birth. All series are (5,1,5) moving averages.

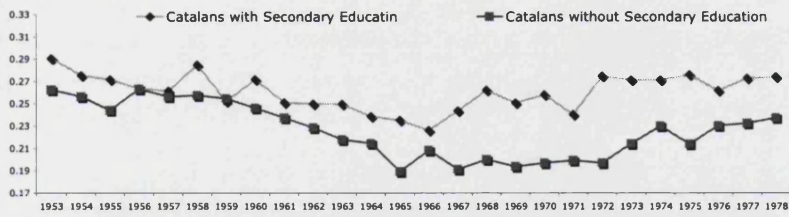
Figure 3.5: The effect of Education on the Identity



All series are (5,1,5) moving averages.

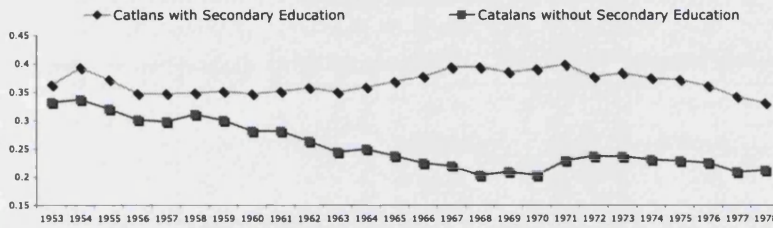
Note: When they are asked to define their identity, respondents can choose among 5 categories: (1) I feel only Spanish, (2) I feel more Spanish than Catalan, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan. The graph plots, for each of the 5 categories, the difference between the fraction of respondents with secondary education who choose category X and the fraction of respondents without secondary education who choose category X against year of birth. All series are (5,1,5) moving averages.

Figure 3.6: Fraction of respondents who feel only Catalan by cohort and level of Education



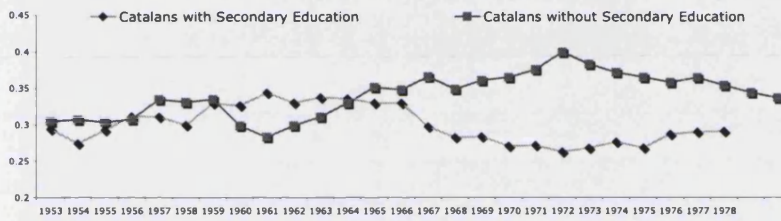
Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel only Catalan* against year of birth. All series are (5,1,5) moving averages.

Figure 3.7: Fraction of respondents who feel more Catalan than Spanish by cohort and level of Education



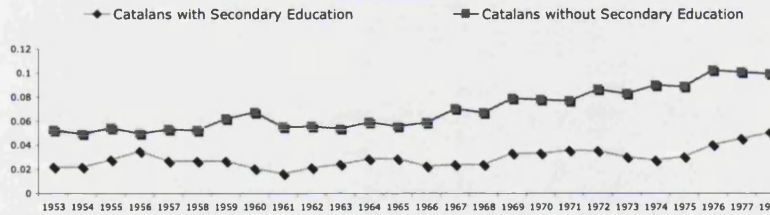
Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel more Catalan than Spanish* against year of birth. All series are (5,1,5) moving averages.

Figure 3.8: Fraction of respondents who feel as Catalan as Spanish by cohort and level of Education



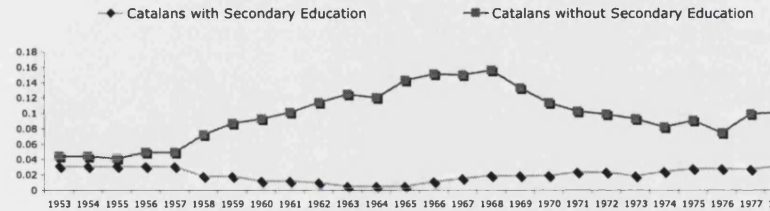
Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel as Spanish as Catalan* against year of birth. All series are (5,1,5) moving averages.

Figure 3.9: Fraction of respondents who feel more Spanish than Catalan by cohort and level of Education



Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel more Spanish than Catalan* against year of birth. All series are (5,1,5) moving averages.

Figure 3.10: Fraction of respondents who feel only Spanish by cohort and level of Education



Note: The graph plots the fraction of respondents with and without secondary education who choose *I feel only Spanish* against year of birth. All series are (5,1,5) moving averages.

Figure 3.11: The effect of Education on feeling only Catalan by year of birth



Note: The graph plots the difference between the fraction of respondents with secondary education who choose *I feel only Catalan* and the fraction of respondents without secondary education who choose *I feel only Catalan* against year of birth. All series are (5,1,5) moving averages.

Figure 3.12: The effect of Education on feeling more Catalan than Spanish by year of birth



Note: The graph plots the difference between the fraction of respondents with secondary education who choose *I feel more Catalan than Spanish* and the fraction of respondents without secondary education who choose *I feel more Catalan than Spanish* against year of birth. All series are (5,1,5) moving averages.

Figure 3.13: The effect of Education on feeling as Catalan as Spanish by year of birth



Note: The graph plots the difference between the fraction of respondents with secondary education who choose *I feel as Spanish as Catalan* and the fraction of respondents without secondary education who choose *I feel as Spanish as Catalan* against year of birth. All series are (5,1,5) moving averages.

Figure 3.14: The effect of Education on feeling more Spanish than Catalan by year of birth



Note: The graph plots the difference between the fraction of respondents with secondary education who choose *I feel more Spanish than Catalan* and the fraction of respondents without secondary education who choose *I feel more Spanish than Catalan* against year of birth. All series are (5,1,5) moving averages.

Figure 3.15: The effect of Education on feeling only Spanish by year of birth



Note: The graph plots the difference between the fraction of respondents with secondary education who choose *I feel only Spanish than Catalan* and the fraction of respondents without secondary education who choose *I feel only Spanish* against year of birth. All series are (5,1,5) moving averages.

Table 3.4: Reform and Identity

Dependent variable: Identity

	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]
intensity	.067** (.029)	.147*** (.055)	.082*** (.026)	.183*** (.058)
non catalan origin			-1.335*** (.1)	-3.038*** (.247)
mixed family			-.493*** (.073)	-1.074*** (.16)
non catalan family			-.958*** (.072)	-2.187*** (.163)
YEARS OF EDUCATION	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES
PROVINCES	NO	NO	YES	YES
No. obs.	1467	1467	1467	1467
Rsq	.088		.324	
Pseudo-Rsq		.032		.149

Note: The specifications reported in Columns 1 and 2 include years of education and year of birth fixed effects. The specifications reported in Columns 3 and 4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

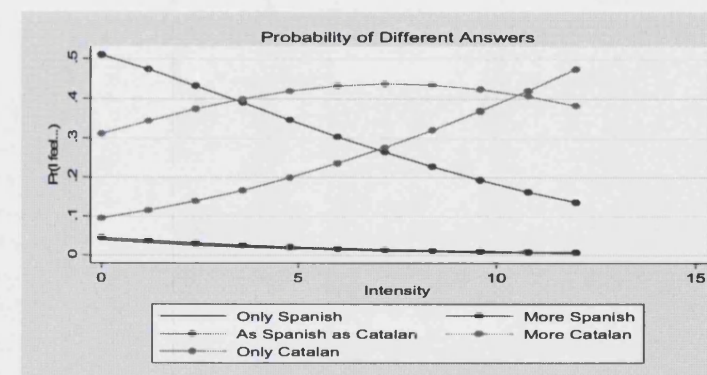
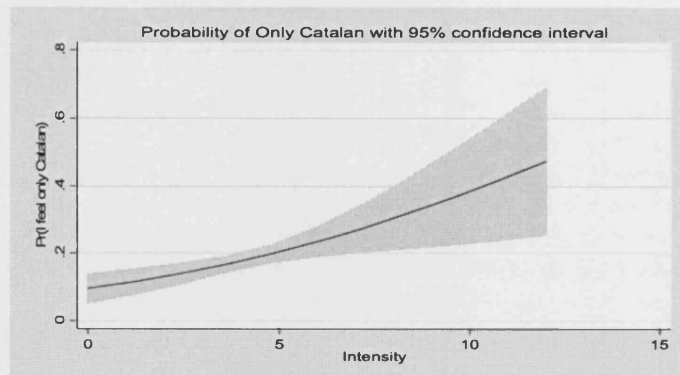
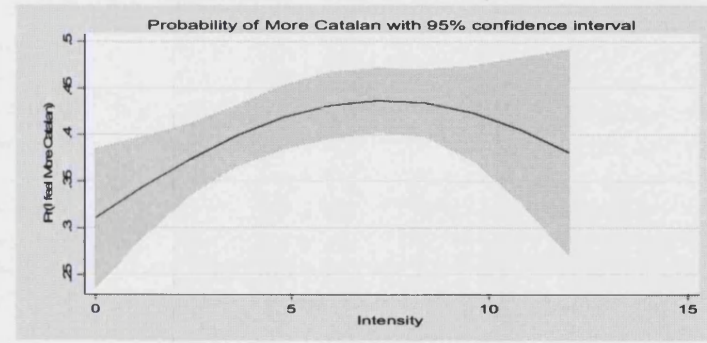
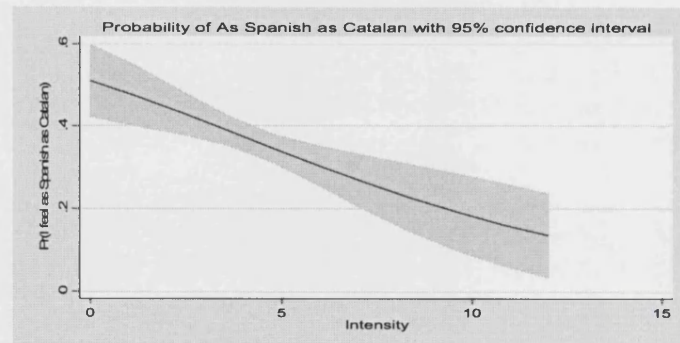
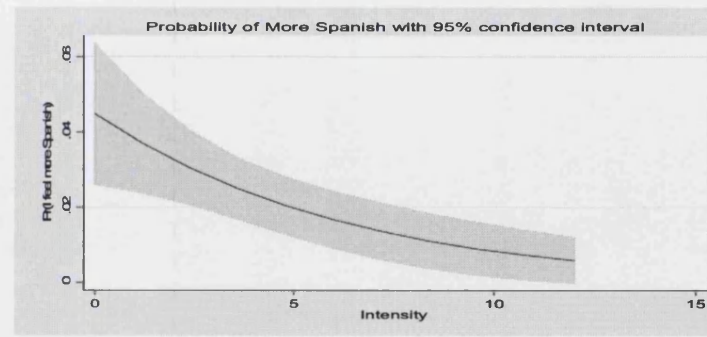
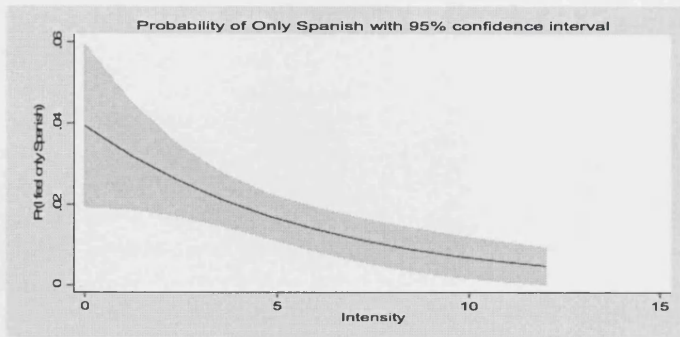


Figure 3.16: Change in probabilities (specification 3.1)

Note: we plot how the probability of each answer changes with the intensity of the reform, i.e. the number of years of Catalan instruction. This corresponds to the ordered logit regression in Column 4 of Table 3.4.

Table 3.5: Effects for each level of treatment

Dependent variable: Identity

	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]
1 YEAR OF TREATMENT	-.185 (.158)	-.343 (.326)	-.117 (.127)	-.445 (.313)
2 YEARS OF TREATMENT	.039 (.185)	.118 (.354)	.318* (.178)	.751* (.422)
3 YEARS OF TREATMENT	.665*** (.162)	1.29*** (.329)	.548*** (.124)	1.147*** (.303)
4 YEARS OF TREATMENT	-0.91 (.21)	.062 (.373)	.285 (.177)	.721* (.398)
5 YEARS OF TREATMENT	1.033*** (.211)	1.981*** (.434)	.785*** (.163)	1.6*** (.407)
6 YEARS OF TREATMENT	.383 (.245)	.904** (.443)	.594*** (.227)	1.479*** (.508)
7 YEARS OF TREATMENT	.992*** (.226)	2.013*** (.447)	.934*** (.194)	1.967*** (.471)
8 YEARS OF TREATMENT	.562** (.28)	1.264** (.505)	.837*** (.264)	1.977*** (.588)
9 YEARS OF TREATMENT	.764*** (.274)	1.561*** (.538)	.577** (.233)	1.185** (.557)
10 YEARS OF TREATMENT	.464 (.329)	1.189** (.6)	.823*** (.303)	2.011*** (.679)
11 YEARS OF TREATMENT	.77** (.317)	1.57** (.624)	.772*** (.262)	1.575** (.63)
12 YEARS OF TREATMENT	.695** (.35)	1.585** (.637)	1.096*** (.321)	2.485*** (.719)
INDIVIDUAL CONTROLS	NO	NO	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES
PROVINCES	NO	NO	YES	YES
No. obs.	1467	1467	1467	1467
Rsquared	.1		.33	
Pseudo Rsquared		.036		.152

Note: The specifications reported in Columns 1 and 2 include years of education and year of birth fixed effects. The specifications reported in Columns 3 and 4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

