# TOPICS IN THE ECONOMICS OF CONFLICT

Dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics (XXI Cycle)

Olaf J. de Groot (21493)
Bocconi University

Thesis Committee:

Prof. Dr. Guido Tabellini, Bocconi University (co-chair)

Prof. Dr. Eliana La Ferrara, Bocconi University (co-chair)

Prof. Dr. Kristian Skrede Gleditsch, University of Essex

# Preface

This dissertation has been written as part of my effort to complete the PhD programme in Economics at Bocconi University in Milan, Italy. The dissertation's topic is the economics of conflict and it consists of four chapters, of which the first is only an introduction to the topics that are discussed in the further chapters.

In the second chapter, I introduce a measure of ethno-linguistic (dis)similarity between nations that I refer to as Ethno-Linguistic Affinity (ELA). I argue that using such a measure can be an improvement on previous work that looks at spill-overs of many types. I then continue to show that the proposed measure has a number of appealing characteristics, such as its clear interpretation. Following the theoretical discussion, I set up an actual version of the measure of ELA regarding Africa, which can be used to measure the spillover of conflict. Finally, using empirical data on conflict presence and conflict initiation, I show that conflict can indeed be concluded to be more likely to spill over between countries that have a high level of Ethno-Linguistic Affinity.

The third chapter looks at a different issue in the conflict literature. In this chapter, I consider the influence of conflict on economic growth in neighbouring countries in Africa. While this is an issue that has been explored before, I believe my effort is more rigorous and my results are therefore stronger. Previously, one of the underlying assumptions was that conflict is unambiguously bad for neighbouring states, whereas I believe this holds true only for directly contiguous states and not for proximate, non-contiguous ones. Additionally, I propose a new way of measuring distance, that combines decreasing returns to distance and the use of a database with distances of closest approach. Employing the newly proposed methodology, I conclude that contiguous nations indeed suffer from the presence of conflict, while non-contiguous, proximate nations in fact receive a small benefit. Including my measure of ELA reinforces the results.

In the fourth chapter, we look at a completely different aspect of conflict, namely the influence civil conflict has on the demand for education. First, we develop a theoretical model on how the presence of civil conflict could distort incentives, particularly in a relatively highly developed, but culturally distant, region. We show that, under relatively general circumstances, it can be argued that only individuals with a medium level of education are going to be changing their behaviour by increasing their desired level of education. Both highly- and lowly-educated individuals will be unaffected. We test this hypothesis empirically using the Basque Region in Spain as an example of a conflict in which there has been no supply-side effect in education and the demand-side effect is thus isolated. We use Census results to create a database with nearly four million individuals and employ a matching method to recreate a synthetic Basque Region with data from before the conflict. Then, when comparing the development of the true and synthetic regions after the civil conflict starts, we find evidence that supports our theoretical model.

## Acknowledgments

Many people have contributed to the development of this dissertation. I am particularly grateful to the members of my Dissertation Committee, Prof. Dr. Guido Tabellini, Prof. Dr. Eliana La Ferrara and Prof. Dr. Kristian Skrede Gleditsch. They have given me useful advice and supported the occasionally unusual ideas I proposed as dissertation chapters. I have also benefited from the discussion I have had with Dr. Elsa Artadi. Of course I am grateful as well to the board of Bocconi University, for their financial support during my thesis work, without which I could not have afforded to engage in my graduate studies.

Furthermore, I would like to express my gratitude for everyone who has read parts of the thesis and has given me useful (or less useful) comments on the different elements. Parts of the dissertation have been presented at conferences in Izmir, Turkey (April 2007 and April 2008), Utrecht, the Netherlands (October 2007), Milan, Italy (February 2008), La Marsa, Tunisia (June 2008), Ankara, Turkey (June 2008), Amsterdam, the Netherlands (July 2008), Copenhagen, Denmark (September 2008) and Maastricht, the Netherlands (November 2008), the participants of which have also helped me improve the different chapters. Two of the people who have particularly contributed to the development of this dissertation are my co-author and friend Idil Göksel (Bocconi University) and discussion partner and friend Jochen Mierau from the University of Groningen.

Finally, a special mention goes to Fedor van Rijn, who has greatly helped me by proof-reading an infinite number of versions of this dissertation and always being there for me when I need him most. Any remaining mistakes are, obviously, my own.