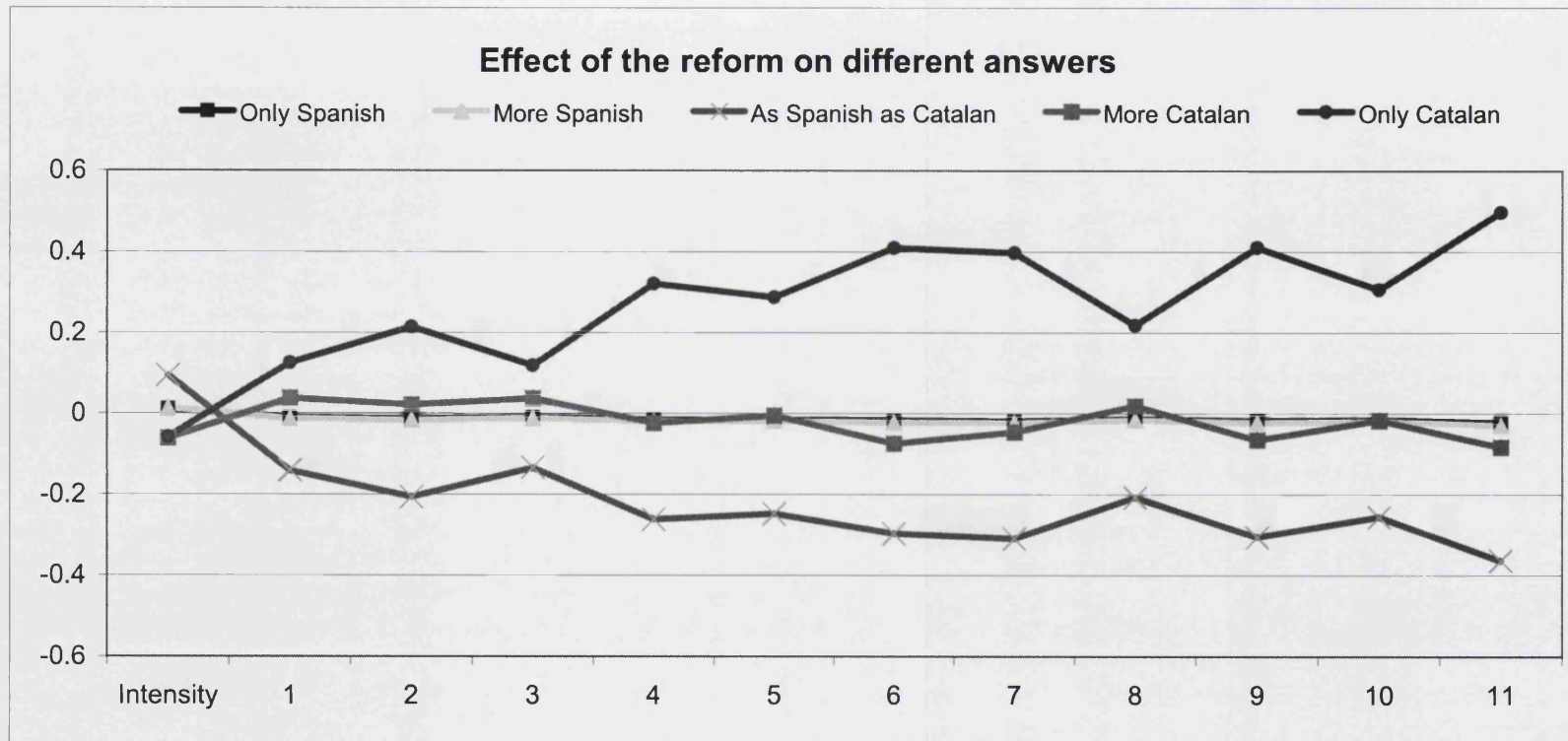


Figure 3.17: Effect for each level of treatment (specification 3.2)

Note: we plot how the effect the reform changes with the intensity of the reform, i.e. the number of years of Catalan instruction. This corresponds to the ordered logit regression in Column 4 of Table 3.5.

Table 3.6: Channels

Dependent variable: Identity

	Language expose of Catalan							
	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]	First stage [5]	First Stage [6]	2SLS [7]	2SLS [8]
Language exposure	0.032*** (.01)	0.075*** (.024)					.113*** (.262)	
use of Catalan			.193*** (.013)	.515*** (.04)				.508*** (.161)
Intensity					.695*** (.08)	.179** (.08)		
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES	YES	YES
No. obs.	1464	1464	1321	1321	1464	1321	1464	1321
Rsq	0.329		0.457		0.524	0.519	0.287	
Pseudo Rsq		0.152		0.232				0.138

Note: Individuals were asked to classify from 1 to 5 how much of their education was in Catalan and how much in Spanish; we then create an index from 0 to 1, with 0 corresponding to education only in Spanish and 1 to education only in Catalan. "Language exposure" is the product between this index and the years of education. We consider questions about the use of the Catalan language in everyday life: 1) when the interviewee meets friends, 2) when he is at home, 3) when he goes shopping, 4) if he is asked for directions in the street, 5) when he answers the phone, 6) when he interacts with civil servants 7) or when he is with the colleagues at work. Use of Catalan is an index that goes from 0 to 7, where 7 indicates that the interviewee uses the Catalan language in all of these circumstances. All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

Table 3.7: Other robustness checks (1)

Dependent variable: Identity

	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]	OLS [5]	Ord. Logit [6]
intensity	.08*** (.028)	0.176*** (.062)	.069** (.027)	0.164*** (.063)	.07** (.033)	0.18*** (.079)
SAMPLE	1948-1983	1948-1983	1948-1983	1948-1983	1953-1978	1953-1978
	all	all	catorigin	catorigin	catorigin	catorigin
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES
No. obs.	1467	1467	1333	1333	994	994
Rsq	0.326		0.288		0.287	
Pseudo Rsq		0.149		0.135		0.138

Note: All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. In the specifications reported in Columns 1 and 2 we include also the fraction of respondents, within each cohort-educational group, belonging to each of the 4 origin categories. In the specifications reported in Columns 3 and 4, the sample considered consists only of respondents born in Catalonia and in the specifications reported in Columns 5 and 6 is further restricted to individuals born between 1953 and 1978. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

Table 3.8: Other robustness checks (2)

	<i>OLS</i> [1]	<i>Ord. Logit</i> [2]	<i>OLS</i> [3]	<i>Ord. Logit</i> [4]
intensity	0.108*** (.027)	0.249*** (.064)	0.115*** (.031)	0.289*** (.072)
SAMPLE	1951-1980	1951-1980	1956-1975	1956-1975
INDIVIDUAL CONTROLS	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES
No. obs.	1245	1245	853	853
Rsq	0.318		0.354	
Pseudo Rsq		0.149		0.166

Note: All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. In the specifications reported in Columns 1 and 2 we consider only individuals born between 1951 and 1980; in the specifications reported in Columns 3 and 4 we consider only individuals born between 1956 and 1975. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%.

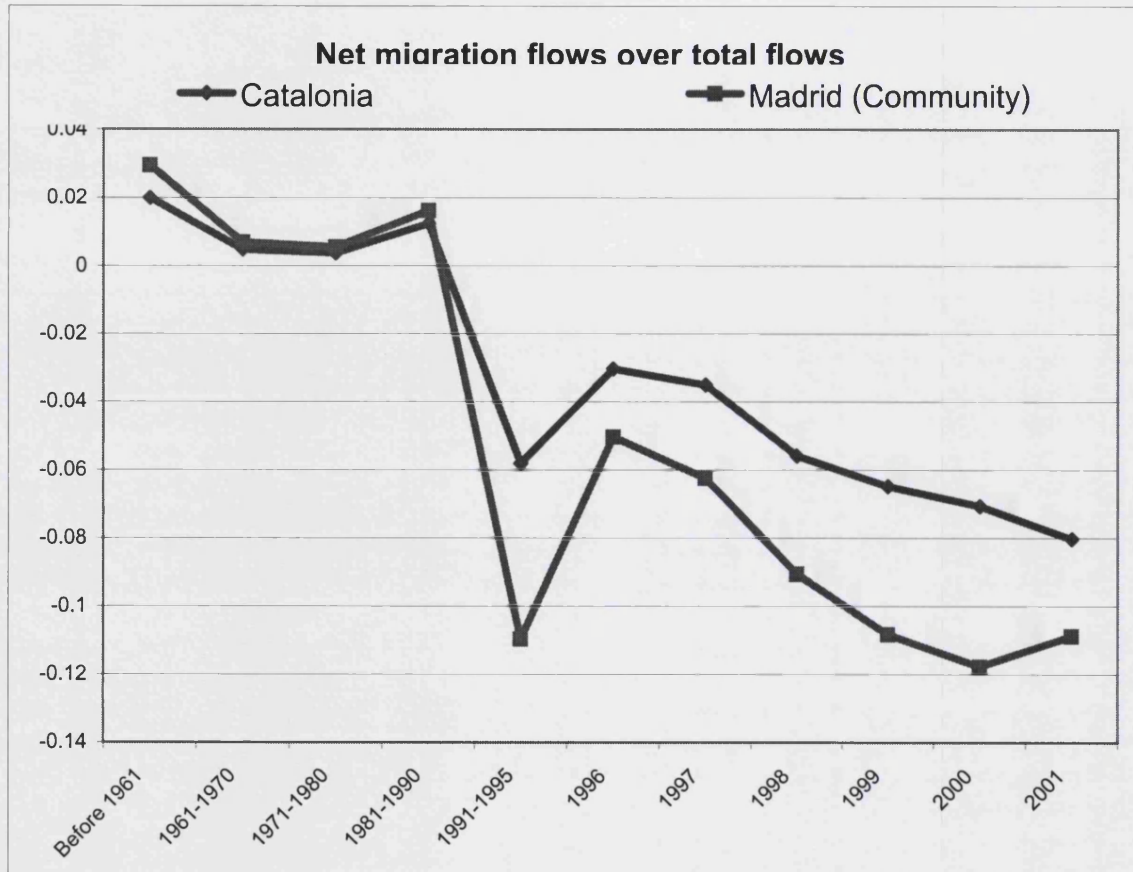


Figure 3.18: Migration

Table 9: Controlling for exogenous trends

Dependent variable: Identity

	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]	OLS [5]	Ord. Logit [6]	OLS [7]	Ord. Logit [8]
intensity			.115*** (.037)	.232** (.93)	.118*** (.033)	.236*** (.076)	.004 (.025)	.021 (.054)
pseudo-intensity	-.015 (.032)	-.07 (.075)	-.022 (.036)	-.024 (.094)				
Franco intensity					.083** (.04)	.122 (.1)		
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES	YES	YES
No. obs.	1930-1965 1008	1930-1965 1008	1921-1980 1860	1921-1980 1860	1467	1467	1260	1260
Rsq	.291		.295		.326		.231	
Pseudo Rsq		.131		.135		.149		.106

Note: In the specifications reported in Columns 1 and 2 we consider only cohorts who are not affected by the reform (1930-1965) and we assign a pseudo-treatment to the younger cohorts (1948 -1965). In the specifications reported in Columns 3 and 4 we consider cohorts born between 1921 and 1980 and the variable *pseudo-intensity* equals years of pseudo treatment for cohorts 1936-1950 and years of real treatment for cohorts 1966-1980. In the specifications reported in Columns 5 and 6 we include the variable *Franco intensity* that measures the number of years for which individuals attended school once Franco had died but before the educational system became bilingual. All the specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. In Columns 7 and 8, we use the standard identification strategy, but a different dataset: we use representative survey data on the institutional attachment of residents in the Basque Country to the Spanish State. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

Table 3.10: First stage*Dependent variable: Intensity*

	OLS [1]
Intensity compulsory	1.058*** (.028)
non catalan origin	.084 (.116)
catalan origin mixed family	-.108 (.089)
catalan origin non catalan family	-.122 (.086)
YEARS OF EDUCATION	YES
YEARS OF BIRTH	NO
PROVINCES	YES
AGE and AGE SQUARED	YES
No. obs.	1467
Rsq	.951

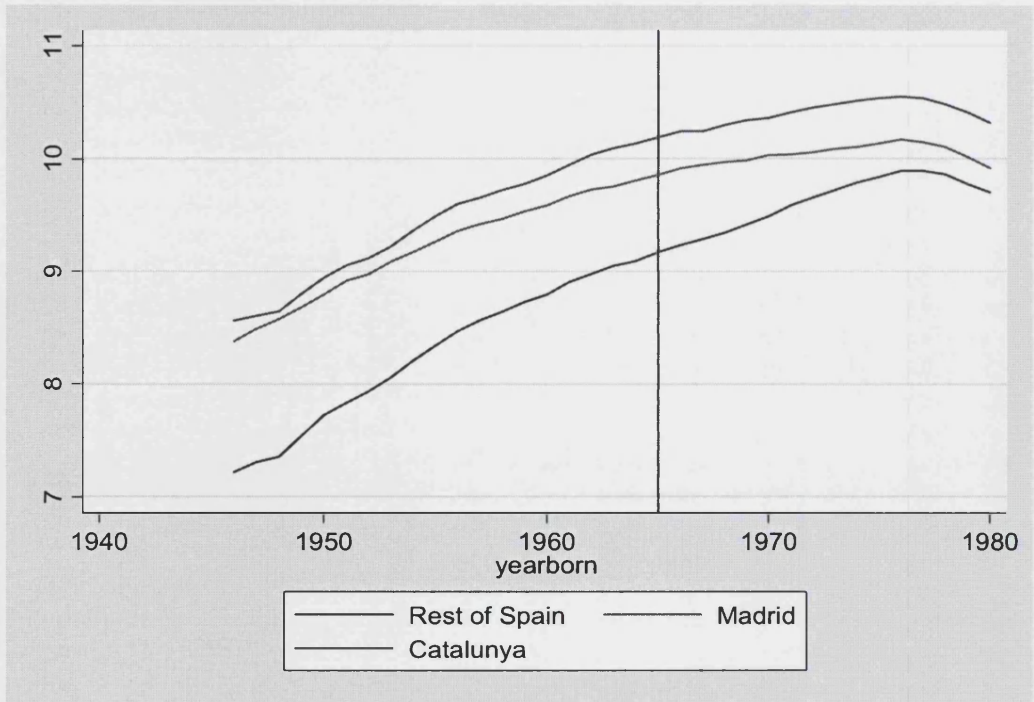
Note: The dependent variable is the number of years of education under the reform. The specification includes the number of years of compulsory education under the reform, years of education fixed effects, province fixed effects, the age and the age squared of the respondent and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

Table 3.11: Controlling for endogeneity of years of education

Dependent variable: Identity

	OLS [1]	Ord. Logit [2]	OLS [3]	Ord. Logit [4]	IV [5]	OLS [6]	Ord. Logit [7]	OLS [8]
intensity	.066** (.026)	.158*** (.059)	.055*** (.020)	.107** (.044)	.054** (.023)			
intensity compulsory						.056** -0.025	.105** (.049)	
language at home	.688*** (.096)	1.669*** (.235)						
intensity (people arriving)								.069** (.038)
AGE AT ARRIVAL	NO	NO	NO	NO	NO	NO	NO	YES
INDIVIDUAL CONTROLS	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF EDUCATION	YES	YES	YES	YES	OLS	YES	YES	YES
YEARS OF BIRTH	YES	YES	NO	NO	NO	NO	NO	YES
PROVINCES	YES	YES	YES	YES	YES	YES	YES	YES
AGE and AGE SQUARED	NO	NO	YES	YES	YES	YES	YES	NO
No. obs.	1467	1467	1467	1467	1467	1467	1467	361
Rsquared	.361		.299		.299	.296		.523
Pseudo Rsquared		.171		.136			.135	

Note: The specifications reported in Columns 1 and 2 include cohort fixed effects. The specifications reported in Columns 3 and 4 include the age and the age squared of the respondents. The specification reported in Columns 5 includes the age and the age squared of the respondents and the variable *intensity* is instrumented by the variable *intensity compulsory* (the first stage is in Table 3.10). The specifications reported in Columns 6 and 7 include the age and the age squared of the respondents. The specification reported in Column 8 includes cohort fixed effects and, as a measure of treatment, the number of years, within compulsory education, for which the respondents have studied in Catalonia after 1983. In this specification we restrict the sample to individuals who were not born in Catalonia. All the specifications include years of education fixed effects, province fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level (Columns 1-2), cohort level (Columns 3-7), cohort-year of arrival level (Column 8). * Significant at 10%, ** significant at 5%, *** significant at 1%.



Note: we plot average years of education by cohort for Catalonia, the Madrid region (a region which have always been at the same level of development as Catalonia) and the rest of Spain. This is done using data from the Spanish Census in 2001

Table 3.12: The role of individual origin

	OLS	Ord. Logit
	[1]	[2]
intensity*non cat.origin	.062* (.035)	.115 (.078)
intensity*cat.family	0.086*** (.026)	0.197*** (.06)
intensity*mixed family	0.081*** (.029)	0.172*** (.065)
intensity*non cat.family	0.075*** (.028)	0.163** (.063)
non catalan origin	-1.272*** (.114)	-2.835*** (.272)
mixed family	-.472*** (.086)	-.976*** (.184)
non catalan family	-.913*** (.095)	-2.047*** (.21)
YEARS OF EDUCATION	YES	YES
YEARS OF BIRTH	YES	YES
PROVINCES	YES	YES
No. obs.	1467	1467
Rsq	.325	
Pseudo Rsq		.15

Note: The specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). The specifications include also interactions between the variable *intensity* and each of the 4 categories. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

Chapter 4

Education and Political behaviour: Evidence from the Catalan Linguistic Reform

(joint with Oriol Aspachs-Bracons and Irma Clots-Figueras)

4.1 Introduction

Many social scientists have studied extensively the role of education and how schooling can create private and social benefits of different kinds. In particular, the economic literature has established a causal link between education and private earnings; and more recently, some scholars (Chiswick et al. 2006) have argued that schooling can increase the earnings of immigrants through its effect on language proficiency.

In addition, both the political science and the economic literature has investigated the link between education and political behaviour. In this respect, in 1978, Samuel Bowles argued that the school system has been instrumental in the interests of the elite and may have helped the consolidation of the capitalist system, reproducing the social relations which define the position of the capitalist class and the other dominant groups. However there is still not much evidence on the nature of the channels through which education can affect political outcomes and the causal link has only recently been carefully studied. Using compulsory education laws as instruments, Milligan et al. (2004) found a robust positive relationship between education and turnout for the United States. However, this was not the case for the United Kingdom.

In this chapter, we focus on the relationship between schooling and political behaviour in ethnically divided societies. We explore how the linguistic reform that took place in Catalonia affected turnout and party choice. In 1983, the Catalan education system became bilingual and the law established the predominant use of Catalan in primary and secondary schools. Using survey data from the “Centro de Investigaciones Sociológicas”, we study the effect of such a reform on the political behaviour of Catalan citizens. Identity is a likely channel through which the linguistic reform affected political choices.

Our empirical argument relies on within and between cohort variation in exposure to Catalan language at school. Between cohorts, the 1983 educational reform affected the younger cohorts (who were more exposed to the language) more than the older generations. Within a given cohort, the exposure to Catalan varies according to the number of years for which a student remains in education, as some students leave school earlier than others.

We find that individuals who have experienced greater exposure to teaching in Catalan are more likely to declare that they voted in the 1999 regional election and to have chosen a party with a Catalanist (i.e Catalan regionalist) platform. Thus, the introduction of a bilingual education system, a standard example of multicultural policy in ethnically divided societies, increased the salience of the ethnic issue in Catalan society and helped the consolidation of a political system organized along ethnic lines.

While we can reject the existence of education specific cohort trends for political behaviour and we tend to rule out the possibility that the reform changed the composition of the Catalan population, we still have to conduct some checks to account for the fact that individual years of schooling are typically an individual and parental choice, so it may be the case they are affected by the reform and, at the same time, correlated with individual political preferences.

The chapter is related to several strands of literature. Firstly, there is the literature that focuses on the determinants of turnout and political behaviour (see Merlo 2006 for a review of the topic) and, in particular, to that which studies the relationship between education and individual political choice (Milligan et al. (2004)).

Secondly, there is a connection with the early political science literature that analyzes the effect of multicultural policies and nation building policies on individual choice. If we interpret the reform as an example of a nation building policy within Catalonia, where individuals living in Catalonia who were born elsewhere, or whose parents do not have Catalan origins are the relevant minority, we find that the nation building policies foster civic duties (turnout) and attachment to nationalistic parties.

The chapter is organized as follows. Section 4.2 provides a brief description of the

Catalan party system and of the data we use. Section 4.3 describes our identification strategy, while Section 4.4 presents the basic empirical evidence. Section 4.5 discusses the role of individual origin. Section 4.6 provides some robustness checks, and Section 4.7 presents the conclusion.

4.2 Background

4.2.1 Institutional Setting

The democratization process that occurred in Spain during the late seventies, after the end of Franco's dictatorship, led to the approval of a Constitution that defined in turn a decentralized government structure. As basic democratic rights came to be recognized, the old political parties, which had existed in secrecy during the dictatorship, were legalized and new political parties also emerged. From the five parties in the Catalan Parliament at the time of the survey, one was created during the transition: *Convergència i Unió* (CIU), the party that was in power from 1980 to 2003. The other four had existed, with variations, prior to the dictatorship: *Partit dels Socialistes de Catalunya* (PSC), *Esquerra Republicana de Catalunya* (ERC), *Partido Popular* (PP), and *Iniciativa per Catalunya Verds* (ICV). Three of these are Catalan-only parties, CIU, ERC and ICV, in the sense that they only present candidates in Catalonia, while the other two exist throughout the Spanish territory. In our wider classification of Catalan and non-Catalan parties, we consider as Catalan all parties that have in their programmes the approval of a law that would give Catalonia the right to self-determination. These parties are CIU, ERC and ICV. Then, in our narrower classification of the Catalan parties, we include only those parties that are in favour of an independent Catalonia: CIU and ERC.

One of the main concerns of the new Catalan administration was the promotion of Catalan in the field of education. The main Educational law that was undertaken by the Catalan Government was the "General Directory of Linguistic Policy" which was described extensively in the previous chapter. The law was approved in 1983 (when CIU was in power) and introduced Catalan (together with Spanish) as the medium of instruction in all of the schools of Catalonia.

4.2.2 Data Description

The empirical analysis draws on survey data provided by the "Centro de Investigaciones Sociológicas" in 2001. In this survey, individuals were asked about their age, education,

origin, origin of their parents, when they arrived in Catalonia and questions about their use of Catalan, as well as some ideological questions: their identity, whom they voted for in the 1999 elections and whether they voted at all.

The linguistic reform was first implemented in the 1983-84 school year in every school in Catalonia. We adopt the same strategy as the previous chapter: to identify the cohorts who are likely to be treated, we assume that all of the respondents started their education at the age of 6. As a result, since, in Spain, primary education lasts for 8 years and secondary education for 4 years, a respondent born in 1966 with secondary education received one year of treatment. The 1966-81¹ birth cohorts are the cohorts who were potentially treated; a natural comparison group is the cohorts born between 1950-1965.

The intensity of the treatment varies across cohorts and level of education. If there is a link between language policy and voting behaviour, the reform will change the voting behaviour for those cohorts who were educated in Catalan: the younger (born after 1965) and more educated cohorts. We expect that individuals who experienced greater exposure to the reform will feel a greater identification with Catalonia and, as a result, to be more likely to vote in the regional elections and choose Catalanist parties. A preliminary look at the data shows that this is indeed the case.

In Figure 4.1, we plot the fraction of individuals who turned out to vote in elections by cohort and level of education (we distinguish between respondents with secondary education and those without). Even if the younger individuals vote less frequently, this negative trend is attenuated for individuals who were more affected by the Catalan reform: those born after 1966 who have secondary education. Figure 4.2 illustrates the difference between these two lines: that is, the within-cohort variation. For cohorts who were not affected by the reform, the line fluctuates around zero, while the difference starts to increase for cohorts born after 1965.

Figure 4.3 shows the fraction of voters who voted for Catalanist parties², by cohort and education. This figure shows a large increase in the proportion of younger and more educated voters who voted for Catalanist parties, while, for less educated voters, there is a positive effect, but less strong. Figure 4.4 plots the difference between these two lines and shows the within-cohort variation for each cohort. The difference is larger for cohorts born after 1965.

In Table 4.1, we show descriptive statistics of the main variables employed in this study. We show descriptive statistics for all individuals, and then for individuals with and

¹Although we have data for people born in 1982 and 1983, these individuals could not vote in the 1999 elections as they were under 18, so we do not include them in the sample.

²We use the wider definition

without secondary education. Within these groups, we show statistics for the different samples that we use in the regressions.

4.3 Empirical strategy

Students' exposure to Catalan language varies according to year of birth and years of education. We obtain differences-in differences estimates using the equation

$$y_{ijm} = \alpha + \beta L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (4.1)$$

L_{ijm} is the length of exposure to teaching in Catalan. γ_j is the cohort effect, δ_m is a dummy for education background and X_{ijm} is a vector of individual-level control variables, including gender, individual's origin, parents' origin and location dummies.³

As previously mentioned, we analyze the effect of the linguistic reform on political behaviour, namely turnout and party preferences. As a consequence, we use several dependent variables: y_{ijm} indicates whether or not an individual i , from cohort j with m years of schooling

(1) declares to have voted in the 1999 regional election. This specification checks if the linguistic reform stimulated turnout to the regional elections.

(2) declares to have voted for a Catalanist party in the 1999 regional election (we restrict the sample only to people who declared to have voted in that election). As a first step we consider IC, ERC and CIU as Catalanist parties, then we look only at ERC and CIU. Finally, we consider only CIU, the party that was in power and promoted the linguistic reform. This specification checks whether the linguistic reform increased the share of votes for Catalanist parties in the regional elections.

(3) declares to have voted for a Catalanist party in the 1999 regional election (we do not restrict the sample). As in point (2), firstly, we consider IC, ERC and CIU as Catalanist parties, then look only at ERC and CIU and, as a last step, we consider only CIU. This specification checks if the linguistic reform increased the total number of votes for the Catalanist parties the in regional elections.

The identification strategy relies on the fact that there is no other variable that affects the political behaviour of some cohort-years of education. We run this specification first as a linear regression and then as a logit, due to the binomial nature of the dependent

³We exclude from the sample any migrants who received some education outside Catalonia.

variable⁴. Robust standard errors are clustered at cohort-years of education level, to control for the fact that observations in a given cohort-years of education group may be correlated.

The second econometric specification to be tested is very similar to the first one. However, the dependent variable can assume three different values. This time y_{ijm} indicates whether individual i , from cohort j and with m years of schooling declared to have (i) voted for a Catalanist party (ii) voted for a non-Catalanist party or (iii) abstained. We run this specification first as a multinomial logit; however, the multinomial logit would assume the independence of irrelevant alternatives. We then run a multinomial probit, assuming the error terms to be independent, and, as a last step, we run it as a multinomial probit but relaxing the assumption of independence of the error terms⁵.

4.4 Results

4.4.1 Difference-in-Difference Estimates

Table 4.2 reports estimates for Equation (4.1) using as dependent variable the probability of voting during the regional election of 1999. In Columns 1 and 2, we show estimates when we only control for years of education and year of birth fixed effects, for the OLS and the logit specifications, respectively. OLS estimates show a positive and significant coefficient, suggesting that an increase in exposure to the reform increases the probability that an individual votes. The logit results go in the same direction and are significant. In Columns 3 and 4, we add province dummies, controls for gender, family and individual origin. The coefficients are still positive and significant. The results suggest that those who received Catalan-intensive instruction tend to be significantly more likely to vote; in fact, an extra year of education in Catalan increases the likelihood that the respondent declared to have voted by 4.5%. The probability of voting is also correlated with the individual origin and with the origin of the parents of the interviewee. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia 4) individuals who were born in Catalonia whose parents were both born in Catalonia. We find that the probability of voting for a Catalanist party is highest among respondents with Catalan origins and with parents with Catalan origins.

Tables 4.3 and 4.4 report estimates for Equation (4.1) using as dependent variable the

⁴The results for probit are very similar.

⁵We use the `asmprobit` command in STATA.

probability of voting for a Catalanist party during the regional election in 1999 when we restrict the sample only to people who declared that they voted in that election (Table 4.3) and when we do not restrict the sample (Table 4.4). In Columns 1 and 2 of both tables we show estimates, for the OLS and the logit specifications, when we only control for years of education and year of birth fixed effects and when we classify IC, ERC and CIU as Catalanist parties. In Columns 3 and 4 of both tables, we consider only CIU and ERC as Catalanist parties. In Columns 5 and 6 of both tables, we focus only on CIU. OLS estimates show a positive and significant coefficient, suggesting that an increase in exposure to the reform increases the probability that an individual will vote for a Catalanist party. Furthermore, the results clearly show that most of the effects we find in Tables 4.2 and 4.3 come from the increase in the probability of choosing CIU. The logit results go in the same direction and are significant. In Columns 7-12 of both tables, we add province dummies, controls for family and individual origin. The coefficients are still positive and significant. The results suggest that those who received Catalan-intensive instruction tend to be significantly more likely to vote for a Catalanist party, both if we consider the sample of voters or the total sample. The probability of voting for a Catalanist party is also correlated, as predictable, with individual origin and with the origin of the parents of the interviewee. We find that probability of voting is highest among respondents with Catalan origins and with parents who have Catalan origins. As shown in Column 8 of Table 4.3 (restricted sample), an increase from 4 years of treatment (the average amount of treatment) to 8 years (the average amount of treatment plus one standard deviation) would increase the likelihood that the respondent declared to have voted for a party with Catalan affiliation by almost 20%. Column 8 of Table 4.4 (unrestricted sample) shows that an increase from 4 to 8 years of treatment would increase the likelihood that the respondent declared to have voted for a party with Catalan affiliation by 27%.