# Contents

# 1 Introduction

Conflict, whether international or civil, has traditionally been a topic studied more in political science and international relations. However, in recent years, attention from development and political economists to the topic has increased significantly. A number of important contributions in the relatively new field of conflict economics have been published during the past ten years. Among them are the excellent treatises of Collier and Hoeffler (2004), who analyse whether conflict is driven by greed or grievance, and Sambanis (2002). The highly relevant investigation of the relationship between conflict and economic development was subject of Collier (1999) and the excellent contribution by Miguel et al. (2004). Others who have contributed in this relatively young field of economics include Alberto Alesina, Dani Rodrik, Stergios Skaperdas, Amihai Glazer, William Easterly and Eliana La Ferrara.

With this dissertation, I hope to contribute to the literature on conflict economics, by proposing several new avenues of research and expanding on others. In particular, the first avenue that interests me is the spatial component of conflict. There are several studies that apply spatial econometric techniques to study some of these spatial impacts, most importantly the studies by Ward and Gleditsch (2002), Gleditsch (2007) and Murdoch and Sandler (2002a, 2002b and 2004), all of which are cited repeatedly throughout this dissertation. I am particularly interested in finding alternative ways of measuring distance and applying these in the economics of conflict. The second avenue of interest is related to the influence of conflict on individual incentives, about which I write in the final chapter.

It was my interest in alternative measures of distance that led me to write the second chapter. In it, I argue that measures of purely geographical distance are inappropriate in many contexts and that these distances ought to be augmented to include an ethno-linguistic component. After all, country studies have been including measures of ethno-linguistic heterogeneity for a long time (see e.g. Easterly and Levine, 1997, and Mauro, 1995), but the few times this has been done in cross-country studies, it has seemed unfocused and ad hoc. The way I am proposing to measure so-called Ethno-Linguistic

Affinity (ELA) between nations is partly based on the work of Laitin (2000), Fearon (2003) and Fearon and Laitin (2003). I assume that all groups can be described according to a number of region- and topic-specific dimensions that I refer to as identity characteristics. Using detailed data on these identity characteristics, the measure I propose is:

$$ELA = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\alpha_i \cdot \beta_j \cdot c_{ij}\right)$$

where $\alpha_i$ is the share of population that ethnic group $i \in I$ has in country A and $\beta_j$ is the share of population that ethnic group $j \in J$ has in country B. $c_{ij}$, finally is percentage of identity characteristics that are shared between groups $i$ and $j$. This measure has several attractive properties, not the least of which is its convenient interpretation: the expected percentage of shared identity characteristics of two individuals randomly drawn from two different populations.

Given my interest in conflict and Africa, I then continue to test the measure on that subject and area. For this, an unorthodox dataset is used: the Joshua Project (2007), which provides detailed data on small ethno-linguistic groups in Africa and across the world. The identity characteristics I choose to include in this measure of ELA are 1) The self-identified ethnic group, 2) The people cluster to which that ethnic group belongs, 3) The traditional language of a group, 4) The linguistic subfamily the group belongs to and 5) The religious sub-affiliation. These five characteristics are able to describe the different groups well and are potentially relevant in the conflict spill-over that may be taking place.

It should be remembered, however, that ELA is to be seen as an augmentation of purely geographic measures of contiguity and not as a replacement. For that reason, I perform an element-wise multiplication between my matrix of dyadic ELA-values and a matrix containing border lengths for all countries in Africa and use that outcome to construct the final contiguity matrix to be used in the analysis. The analysis itself consists of a series of logit Maximum Pseudo-Likelihood estimations on the probability of conflict initiation. Standard control variables, such as GDP level, population size and the number of years since the last conflict are included and the results clearly show that conflict in neighbouring, ethno-linguistically similar countries influences conflict initiation significantly. Moreover, when using either a purely ELA based or a purely border length-based contiguity matrix, there is no significant spill-over effect. Several robustness checks are unable to significantly alter the results.

In the third chapter, I approach a different element in the spatial econometric conflict literature: the influence of conflict on growth in other nations. This is exactly the topic that Murdoch and Sandler (2002a, 2002b and 2004) have worked on extensively, concluding that the negative effects of conflict affect a wide region around the core conflict nation in all specifications of their models. I have, however, some doubts about the conclusions reached by Murdoch and Sandler. First, they assume by construction that all nations are affected negatively, while there are strong arguments why nations that are proximate, yet non-contiguous, may actually benefit instead. Second, again by construction, the authors assume that all nations within their cut-off are affected in the same way, independent of the actual distance between the conflict nation and the affected nation.