As a final step, we show results using the second econometric specification mentioned in the previous section. The dependent variable is now a variable that can assume three different values depending on whether the respondent: (i) voted for a Catalanist party (ii) voted for a non-Catalanist party or (iii) abstained. For this specification we consider CIU, ERC and IC as Catalanist parties. Columns 1-3 of Table 4.5 reports the results when we run this specification as a multinomial logit, Columns 4-6 when we run it as a multinomial probit, but assuming the error terms to be independent. Columns 7-9 report the results when we run it as a multinomial probit but removing the assumption of the independence of the error term. The picture the table shows is, however, quite homogeneous. We observe that the reform caused a significant decline in abstention, a decline (but not always significant at standard level, i.e. significant at 10 per cent level only with the multinomial logit specification) in voting for a non Catalanist party and also a contemporaneous and significant increase in voting for a Catalanist party.

The results in Columns 7-9 show how an increase in years of Catalan instruction from 4 to 8 increases the likelihood that the respondent declared to have voted for a Catalanist party by 25% and decreases the likelihood that the respondent declared to have voted for a non-Catalanist party by 16%. However, since the decline in the amount of votes for non-Catalanist parties is not always significant, the data do not allow us to distinguish between two different patterns of voting behavior:

1) a simple switch from abstention to voting for a Catalanist party

2) a more complex scenario in which

(i) on the one hand, the reform contributed to a general rise in electoral participation and, then, as a result, led to a switch from abstention to voting for both Catalanist and non-Catalanist parties

(ii) but, on the other hand, the reform increased individual preferences for a Catalanist party and, as a result, led to a switch from voting for a non-Catalanist party to a Catalanist party.

4.5 Heterogenous effects

In this section, we study whether the effect of the reform was homogenously spread across the population. We then interact our intensity measure with the four origin dummies we created and, as Table 4.6 shows, we find that the effect of the reform on both turnout and preference for a Catalanist party is positive not only for those respondents with Catalan origin and whose parents were both born in Catalonia, but also for individuals born in Catalonia but whose parents were not, for individuals born in Catalonia with only one parent with Catalan origin and for individuals who were not born in Catalonia. As is shown in Figures 4.5, 4.6 and 4.7, while political choices vary widely depending on origin among individuals with no treatment, origin plays almost no role among individuals with full treatment.

Since, as already mentioned in the introduction, we can interpret the reform as an example of a nation building policy within Catalonia, where individuals living in Catalonia who were born elsewhere, or whose parents do not have Catalan origins are the relevant minority, we conclude that nation building policies foster civic duties (turnout) and attachment to nationalistic parties.

4.6 Robustness

In this section we perform several robustness checks on the main specification. Firstly, we show that our results are not driven by education-specific cohort trends. Then, we deal with the possibility that years of education may be endogenous. Finally, we restrict the sample and conduct some extra checks to control for the fact that the reform could have encouraged migration in and out of Catalonia.

4.6.1 Exogenous Trends

The estimates in Tables 4.2, 4.3 and 4.4, which show how intensive Catalan instruction affects political behaviour, may be confounded by education-specific cohort trends for turnout and political preferences. In particular, these trends may exist if different cohorts have been raised by parents with different values and political preferences (the younger parents being more pro-Catalan), and if those parents with stronger preferences for Catalanist parties have invested more in their children's education.

As a formal falsification exercise, we run the specification proposed by Equation (4.1) on cohorts unaffected by the reform, namely cohorts 1934 to 1965. Cohorts 1950 to 1965 received a pseudo treatment that consists of the intensity of the reform received by individuals with the same years of education but born 16 years later. Columns 1-6 of panel A in Table 4.7 report the results of this exercise. We adopt both an ordinary least squares and a logit specification. The coefficient of the pseudo treatment variable is never significant, negative and very small when we consider turnout as dependent variable and positive but sufficiently small when we use as dependent variable whether or not the respondent declared that he/she voted for a Catalanist (IC, ERC and CIU) party in the 1999 regional election (both with the sample restricted only to people who declared to have voted in that election and with the unrestricted sample).

In addition, in order to correct the estimates in Tables 4.1-4.3 for education-cohort trends, we perform a triple differences-type of analysis similar to that of Angrist et al (2006), using cohorts born between 1921 and 1980. We use the following equation:

$$y_{ijm} = \alpha + \beta L_{ijm} + \theta l_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + \varepsilon_{ijm} \quad (4.2)$$

where L_{ijm} is a term that captures real exposure to teaching in Catalan, while l_{ijm} captures spurious effects.

L_{ijm} are the years of real treatment for cohorts 1966-1980; while l_{ijm} equals years of

pseudo treatment for cohorts 1936-1950 (where now the pseudo-treatment consists of the intensity of the reform received by individuals with the same years of education but born 30 years later) and years of real treatment for cohorts 1966-1980⁶. β are the triple difference estimates, i.e. the treatment effects from the real experiment net of the pseudo treatment effects estimated using cohorts 1921-1950. Columns 7-12 of panel A in Table 4.7 report the results of this specification. Triple difference estimates are positive and significant, while the coefficient that would be capturing spurious trends is negative and not significant. Both exercises are reassuring, as they may indicate that our results are not capturing the effect of an education-specific cohort trend.

As a further step, we study the possible interactions between the reform and Franco's death in 1975. One may think that the end of the dictatorship could have affected the content of the textbooks and this, in turn, could have modified the political behaviour of the younger and more educated cohorts. In order to check whether this is the case, we propose the following exercise. We generate a variable called "Franco Intensity" that measures the number of years for which individuals attended school after Franco died but before the educational system became bilingual. Then, we run the Specification (4.1) with our intensity measure and this additional variable. The results are shown in panel B of Table 4.7 and show how our measure of intensity remains positive and significant, while the Franco intensity variable is not significant and smaller. The coefficient of interest is not significant at standard level (but it is at 10,3% level!) in only one of the six specifications that we run, i.e. in the one where, within an OLS estimation strategy, we use as dependent variable the probability that the respondent declares that he/she voted for a Catalanist party (restricting the sample only to people who declared that they voted in that election and considering IC, ERC and CIU as Catalanist parties). The coefficient of the variable that measures the exposure to the language reform, however, is similar in magnitude to the one we found in the main specification and, importantly, the coefficient of the variable "Franco intensity" is negative, pointing out that years of education after Franco's death could not have affected people's political preferences. This test further supports the theory that political behaviour changed with the language reform.

Like in the previous chapter, as a further test to rule out the possibility that our results are driven either by the existence of education-specific cohort trends or by the effect of the end of the dictatorship on the political behaviour of the younger and more educated cohorts, we consider an other Spanish region: the Basque Country. We use representative survey data on the institutional attachment of residents in the Basque Country to the Spanish State. This survey was also published by CIS in 2001 and is very similar to the survey conducted in Catalonia. The only difference is that the

⁶Due to lack of data for very old cohorts we could not use a larger number of cohorts.

Basque survey was conducted one month before the Catalan one. We then apply the same empirical strategy to the Basque sample and we find that the coefficient of the variable of interest (years of treatment) is not significantly different from zero (Columns 7-12 of panel B in Table 4.7) in all the specifications considered. Since the death of Franco is likely to have affected residents of Catalonia and Basque Country in the same way and education-specific cohort trends are likely to follow similar patterns in both the regions, the comparison between Catalonia and Basque Country helps us to exclude the possibility that the findings discussed in the previous sections are driven either by the existence of education-specific cohort trends or by the end of the Franco's dictatorship.⁷

4.6.2 Endogeneity of years of education

In this section, we discuss the possibility of an endogenous response of schooling to the 1983 language reform. This is a concern, since we rely on the differences between schooling groups across cohorts to identify the effect of the reform. Indeed, we may think that respondents with very intense anti-Catalanist political preferences may have decided to drop out of school because of the new legislation and because they may felt less comfortable about receiving education in the Catalan language.

As a preliminary check, we include in our specification also a dummy variable that is equal to one if the respondent used to speak Catalan at home with their parents. The results are reported in Columns 1 and 2 for all sets of regressions in Table 4.8 and these are consistent with those shown in Tables 4.1-4.3. We believe that this is important, since individuals who are less confident in the Catalan language may drop out of school earlier as a consequence of the reform and, at the same time, are also the ones more likely to vote for a non-Catalanist party.

We then remove cohort fixed effects from Equation (4.1); in this way, our identification strategy relies less on within cohort variation in years of education, that is likely to be endogenous. In order to control for age trends correlated with the dependent variables, however, we include in our specification also the age and the squared age of the respondents. The results are reported in Columns 3 and 4 in Table 4.8 and are similar to those obtained previously, even though the coefficients are quite smaller.

Moreover, as a further (but unsuccessful) robustness check, we consider only compulsory education and eliminate all within cohort variation. We build a measure of exposure to the reform that is not the result of individual choice: the number of years of compulsory education received in Catalan language. Each respondent belonging to the same cohort will be subject to the same amount of exposure to compulsory education. This index

⁷We consider pnv, eh and ea as ethnic/ Basque regionalist parties.

should provide an exogenous measure of years of education, but it should also increase the measurement error and cause attenuation bias in the coefficient of our variable of interest. Unfortunately, as Columns 5-6 in Table 4.8 show, the coefficients are positive but smaller in size and not significant.

As shown in Figures 4.1-4.4, the effect of the reform varies according to the amount of exposure: that is, according to the number of years for which a given individual has been taught in Catalan. In the previous regressions, we do not take this variation into account and so we cannot add cohort fixed effects. Given that voting behaviour may change with age, it would seem appropriate to include cohort fixed effects in the regression. Thus, instead of using years of education as intra-cohort variation in years of treatment, we use variation in the timing of arrival as a determinant of more or less exposure to the electoral reform⁸. In order to perform this exercise, we restrict the sample to individuals who were not born in Catalonia. These individuals arrived at different points in time and at different stages during their education. For each individual, we compute the number of years, within compulsory education, that they have studied in Catalonia after 1983. We believe that this specification is particularly useful, since it allows us to use both compulsory years of education and cohort fixed effects. In this way, we are able to control for any issue related to the endogeneity of the education decision and, at the same time, and rule out any potential problem of omitted variables correlated with both identity and age (although in a non linear fashion). The results, reported in Column 7 in Table 4.8, are fairly similar to those obtained when the within-cohort variation used was given by differences in years of education⁹.

Morover, as Figure 3.19 in the previous chapter shows, when we plot average years of education by cohort for Catalonia, the Madrid region and the rest of Spain, we do not find an anomalous pattern for cohorts affected by the reform (those born in Catalonia after 1965).

4.6.3 Restricting the sample

Cohorts affected and not affected by the reform are actually individuals of very different ages. In order to compare cohorts who are more similar to each other and who are more likely to be affected by similar shocks, we restrict the sample and compare cohorts of similar ages. The results are shown in panel A of Table 4.9. A comparison between individuals aged 26 (cohort 1975) and individuals aged 45 (cohort 1956) should be more reliable than a comparison between individuals aged 20 (cohort 1981) and individuals

⁸To this purpose, we also include in the sample migrants who received some education outside of Catalonia and who were not included in any of the previous specifications.

⁹Due to lack of data, we use only an OLS specification.

aged 51 (cohort 1950). The results are robust. The coefficients are always positive and significant and their size is also similar in magnitude to the ones we found in the main specification. The coefficient of interest is not significant at standard level (only at 15% level!) in only one of the six specifications that we run, i.e. the one where, within an OLS estimation strategy, we use as dependent variable the probability that the respondent declares to have voted for a Catalanist party (restricting the sample only to people who declared that they voted in that election and considering IC, ERC and CIU as Catalanist parties).

4.6.4 Migration patterns

Migration patterns in and out of Catalonia changed the composition of the Catalan population. As table 4.1 shows, the fraction of respondents whose parents were both born in Catalonia is larger among control cohorts than among treated cohorts. In general the family characteristics of a respondent belonging to the control cohorts are different from the family characteristics of a respondent belonging to the treated (or potentially treated) cohorts. Although in the main specification we control for the type of family of the respondents, migration patterns, affecting the composition of each cohort educational group, might have biased our results. In order to control for this possible source of bias, we add in the regressions a new set of controls. As already mentioned, we classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia. As in the previous chapter, we calculate, within each cohort-educational group, the fraction of respondents belonging to each of the 4 categories. We then adopt the following econometric specification

$$y_{ijm} = \alpha + \beta_k L_{ijm} + \gamma_j + \delta_m + X_{ijm}\mu + s_{jm} + \varepsilon_{ijm} \quad (4.3)$$

Where the dependent variable y_{ijm} indicates, as usual, whether individual i , from cohort j and with m years of schooling answers (1) I feel only Spanish, (2) I feel more Catalan than Spanish, (3) I feel as Spanish as Catalan, (4) I feel more Catalan than Spanish, (5) I feel only Catalan. s_{jm} is a vector of 4 explanatory variables that control for the composition of each cohort educational group jm . Each variable measures the size, within the cohort educational group jm , of one of the 4 family groups.

It turns out, as columns 1-6 of Panel B in Table 4.9 show, that β_k , the coefficient of

our variable of interest, is still positive and significant and the size of the coefficient is very similar to the size of the coefficient in our baseline specification.

A further concern is endogenous migration; moving costs could also be related to schooling opportunities. Thus, the effect of the reform on migration costs could be particularly strong for potential migrants with highly anti-Catalanist political preferences. This could introduce a selection problem in the sample we study.

We therefore carry out the following exercise: we restrict our sample to individuals born in Catalonia. In this way, since the youngest cohort is the one born 1981, we consider only respondents whose parents were already in Catalonia when the reform was implemented. As a result, their parents' decision to migrate should be less likely to have been affected by the linguistic reform. Columns 1-6 of panel B in Table 4.9 show that the results are robust to this check.

While panel B in Table 4.9 successfully controls for the possible endogeneity of the migration inflows, unfortunately we have not been able to perform any tests that consider the possible effect of the reform on patterns of migration outflows. However, as reported in the previous chapter (Figure 3.18), when we use data from INE and compare net migration flows during the last 40 years in Catalonia with those in Madrid, the other main Spanish region, we find that migration patterns are fairly similar.

4.7 Conclusions

In ethnically divided societies, education can affect political behaviour through the language used for teaching. This chapter presents evidence of a positive effect of the educational reform, whereby the Catalan education system became bilingual, on turnout and on the likelihood of choosing a party with a Catalanist nature. Identity is a likely channel through which the linguistic reform affected political choices.

4.8 Appendix

4.8.1 Definition of the variables

-Years of education: (i) if the respondent answered that he did not complete primary education, we assume that he received 4 years of schooling; (ii) if the respondent answered that he completed only primary education, we assume that he received 8 years of education; (iii) if the respondent answered that he started but did not complete

secondary education or that he received some professional training, we assume that he received 10 years of education; and (iv) if the respondent answered that he completed secondary education or a higher level of professional training, we assume that he received 12 years of education. Source: CIS

-*Catalan origin, Catalan family*: dummy equal to (1), if the respondent answered that he was born in Catalonia and both his parents were born in Catalonia. Source: CIS

-*Catalan origin, mixed family*: dummy equal to (1), if the respondent answered that he was born in Catalonia and only one of his parents were born in Catalonia. Source: CIS

-*Catalan origin, non Catalan family*: dummy equal to (1), if the respondent answered that he was born in Catalonia and both his parents were not born in Catalonia. Source: CIS

-*non Catalan origin*: dummy equal to (1), if the respondent answered that he was not born in Catalonia. Source: CIS

-*Voting*: dummy equal to (1), if the respondent declares to have voted in the 1999 regional election. Source: CIS

-*Catalan Voting (IC+ERC+CIU)*: dummy equal to (1) if the respondent declares to have voted for a Catalanist party in the 1999 regional election. We consider IC, ERC and CIU as Catalanist parties Source: CIS

-*Catalan Voting (ERC+CIU)*: dummy equal to (1) if the respondent declares to have voted for a Catalanist party in the 1999 regional election. We consider ERC and CIU as Catalanist parties Source: CIS

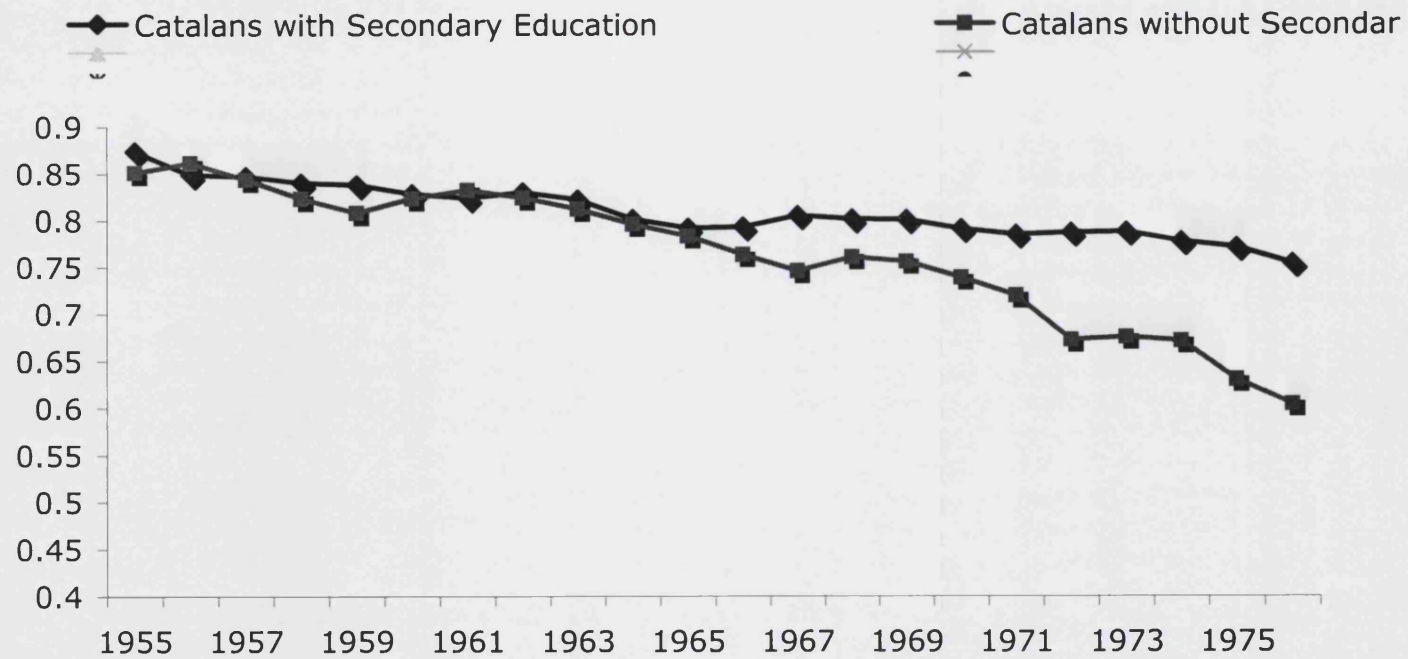
-*Catalan Voting (CIU)*: dummy equal to (1) if the respondent declares to have voted for a Catalanist party in the 1999 regional election. We consider CIU as the only Catalanist parties Source: CIS

-*Language at home*: dummy equal to (1), if the individual spoke Catalan at home with his parents

-*Female*: dummy equal to (1), if the respondent is female

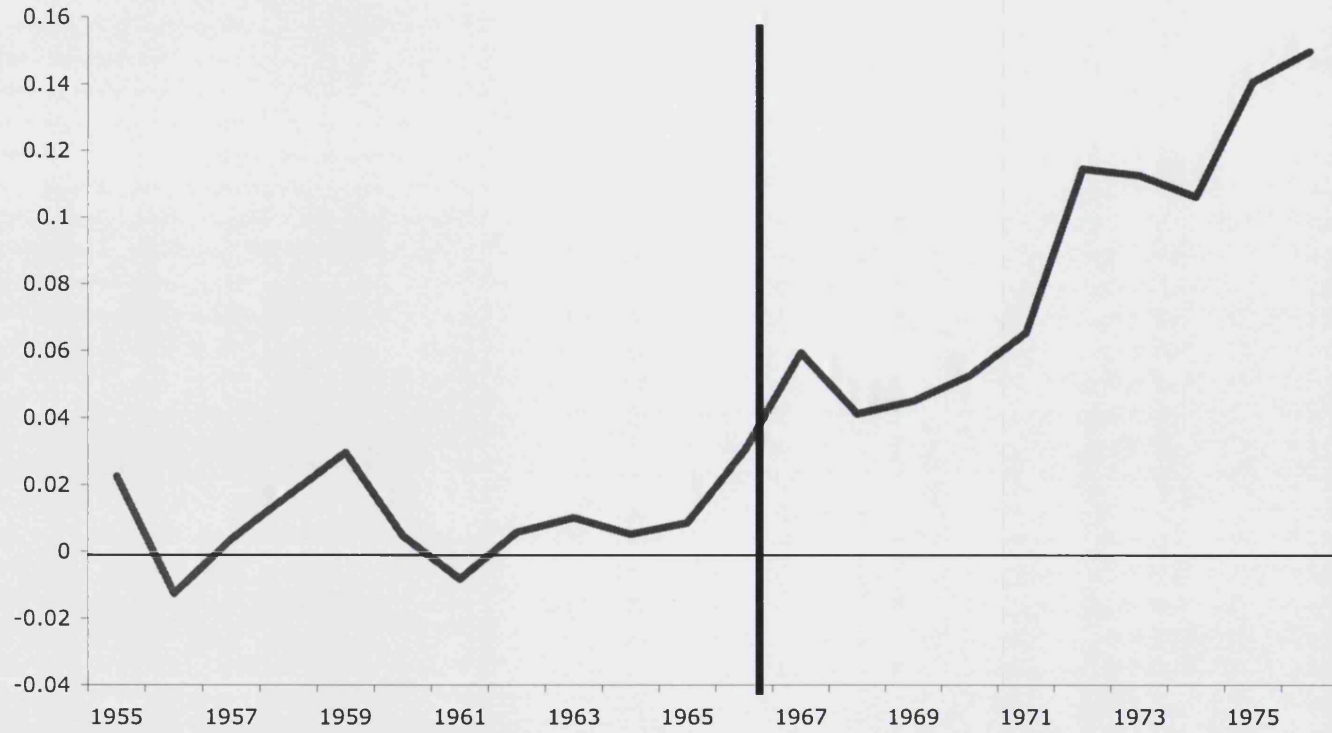
4.9 Figures and Tables

Figure 4.1: Participation rate by cohort and level of education



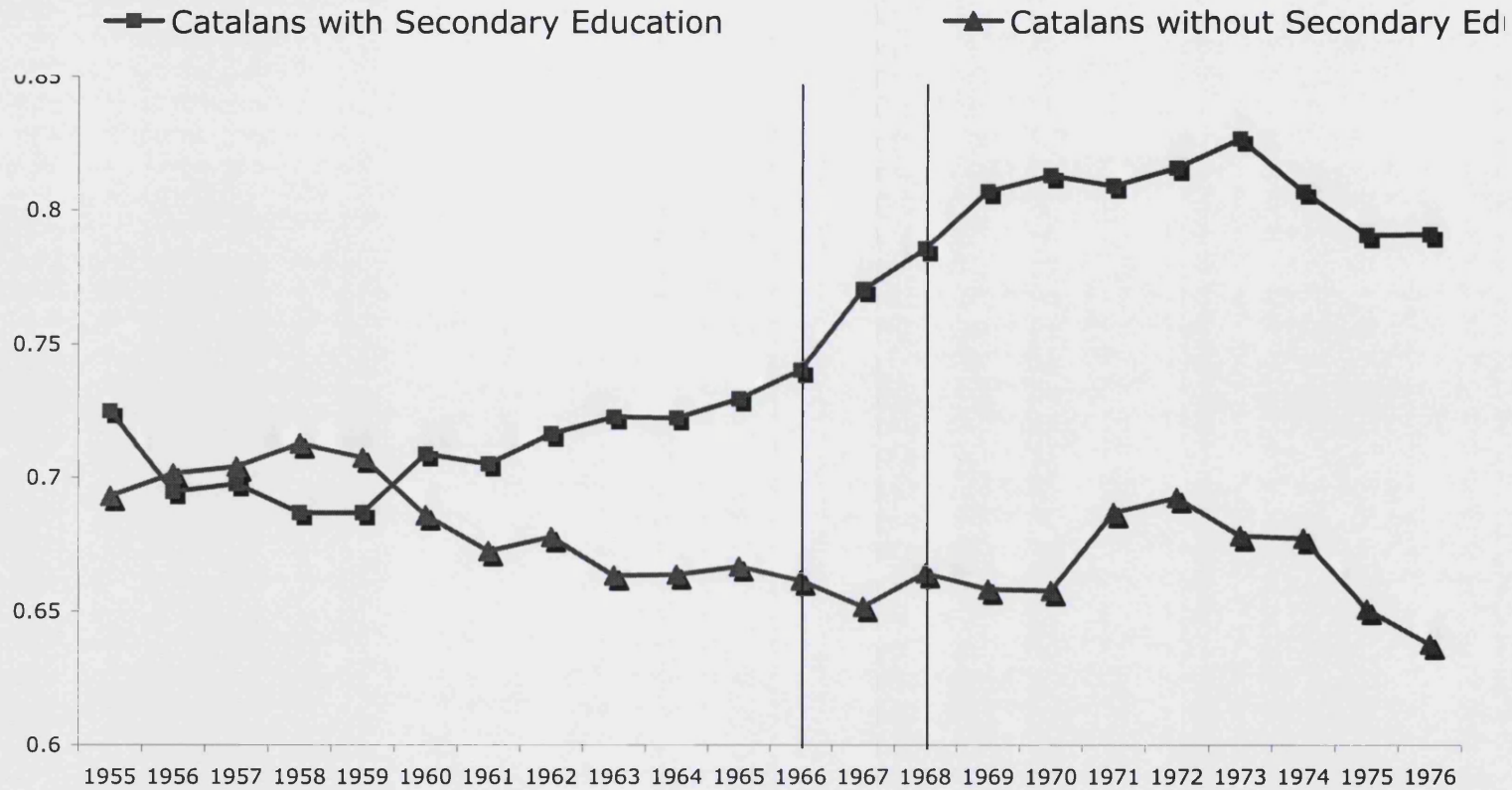
Note: The graph plots the fraction of respondents who voted during the regional elections against year of birth. All series are (5,1,5) moving averages.

Figure 4.2: Effect of the reform on the participation rate by cohort



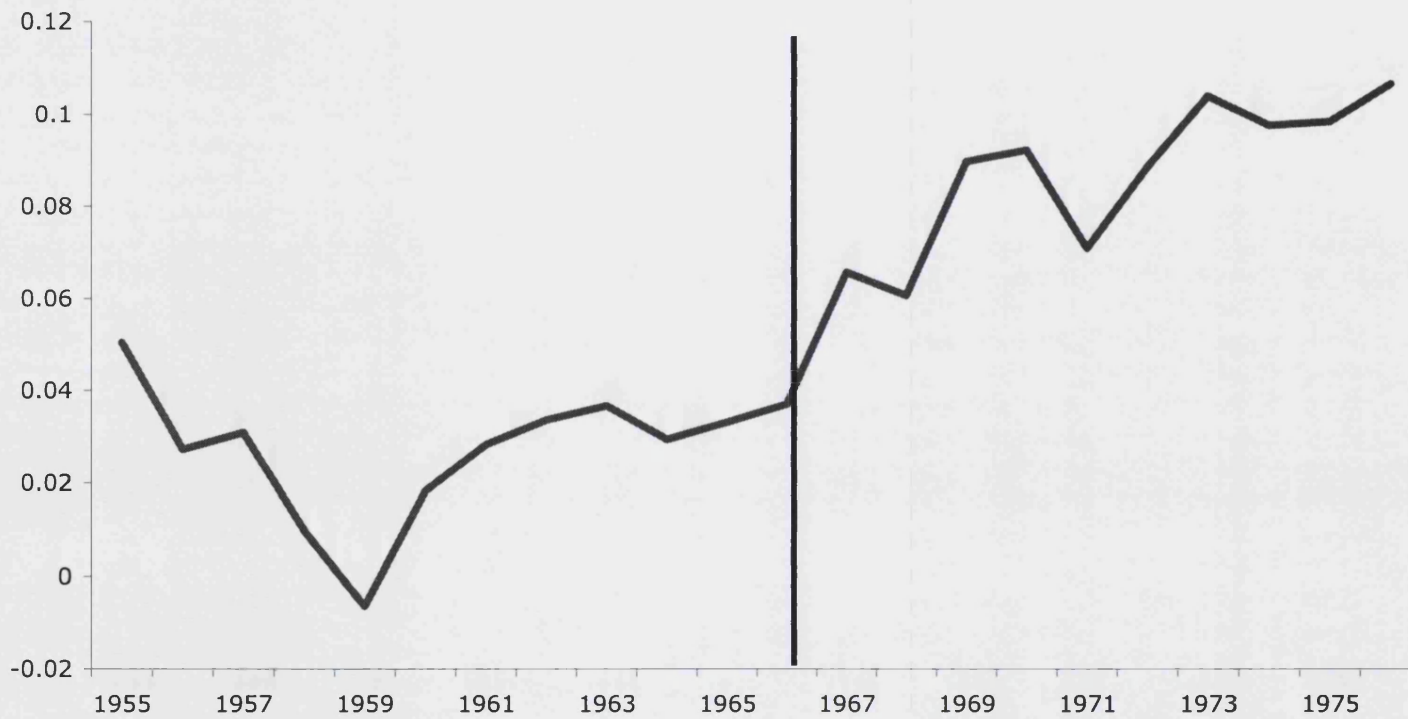
Note: The graph plots the difference between the fraction of respondents with secondary education who voted during the regional elections and the fraction of respondents without secondary education who voted during the regional elections against year of birth. All series are (5,1,5) moving averages.

Figure 4.3: Ethnic Voting by cohorts and level of education



Note: The graph plots the fraction of respondents who voted for a Catalanist party during the regional elections against year of birth. All series are (5,1,5) moving averages.

Figure 4.4: Effect of the reform on Ethnic Voting by cohort



Note: The graph plots the difference between the fraction of respondents with secondary education who voted for a Catalanist party during the regional elections and the fraction of respondents without secondary education who voted for a Catalanist party during the regional elections against year of birth. All series are (5,1,5) moving averages.

Table 4.1

		Cohorts (total sample)			Cohorts (with sec. edu)			Cohorts (without sec. edu)		
		1950-81	1950-65	1966-81	1950-81	1950-65	1966-81	1950-81	1950-65	1966-81
Identity	Average	3.52	3.54	3.50	3.7	3.71	3.7	3.3	3.39	3.21
	St. Dev.	1.02	1.03	1.01	0.9	0.9	0.9	1.1	1.1	1.09
	Observations	1326	557	769	680	236	444	646	321	325
Identity (binomial)	Average	0.49	0.49	0.48	0.57	0.55	0.58	0.39	0.45	0.34
	St. Dev.	0.50	0.50	0.50	0.49	0.49	0.49	0.48	0.49	0.47
	Observations	1326	557	769	680	236	444	646	321	325
Years of Education	Average	10.38	9.97	10.67	12	12	12	8.46	8.27	8.65
	St. Dev.	1.88	1.99	1.74	0	0	0	0.99	0.98	0.97
	Observations	1349	565	784	697	241	456	652	324	328
Cat. Origin, Cat. Family	Average	0.43	0.46	0.40	0.49	0.56	0.45	0.35	0.38	0.32
	St. Dev.	0.49	0.49	0.49	0.5	0.49	0.49	0.47	0.48	0.46
	Observations	1349	565	784	697	241	456	652	324	328
Cat. Origin, mixed Family	Average	0.17	0.14	0.19	0.17	0.14	0.19	0.17	0.14	0.19
	St. Dev.	0.38	0.35	0.39	0.38	0.35	0.39	0.37	0.35	0.4
	Observations	1349	565	784	697	241	456	652	324	328
Cat. Origin, no Cat. Family	Average	0.28	0.21	0.32	0.24	0.16	0.28	0.32	0.25	0.39
	St. Dev.	0.44	0.41	0.46	0.43	0.37	0.45	0.46	0.43	0.48
	Observations	1349	565	784	697	241	456	652	324	328
No Catalan Origin	Average	0.11	0.17	0.06	0.08	0.13	0.05	0.14	0.2	0.08
	St. Dev.	0.31	0.37	0.25	0.27	0.33	0.23	0.35	0.4	0.27
	Observations	1349	565	784	697	241	456	652	324	328
Voting	Average	0.77	0.83	0.72	0.79	0.84	0.77	0.74	0.83	0.65
	St. Dev.	0.41	0.36	0.44	0.40	0.36	0.41	0.43	0.37	0.47
	Observations	1248	528	720	642	223	419	606	305	301
Catalan voting (CIU+ERC+IC)	Average	0.50	0.52	0.48	0.56	0.57	0.55	0.42	0.48	0.37
	St. Dev.	0.50	0.49	0.50	0.49	0.49	0.49	0.49	0.5	0.48
	Observations	1248	528	720	642	223	419	606	305	301
Catalan voting (CIU+ERC)	Average	0.44	0.47	0.42	0.49	0.49	0.48	0.39	0.45	0.33
	St. Dev.	0.19	0.49	0.49	0.5	0.5	0.5	0.48	0.49	0.47
	Observations	1248	528	720	642	223	419	606	305	301
Catalan voting (CIU)	Average	0.31	0.33	0.29	0.31	0.3	0.31	0.3	0.35	0.25
	St. Dev.	0.46	0.47	0.45	0.46	0.46	0.46	0.46	0.47	0.43
	Observations	1248	528	720	642	223	419	606	305	301
Catalan voting (CIU+ERC+IC)—only voters	Average	0.66	0.63	0.68	0.71	0.68	0.73	0.59	0.58	0.59
	St. Dev.	0.47	0.48	0.46	0.45	0.46	0.43	0.49	0.49	0.49
	Observations	941	439	502	503	190	313	438	249	189
Catalan voting (CIU+ERC)—only voters	Average	0.58	0.57	0.60	0.62	0.59	0.64	0.54	0.55	0.53
	St. Dev.	0.49	0.49	0.48	0.48	0.49	0.48	0.49	0.49	0.5
	Observations	941	439	502	503	190	313	438	249	189
Catalan voting (CIU)—only voters	Average	0.41	0.40	0.41	0.4	0.36	0.42	0.42	0.43	0.41
	St. Dev.	0.49	0.49	0.49	0.49	0.18	0.49	0.49	0.49	0.49
	Observations	941	439	502	503	190	313	438	249	189
Language at home	Average	0.44	0.47	0.41	0.5	0.54	0.47	0.36	0.42	0.31
	St. Dev.	0.49	0.49	0.49	0.5	0.49	0.49	0.48	0.49	0.46
	Observations	1340	561	779	692	238	454	648	323	325
Intensity of treatment	Average	3.83	0.00	6.47	5.17	0	7.85	2.23	0	4.36
	St. Dev.	4.54	0.00	4.21	4.89	0	3.9	3.47	0	3.78
	Observations	1349	565	784	697	241	456	652	324	328
Intensity of (compulsory) treatment	Average	2.41	0.00	4.07	2.8	0	4.26	1.94	0	3.79
	St. Dev.	3.26	0.00	3.35	3.37	0	3.32	3.08	0	3.39
	Observations	1349	565	784	697	241	456	652	324	328
Female	Average	0.49	0.48	0.49	0.49	0.44	0.52	0.48	0.52	0.45
	St. Dev.	0.50	0.50	0.50	0.5	0.49	0.49	0.5	0.48	0.49
	Observations	1349	565	784	697	241	456	652	324	328

Table 4.2: Reform and Turnout

Dependent variable: Turnout

	OLS [1]	Logit [2]	OLS [3]	Logit [4]
intensity	.056*** (.018)	.045*** (.015)	.057*** (.018)	.045*** (.015)
non catalan origin			-.113* (.057)	-.142* (.073)
cat origin mixed family			-.084** (.039)	-.095*** (.045)
cat origin non catalan family			-.1291*** (.038)	-.144*** (.043)
YEARS OF EDUCATION	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES
PROVINCES	NO	NO	YES	YES
No. obs.	1248	1248	1248	1248
Rsq	.078		.099	
Pseudo-Rsq		.071		.093

Note: The dependent variable is a Dummy variable which is equal to 1 if the respondent voted during the regional elections. The specifications reported in Columns 1 and 2 include years of education and year of birth fixed effects. The specifications reported in Columns 3 and 4 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions.

Table 4.3: Reform and Ethnic voting (1)

Dependent variable: Probability of ethnic voting (among voters)

	CIU+ERC+IC OLS [1]	CIU+ERC+IC Logit [2]	CIU+ERC OLS [3]	CIU+ERC Logit [4]	CIU OLS [5]	CIU Logit [6]	CIU+ERC+IC OLS [7]	CIU+ERC+IC Logit [8]	CIU+ERC OLS [9]	CIU+ERC Logit [10]	CIU OLS [11]	CIU Logit [12]
intensity	.041*** (.018)	.046*** (.017)	.047** (.020)	.049** (.087)	.049*** (.020)	.053*** (.022)	.040** (.016)	.051*** (.017)	.046** (.019)	.057** (.023)	.049** (.020)	.052** (.022)
non catalan origin							-.359*** (.072)	-.408*** (.074)	-.399*** (.072)	-.423*** (.063)	-.169** (.075)	-.171** (.070)
cat origin mixed family							-.135*** (.042)	-.179*** (.054)	-.183*** (.021)	-.220*** (.053)	-.044 (.061)	-.046 (.061)
cat origin non catalan family							-.382*** (.045)	-.429*** (.051)	-.412*** (.051)	-.442*** (.050)	-.156*** (.048)	-.160*** (.047)
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCES	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
No. obs.	941	941	941	941	941	941	941	941	941	941	941	941
Rsqr	.057		.04		.054		.194		.194		.084	
Pseudo-Rsq		.046		.030		.042		.161		.153		.065

Note: The dependent variable is a Dummy variable which is equal to 1 if the respondent voted for a Catalanist party during the regional elections. In Columns 1-2, 7-8 we consider CIU, ERC and IC as Catalanist parties; in Columns 3-4, 9-10 we consider CIU, ERC as Catalanist parties; in columns 5-6, 11-12 we consider only CIU as a Catalanist party. We restrict the sample only to respondents who voted. The specifications reported in Columns 1-6 include years of education and year of birth fixed effects. The specifications reported in Columns 7-12 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions.

Table 4.4: Reform and Ethnic voting (2)

Dependent variable: Probability of ethnic voting (among all citizens)

	CIU+ERC+IC	CIU+ERC+IC	CIU+ERC	CIU+ERC	CIU	CIU	CIU+ERC+IC	CIU+ERC+IC	CIU+ERC	CIU+ERC	CIU	CIU
	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
intensity	.049*** (.015)	.057*** (.017)	.053*** (.017)	.059*** (.019)	.048*** (.016)	.053*** (.091)	.054*** (.014)	.068*** (.018)	.057*** (.016)	.071*** (.020)	.050*** (.016)	.055** (.019)
non catalan origin							-.363*** (.067)	-.347*** (.053)	-.381*** (.064)	-.335*** (.045)	-.182*** (.063)	-.158*** (.049)
cat origin mixed family							-.176*** (.043)	-.190*** (.044)	-.207*** (.045)	-.204*** (.040)	-.077 (.051)	-.073 (.045)
cat origin non catalan family							-.368*** (.038)	-.379*** (.035)	-.377*** (.043)	-.370*** (.037)	-.164*** (.041)	-.157*** (.035)
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCES	NO	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES	YES
No. obs.	1248	1248	1248	1248	1248	1248	1248	1248	1248	1248	1248	1248
Rsqr	.062		.053		.055		.174		.176		.087	
Pseudo-Rsq		.047		.004		.048		.134		.137		.075

Note: The dependent variable is a Dummy variable which is equal to 1 if the respondent voted for a Catalanist party during the regional elections. In Columns 1-2, 7-8 we consider CIU, ERC and IC as Catalanist parties; in Columns 3-4, 9-10 we consider CIU, ERC as Catalanist parties; in columns 5-6, 11-12 we consider only CIU as a Catalanist party. The specifications reported in Columns 1-6 include years of education and year of birth fixed effects. The specifications reported in Columns 7-12 include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions.

Table 4.5: Reform and voting patterns

Dependent variable: Voting choice (marginal effects)

	Abstain MLOGIT [1]	Ethnicvoting MLOGIT [2]	Non Ethnicvoting MLOGIT [3]	Abstain MPROBIT [4]	Ethnicvoting MPROBIT [5]	Non Ethnicvoting MPROBIT [6]	Abstain ASMPROBIT [7]	Ethnicvoting ASMPROBIT [8]	Non Ethnicvoting ASMPROBIT [9]
intensity	-.024* (.014)	.065*** (.017)	-.042** (.017)	-.022 (.014)	.065*** (.017)	-.043** (.017)	-0.022 (0.031)	.064*** (.021)	-.042* (.022)
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES	YES	YES	YES
No. obs.	1248			1248			1248		

Note: The dependent variable can assume three different values: (i) abstain (ii) ethnic voting or (iii) non ethnic voting. We run this specification first as a multinomial logit (Columns 1-3); as a multinomial probit, assuming the error terms to be independent (Columns 4-6), and also relaxing the assumption of independence of the error terms (Columns 7-9). We consider CIU, ERC and IC as Catalanist parties. Each group of three columns corresponds to a regression. All the specifications include years of education fixed effects, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%. Marginal effects evaluated at the average are reported for the logit regressions. Each group of three columns corresponds to a regression.

Table 4.6: The role of individual origin

<i>Dependent variable:</i>	<i>voting</i>		<i>ethnic voting (share)</i>		<i>ethnic voting (all)</i>	
	OLS	Logit (coeff)	OLS	Logit (coeff)	OLS	Logit (coeff)
	[1]	[2]	[1]	[2]	[1]	[2]
intensity*non cat.origin	.069*** (.022)	.355*** (.111)	.045* (.024)	.252** (.114)	.066** (.020)	.325*** (.100)
intensity*cat.family	.056*** (.018)	.268*** (.094)	.042** (.016)	.265*** (.084)	.054*** (.014)	.274*** (.071)
intensity*mixed family	.070*** (.018)	.366*** (.095)	.026 (.016)	.169** (.085)	.052*** (.016)	.263*** (.078)
intensity*non cat.family	.047** (.021)	.247** (.104)	.043** (.018)	.246*** (.090)	.054*** (.016)	.261*** (.081)
non catalan origin	-.136*** (.065)	-.992*** (.413)	-.36*** (.084)	-1.685*** (.398)	-.168*** (.062)	-.738*** (.265)
mixed family	-.146*** (.05)	-1.012*** (.322)	-.066 (.06)	-.404 (.321)	-.369*** (.054)	-1.599*** (.255)
non catalan family	-.091** (.043)	-.712** (.307)	-.385*** (.061)	-1.846*** (.305)	-.386*** (.079)	-1.675** (.379)
YEARS OF EDUCATION	YES	YES	YES	YES	YES	YES
YEARS OF BIRTH	YES	YES	YES	YES	YES	YES
PROVINCES	YES	YES	YES	YES	YES	YES
No. obs.	1248	1248	941	941	1248	1248
Rsq	.107		.197		.175	
Pseudo Rsq		.100		.165		.135

Note: In Columns 1 and 2, the dependent variable is a Dummy variable which is equal to 1 if the respondent voted during the regional elections; in Columns 3 and 4, the dependent variable is a Dummy variable which is equal to 1 if the respondent voted for a Catalanist party (we consider CIU, IC and ERC as Catalanist parties) during the regional elections (we restricted the sample only to voters); in Columns 5 and 6, the dependent variable is a Dummy variable which is equal to 1 if the respondent voted for a Catalanist party (we consider CIU, IC and ERC as Catalanist parties) during the regional elections (unrestricted sample). The specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. We classified respondents into 4 categories: 1) individuals who were not born in Catalonia 2) individuals who were born in Catalonia but whose parents were not 3) individuals who were born in Catalonia but with only one parent born in Catalonia and 4) individuals who were born in Catalonia whose parents were both born in Catalonia (omitted category). The specifications include also interactions between the variable intensity and each of the 4 categories. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. * Significant at 10%, ** significant at 5%, *** significant at 1%

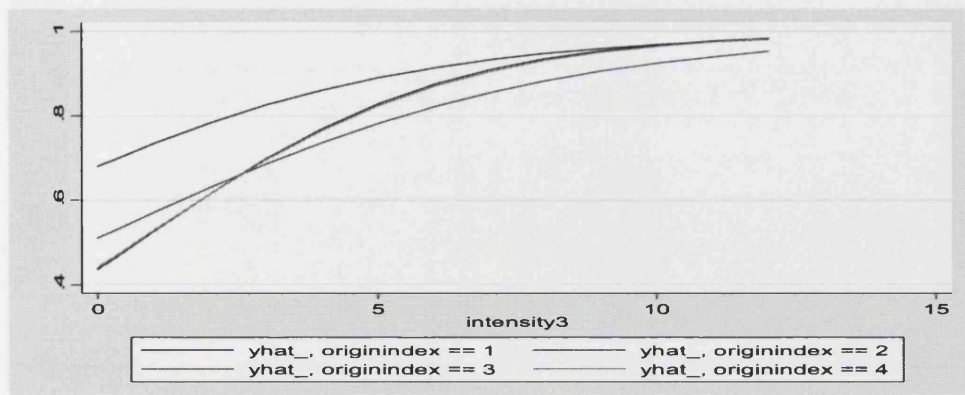


Figure 4.5: Probability of voting by intensity and origin: 1=catalan origin with catalan parents, 2=catalan origin with mixed parents, 3=catalan origin with parents that are not catalan, 4= not catalan origin. This corresponds to the logit regression in Column 2 of Table 4.6.

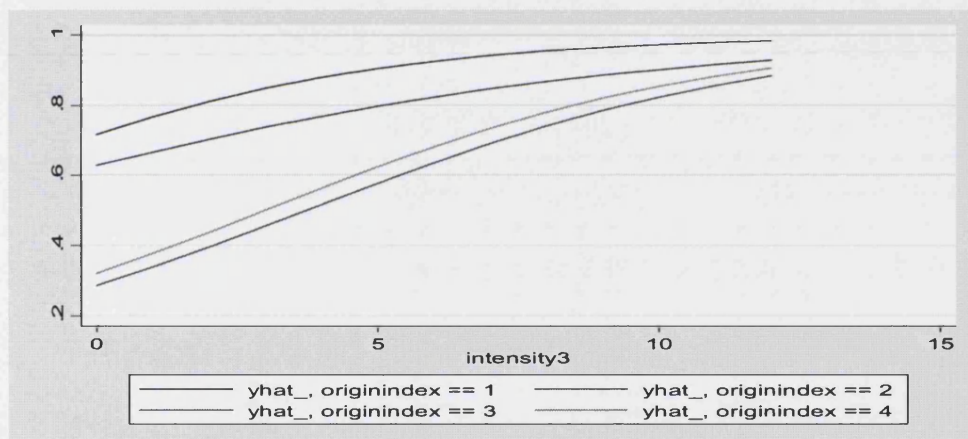


Figure 4.6: Probability of voting for an ethnic party (among voters) by intensity and origin: 1=catalan origin with catalan parents, 2=catalan origin with mixed parents, 3=catalan origin with parents that are not catalan, 4= not catalan origin. This corresponds to the logit regression in Column 4 of Table 4.6.

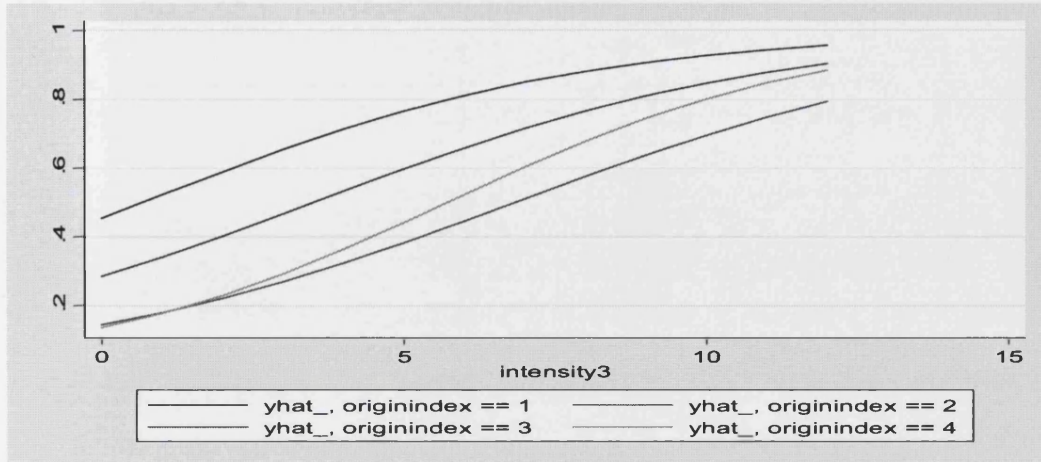


Figure 4.7: Probability of voting for an ethnic party (among all) by intensity and origin: 1=catalan origin with catalan parents, 2=catalan origin with mixed parents, 3=catalan origin with parents that are not catalan, 4= not catalan origin. This corresponds to the logit regression in Column 6 of Table 4.6.

Table 4.7: Robustness checks (1)

PANEL A												
	1934-1965	1934-1965	1934-1965	1934-1965	1934-1965	1934-1965	1921-1980	1921-1980	1921-1980	1921-1980	1921-1980	1921-1980
	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>
			(<i>voters</i>)	(<i>voters</i>)	(<i>all</i>)	(<i>all</i>)			(<i>voters</i>)	(<i>voters</i>)	(<i>all</i>)	(<i>all</i>)
	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
intensity							.049**	.032***	.044**	.049**	.056***	.066***
							(.020)	(.017)	(.021)	(.023)	(.019)	(.024)
pseudo-intensity	-.001	-.013	.017	.019	.016	.018	.002	-0.00	-.012	-.013	-.011	-.013
	(.008)	(.092)	(.013)	(.013)	(.012)	(.013)	(.011)	(.014)	(.018)	(.020)	(.017)	(.021)
No. obs.	862	862	734	734	862	862	1674	1674	1386	1386	1674	1674
Rsq	.337		.171		.163		.107		.178		.179	
Pseudo-Rsq		.097		.148		.130		.113		.153		.141
PANEL B												
	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>	<i>ethnic voting</i>
			(<i>voters</i>)	(<i>voters</i>)	(<i>all</i>)	(<i>all</i>)			(<i>voters</i>)	(<i>voters</i>)	(<i>all</i>)	(<i>all</i>)
	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit	OLS	Logit
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
intensity	.057***	.046**	.03	.04**	.047**	.06***	-.007	-.01	.003	.000	.008	.009
	(.019)	(.018)	(.018)	(.02)	(.019)	(.023)	(.017)	(.017)	(.012)	(.011)	(.015)	(.017)
Franco-intensity	.001	.002	-.021	-.024	-.013	-.017						
	(.017)	(.02)	(.018)	(.02)	(.02)	(.023)						
No. obs.	1248	1248	941	941	1248	1248	967	967	647	608	967	967
Rsq	.099		.194		.174		.108		.132		.129	
Pseudo-Rsq		.092		.161		.134		.093		.138		.099

Note: In the specifications reported in Columns 1-6 of Panel A we consider only cohorts who are not affected by the reform (1934-1965) and we assign a pseudo-treatment to the younger cohorts (1950-1965). In the specifications reported in Columns 7-12 of Panel A we consider cohorts born between 1921 and 1980 and the variable pseudo-intensity equals years of pseudo treatment for cohorts 1936-1950 and years of real treatment for cohorts 1966-1980. In the specifications reported in Columns 1-6 of Panel B we include the variable Franco intensity that measures the number of years for which individuals attended school once Franco had died but before the educational system became bilingual. The specifications include years of education, province fixed effects, year of birth fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. In Columns 7-12 of Panel B, we use the standard identification strategy, but a different dataset: we use representative survey data on the institutional attachment of residents in the Basque Country to the Spanish State.

Table 4.8: Robustness checks (2)
Endogeneity of years of education

	OLS	Logit	OLS	Logit	OLS	Logit	OLS
Dep var: Turnout	[1]	[2]	[3]	[4]	[5]	[6]	[7]
intensity	.054*** (.018)	.041*** (.015)	.025** (.010)	.19** (.009)			
intensity compulsory					0.001 (.016)	.001 (.014)	
language at home	.154*** (.054)	.155*** (.052)					
intensity (people arriving)							.035* (.019)
No. obs.	1240	1240	1248	1248	1248	1248	317
Rsq/Pseudo R sq	.11	.104	.067	.062	.061	.058	.503
	OLS	Logit	OLS	Logit	OLS	Logit	OLS
<i>Dep var: Ethnic voting (share)</i>	[1]	[2]	[3]	[4]	[5]	[6]	[7]
intensity	.035** (.015)	.045*** (.016)	.028** (.012)	.032** (.014)			
intensity compulsory					0.01 (.016)	.01 (.019)	
language at home	.242*** (.067)	.269*** (.066)					
intensity (people arriving)							.121** (.048)
No. obs.	935	935	941	941	941	941	195
Rsq/Pseudo R sq	.216	.182	.172	.139	.167	.167	.578
	OLS	Logit	OLS	Logit	OLS	Logit	OLS
<i>Dep var: Ethnic voting (all)</i>	[1]	[2]	[3]	[4]	[5]	[6]	[7]
intensity	.047*** (.013)	.062*** (.018)	.032*** (.011)	.037*** (.013)			
intensity compulsory					0.01 (.019)	.011 (.022)	
language at home	.299*** (.057)	.32*** (.056)					
intensity (people arriving)							.04** (.019)
No. obs.	1240	1240	1248	1248	1248	1248	317
Rsq/Pseudo R sq	.205	.158	.154	.117	.148	.148	.44

Note: The specifications reported in Columns 1 and 2 include cohort fixed effects. The specifications reported in Columns 3 and 4 include the age and the age squared of the respondents. The specifications reported in Columns 5 and 6 include the age and the age squared of the respondents. The specification reported in Column 7 includes cohort fixed effects and, as a measure of treatment, the number of years, within compulsory education, for which the respondents have studied in Catalonia after 1983. In this specification we restrict the sample to individuals who were not born in Catalonia. All the specifications include years of education fixed effects, province fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. * Significant at the 10%, ** significant at the 5%, *** significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level (Columns 1-4), cohort level (Columns 5-6) and cohort-year of arrival level (Column 7). Column 7 includes also dummy variables for the age of arrival. Marginal effects evaluated at the average are reported for the logit regressions.

Table 4.9: Robustness checks (3)

	<i>OLS</i> [1]	<i>Logit</i> [2]	<i>OLS</i> [3]	<i>Logit</i> [4]	<i>OLS</i> [5]	<i>Logit</i> [6]						
PANEL A: Restricted sample	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i> (voters)	<i>ethnic voting</i> (voters)	<i>ethnic voting</i> (all)	<i>ethnic voting</i> (all)						
intensity	.038* (.021)	.029* (.017)	.022 (.016)	.032* (.019)	.035** (.016)	.042** (.02)						
SAMPLE	1956-1975	1956-1975	1956-1975	1956-1975	1956-1975	1956-1975						
No. obs.	804	804	626	626	804	804						
Rsquared/Pseudo R squared	0.082	0.082	0.206	0.172	0.177	0.136						
	<i>OLS</i> [1]	<i>Logit</i> [2]	<i>OLS</i> [3]	<i>Logit</i> [4]	<i>OLS</i> [5]	<i>Logit</i> [6]	<i>OLS</i> [7]	<i>Logit</i> [8]	<i>OLS</i> [9]	<i>Logit</i> [10]	<i>OLS</i> [11]	<i>Logit</i> [12]
PANEL B: migration	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i> (voters)	<i>ethnic voting</i> (voters)	<i>ethnic voting</i> (all)	<i>ethnic voting</i> (all)	<i>voting</i>	<i>voting</i>	<i>ethnic voting</i> (voters)	<i>ethnic voting</i> (voters)	<i>ethnic voting</i> (all)	<i>ethnic voting</i> (all)
intensity	.045*** (.016)	.032** (.013)	.043*** (.015)	.057*** (.016)	.051*** (.012)	.065*** (.016)	.053*** (.019)	.041** (.016)	.051*** (.016)	.06*** (.016)	.062*** (.015)	.078*** (.019)
SAMPLE	1950-1981 all	1950-1981 all	1950-1981 all	1950-1981 all	1950-1981 all	1950-1981 all	1950-1981 catorigin	1950-1981 catorigin	1950-1981 catorigin	1950-1981 catorigin	1950-1981 catorigin	1950-1981 catorigin
No. obs.	1248	1248	941	941	1131	1248	1131	1114	857	857	1131	1131
Rsquared/Pseudo R squared	0.106	0.099	0.201	0.167	0.18	0.139	0.117	0.107	0.197	0.167	0.184	0.143

Note: All the specifications include years of education fixed effects, province fixed effects and dummy variables that control for gender, the origin of the respondent and of the parents of the respondents. In the specifications reported in Columns 1-6 of panel B we include also the fraction of respondents, within each cohort-educational group, belonging to each of the 4 origin categories. * Significant at the 10%, ** significant at the 5%, *** significant at the 1%. Robust standard errors are reported between parenthesis and are clustered at the cohort-years of education level. Marginal effects evaluated at the average are reported for the logit regressions.

Chapter 5

Conclusion

Overall, the evidence of this thesis is suggestive of the following:

- Using data from the World Values Surveys, we find that there is no negative relationship between ethnic diversity and average national identity. Once we distinguish between minority and majority groups, however, it is possible to claim that: on the one hand, in countries with low ethnic diversity individuals from minorities have stronger national sentiment than individuals from majorities; on the other hand, in countries with high ethnic diversity, individuals from minorities have weaker national sentiment than individuals from majorities. We then provide a theoretical model that builds on a recent contribution by Bisin et al (2006) and that helps explaining the empirical findings.
- We present evidence of a positive effect of the educational reform (whereby the education system became bilingual and Catalan -together with Spanish- was taught in schools) on Catalan feelings.
- The positive effect extends also to individuals whose parents have no Catalan origins. We conclude that multicultural policies tend to favour the development of regional identities while nation building policies tend to promote the expansion of a common national feeling.
- The introduction of a bilingual education system in Catalonia increased turnout and the likelihood of choosing a party who appeals to the electorate in ethnic terms.

Bibliography

- [1] Akerlof, G.A. (1980), "A Theory of Social Custom of which Unemployment May be One Consequence" *Quarterly Journal of Economics* 94, 749-775.
- [2] Akerlof, G.A. (1997), "Social Distance and Social Decisions", *Econometrica* 65, 1005-1027.
- [3] Akerlof, G.A. (2007) "The Missing Motivation in Macroeconomics" *American Economic Review* 97(1), 5-36
- [4] Akerlof, G.A. and R.E. Kranton (2000), "Economics and identity" *Quarterly Journal of Economics* 115, 715-753
- [5] Akerlof, G.A. and R.E. Kranton (2002), "Identity and Schooling: Some Lessons for the Economics of Education" *Journal of Economic Literature*, 40(4), 1167-1201
- [6] Akerlof, G.A. and R.E. Kranton (2005), "Identity and Economics of Organization" *Journal of Economic Perspectives*, 19, 9-32
- [7] Alesina, A., R. Baqir and W. Easterly (2000), "Redistributive Public Employment", *Journal of Urban Economics* 48, 219-241.
- [8] Alesina A, R. Baqir and B. Easterly (1999), "Public goods and ethnic divisions", *Quarterly Journal of Economics* 114(4), 1243-1284
- [9] Alesina A., A. Devleeschauwer, W. Easterly, S. Kurlat and R. Wacziarg (2003) "Fractionalization", *Journal of Economic Growth* , 8(2), 155-194
- [10] Alesina, A., E. Glaeser, and B. Sacerdote (2001), "Why doesn't the US have a European-style welfare state?", *Brookings Papers on Economic Activity*, 187-278.
- [11] Alesina A. and E. La Ferrara (2000), "Participation in Heterogeneous communities" , *Quarterly Journal of Economics* 115(3), 847-904
- [12] Alesina A. and E. La Ferrara (2005), "Ethnic diversity and Economic Performance", *Journal of Economic Literature* 43, 721-761

- [13] Amott, T. and J. Matthaei. (1996). "Race, Gender, and Work: A Multi-Cultural Economic History of Women in the United States" *Rev. ed. Boston, MA: South End Press*
- [14] Anderson B.(1983) "Imagined communities." *London, Verso*
- [15] Angrist, J., A. Chin and R. Godoy (2006) "Is Spanish-only schooling responsible for the Puerto Rican language gap?" *Journal of Development Economics* forthcoming
- [16] Angrist, J. and V. Lavy (1997) "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco" *Journal of Labor Economics*, 15, S48-S76 (January, Part 2)
- [17] Armstrong, J. (1982) "Nations before Nationalism", *Chapel Hill*
- [18] Ballester, C., Calv´o-Armengol, A. and Y. Zenou (2005), "Who's Who in Networks.Wanted: The Key Player," *CEPR Discussion Paper No. 5329*.
- [19] Bates, R (1983) "Modernization, Ethnic competition and the rationality of Politics in Contemporary Africa" in *State versus Ethnic Claims: African Policy Dilemmas*
- [20] Battu, H., M. Mwale and Y. Zenou (2005), "Oppositional Identities and the Labor Market," *CEPR Discussion Paper No. No. 5351*.
- [21] Bénabou, R. and Tirole, J. 2007. "Identity, Dignity and Taboos: Beliefs as Assets," *CEPR Discussion Papers 6123, C.E.P.R. Discussion Papers*
- [22] Bernheim, B.D. (1994), "A Theory of Conformity," *Journal of Political Economy* 102, 841-877.
- [23] Berry, C. (2002) "Piling on: The fiscal effects of jurisdictional overlap". *Unpublished manuscript*
- [24] Bhavnani, R. (2006) "Ethnic Norms and Interethnic Violence: Accounting for Mass Participation in the Rwandan Genocide" *Journal of Peace Research* 43:6 (2006): 651-669
- [25] Bisin A and T. Verdier (2000) "The economic of cultural transmission and the dynamics of preferences" *Journal of Economic Theory* 97, 298-319
- [26] Bisin A, E. Patacchini , T. Verdier , Y. Zenou (2006) "'Bend it like Beckham': identity, socialization and assimilation" *CEPR Discussion paper No 5662*
- [27] Bowles, S (1978) "Capitalist Development and Educational Structure" *World development*, 6, 783-796.

- [28] Breton, A (1964) "The Economics of Nationalism" *Journal of Political Economy*, 72, 4, 376-386.
- [29] Breton, A (1978). "Nationalism and Language Policies" *Canadian Journal of Economics*, 11(4): 656-668
- [30] Breuilly, J. (1982) " Nationalism and the State", *Manchester*
- [31] Caselli F. and W. Coleman (2002) "On the theory of ethnic conflict" *working paper*.
- [32] Chiswick, B and Miller, P (1999). "Language skills and earnings among legalized aliens" *Journal of Population Economics*, 1, 12
- [33] Coleman J and C. Rosberg (1964) "Political parties and national integration in tropical Africa" *Univ. of California Press, Los Angeles*
- [34] Easterly, W. and R. Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions", *Quarterly Journal of Economics* 112(4), 1203-1250.
- [35] Deutch, K. and W. Foltz (1963) "Nation building" *Atherton press, New York*
- [36] Epstein, G. S. (2006). "Migrants, Ethnicity and Employment". *unpublished manuscript*
- [37] Esteban, J and D. Ray (1994) "On the measurement of polarization" *Econometrica*, 62, 4, 819-851
- [38] Esteban, J and D. Ray (2007) "On the salience of ethnic conflict" *unpublished manuscript*
- [39] European Values Study Group and World Values Survey Association. EUROPEAN AND WORLD VALUES SURVEYS INTEGRATED DATA FILE, 1999-2002, RELEASE I [Computer file]. 2nd ICPSR version. Cologne, Germany: Zentralarchiv für Empirische Sozialforschung (ZA)/Tilburg, Netherlands: Tilburg University/Amsterdam, Netherlands: Netherlands Institute for Scientific Information Services (NIWI)/Madrid, Spain: Analisis Sociologicos Economicos y Politicos (ASEP) and JD Systems (JDS)/Ann Arbor, MI: Interuniversity Consortium for Political and Social Research [producers], 2004. Cologne, Germany: Zentralarchiv für Empirische Sozialforschung (ZA)/Madrid, Spain: Analisis Sociologicos Economicos y Politicos (ASEP) and JD Systems (JDS)/Ann Arbor, MI: Interuniversity Consortium for Political and Social Research [distributors], 2004.
- [40] Fearon, J. "Ethnic and Cultural Diversity by Country." (2003) *Journal of Economic Growth*, 8(2), 195-222.

- [41] Fernandez, R and G. Levy (2005) "Diversity and Redistribution " *NBER Working Paper No. W11570*
- [42] Fershtman, C. and Y. Weiss (1998), "Social Rewards, Externalities and Stable Preferences," *Journal of Public Economics* 70, 53-73.
- [43] Fishman, J (1968) "Language Problems of Developing Countries", *New York*
- [44] Folbre, N (1994) "Who Pays for the Kids: :Gender and the Structure of Constraints" *New York, Routledge*
- [45] Gellner E. (1983) "Nations and nationalism" *Basil Blackwell Ltd., England*
- [46] Glaeser, E.L. and J. Scheinkman (2001), "Measuring social interactions," in S.N. Durlauf and H.P. Young (Eds.), *Social Dynamics, Cambridge: MIT Press*, pp. 83-132.
- [47] Greif A. (1993) "Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition" *American Economic Review* 83(3), 525-548
- [48] Grin F (1999). "Economics" in Fishman J (ed) *Handbook of language and ethnic identity*. Oxford University Press, Oxford/ NewYork.
- [49] Hobsbawn, E.J. (1990) "Nations and Nationalism since 1780" *New York : Cambridge University Press, 1992*.
- [50] Hoffman, C (2000). "Balancing Language Planning and Language Rights Catalonia's Uneasy Juggling Act." *Journal of Multilingual and Multicultural Development*, 21 (5): 425-441.
- [51] Hroach, M. (1985) "Social Preconditions of National Revival in Europe" Cambridge 1985
- [52] Humphreys, M. and Weinstein, J. (2004). "What the Fighters Say: A Survey of Ex-Combatants in Sierra Leone", *CGSD Working Paper no. 20*, New York.
- [53] Jacob,S. (2004) "Public spending and the politics of identity", *unpublished manuscript*
- [54] Janeba,E (2004) "International trade and cultural identity" *unpublished manuscript*
- [55] Jacob,P and J.Toscano (1964) "The integration of political communities" *J.B.Lippincott, Philadelphia*
- [56] Kandel, E. and E.P. Lazear (1992), "Peer Pressure and Partnerships," *Journal of Political Economy* 100, 801-817.

- [57] Kevane, M " Can There Be an Identity Economics?" mimeo, *Harvard Academy for International and Area Studies*, 1994
- [58] Knack S. and P. Keefer " Does social capital have an economic payoff? A cross-country investigation" *Quarterly Journal of Economics* 112(4), 1251-1288
- [59] Knack, S. and P. Zack "Trust and growth" *Economic Journal*, 111, 295-321
- [60] Konya I. "A dynamic model of cultural assimilation" *unpublished manuscript*
- [61] Kuran, T. (1998). 'Ethnic Norms and Their Transformation Through Reputational Cascades', *Journal of Legal Studies* 27(2): 623-659
- [62] Jeffres, L. (1983) "Communication, Social Class and Culture" *Communication Research*, Vol. 10, No. 2, 219-246 (1983)
- [63] Johnson, H.G (1965). "A theoretical model of economic nationalism in new developing states" *Political Science Quarterly*, 80: 169-85.
- [64] Johnson,H.G (1975). "Technology and Economic Interdependence" *London McMillan*.
- [65] Jones, S. ""The Economics of Conformism" *Oxford & New York: Basil Blackwell, 1984*
- [66] La Ferrara E. (2003) "Kin groups and reciprocity: a model of credit transactions in Ghana" *American Economic Review* 93(5), 1730-1751
- [67] Landa, J "Trust, Ethnicity and Identity: Beyond the New Institutional Economics of Trading Network" *Ann Harbor: University of Michigan Press*, 1994
- [68] Lazear, E.(1999). "Culture and Language," *Journal of Political Economy*, 107(S6), S95-29
- [69] La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999)"The Quality of Government" *Journal of Law, Economics and Organization*, 222-279
- [70] Li X.(2005) "Ethnic diversity, social identity and tax compliance" *unpublished manuscript*
- [71] McCaa, R. (1993). "Ethnic Intermarriage and Gender in New York City". *Journal of Interdisciplinary History*, 24, 207-231.
- [72] Merlo, A (2007) "Whither Political Economy? Theories, Facts and Issues", in R. Blundell, W. Newey and T. Persson (eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Cambridge: Cambridge University Press. -

- [73] Miguel T and Posner D. (2004) "Sources of ethnic identification in Africa" *unpublished manuscript*
- [74] Miguel E.(2006) "Tribe or nation? Nation-Building and Public goods in Kenya versus Tanzania" *World Politics*
- [75] Milligan, K. , Moretti, E. and Oreopulos P. (2004) "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom" *Journal of Political Economy* 88, 1667-1695
- [76] Montalvo J. and Reynal M. (2005) "Ethnic diversity and economic development" *Journal of Development economics*, 76, 293-323
- [77] Montalvo J. and Reynal M. (2006) "Ethnic polarization, potential conflict and civil wars" *American Economic Review* 95 (3), 796-816
- [78] Padilla, A (2004). "Quantitative Methods in Multicultural Education Research." In J. Banks & C. Banks (Eds.), *Handbook of Research in Multicultural Education*. San Francisco: Josey-Bass.
- [79] Padro, G. (2006) "The Control of Politicians in Divided Societies: The Politics of Fear" *NBER Working Paper No. W12573*
- [80] Pagano, U (1995). "Can economics explain nationalisms? In Bretton, A , Galeotti, G, Salmon, P and Wintrobe, R (eds). *Nationalism and Rationality*" *Cambridge University Press, Cambridge*.
- [81] Pagano, U (2003). "Nationalism, development and integration: the political economy of Ernest Gellner" *Cambridge Journal of Economics*, 27: 623-646.
- [82] Posner, D (2004) "The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi" *American Political science Review* 98(4), 529-545
- [83] Roemer, J. (1998), "Why Don't The Poor Expropriate the Rich", *Journal of Public Economics* 70, 399-424.
- [84] Sen, A (1985). "Goals, commitment and Identity". *Journal of Law, Economic and Organisation*, 1: 345-355.
- [85] Smith (1983) "Theories of Nationalism", *London*
- [86] Stalin (1913) "marxism and the national and Colonial Question", *London*
- [87] Tilly, C (1975) " The formation of national States in Western Europe" *Princeton University Press, Princeton*

- [88] UN report (2004) "Building multicultural democracy"
- [89] Verwimp, P. 2005. "An Economic Profile of Peasant Perpetrators of Genocide: Micro-Level Evidence from Rwanda", *Journal of Development Economics* 77: 297-323
- [90] Watkins, M (1978). "The Economics of Nationalism and the Nationality of Economics: A Critique of Neoclassical Theorizing" *Canadian Journal of Economics*, 11: 87-120.
- [91] Wildsmith, E, M.Gutmann and B. Gratton "Assimilation and intermarriage for U.S. immigrant groups, 1880-1990" *The History of the Family* Vol. 8, Issue 4, 2003, 563-584