For this chapter, I follow the basic ideas of Murdoch and Sandler, except when it comes to the problems mentioned before, which concerns the spatial econometric challenges. In particular, instead of one measure of spillover, I propose to use two different ones: one measure for directly contiguous nations only and one for all nations that belong to the generic neighbourhood of the conflict state. Like Murdoch and Sandler, I test the definition of 'generic neighbourhood' by using different cut-offs for the distance of closest approach, varying from 100 until 950 kilometres. Unlike Murdoch and Sandler however, I do not simply use a dummy variable for all countries that fall within the fixed distance, but instead propose a measure that decreases in distance:

$$\delta_{ij} = [(cut + 50) - (mindist_{ij} \mid mindist_{ij} < cut)]$$

with $mindist_{ij}$ as the distance of closest approach between $i$ and $j$. The cut-off value $cut$ is increased in steps of 50 km, from 100 to 950 km, leading to a total of 18 different minimal-distance contiguity matrices.

When testing my proposed spill-over channels, it turns out that the methods indeed work. Using a maximisation method similar to Murdoch and Sandler's, I find out that a 250 km distance of closest approach works best and that the effects are strongest for the heaviest conflict type. For that conflict type, per five-year period, a directly contiguous nation loses approximately 1 percentage point of GDP growth, while countries that are non-contiguous but whose distance of closest approach is within 250 km, benefit by a few percentage points, depending on the actual distance. A multitude of robustness tests do not significantly alter the results and the addition of the previously introduced measure of Ethno-Linguistic Affinity leads to a small improvement in the results.

The results from this chapter are relevant for two reasons. First, it provides evidence that there are situations in which proximate nations may in fact benefit from conflict, contrary to the conclusions drawn by Murdoch and Sandler. Second, it proposes a new way of using data on the distance of closest approach, which reduces the necessity of converting the data into dummy variables and losing a significant portion of the available information along the way.

The fourth chapter[1] of this dissertation looks at an entirely different aspect of conflict, and does not have a spatial econometric component. It looks at the influence of civil conflict on the demand for education. However, unlike previous work (e.g. Lai and Thyne), we are not interested in the supply-side effects that may take place due to a lack of teachers, difficulty in the physical access to schools, the necessity for students to work, etc. Instead, we look at the demand-side effects, relating this to the literature on brain gains and brain drains.

We start by exploring the suspected effects in a theoretical model. This naive model assumes innate ability for individuals, who then decide on their level of education. The level of education, finally, determines the wage level that one can expect. Additionally, one can decide whether to work in the home region or in a migration region, subject to a cost of migration that depends negatively on the level of education. Finally, there is a bonus to living outside your home region, which can be either positive or negative. We then show that, under relatively general conditions, the initiation of a small-scale civil conflict (not causing any supply-side effects in education) that increases the bonus of living outside the home region will increase the incentive for migration. However, it is particularly interesting to see that the difference in education demand takes place in one specific part of the ability distribution. It is the group with a medium level of ability who are more likely to migrate and increase their level of education. Neither the high-ability, nor the low-ability types show this same effect.

In order to study the veracity of our theoretical model, we look at the Basque Region in Spain, where a relatively low-scale civil conflict broke out in the late 1960s, and which fits the other characteristics that our model requires. The method we employ is inspired by Abadie and Gardeazabal (2003), who use the other Spanish regions to construct a synthetic Basque Region that has the same features as the true Basque Region before the conflict starts. They then look at the development of the true and synthetic regions after the conflict started and conclude that the conflict is responsible for a GDP loss of approximately 10%. However, contrary to Abadie and Gardeazabal, we are not looking at average outcomes, particularly because of our suspicion that different effects are at

[1] This chapter is co-authored with Idil Göksel of Bocconi University.

play at different locations within the distribution. Instead, we plan to replicate the entire educational distribution. But before doing so, it is important to take out other obvious explanatory factors that may explain some of the educational variance between different parts of the country.

We employ a dataset that consists of an amalgamation of the Census results from 1990 and 2000 and includes a total of six million observations. Approximately two million of these observations are ineligible, but with the rest, we perform an ordered probit analysis, that aims to take out the most obvious explanatory variables for educational achievements. What is left is the so-called Residual Education, for which we continue to calculate the average decile values for all Spanish Regions.

In the second stage, these regional decile values for all regions except the Basque Region are used as inputs to replicate each of the Basque Region's deciles between birth years 1910 and 1955. Most deciles use a variety of inputs to create the artificial region's distribution of Residual Education. For a sense of clarity, we then add up the different components in three different groups: the three lowest, the four middle and the three highest deciles. Graphically, it is obvious that the lower and higher deciles are indeed unaffected by the start of the civil conflict and education in the true and artificial regions continue to move very similarly. For the middle deciles, on the other hand, the true Basque Region's level of education is significantly higher than that of the artificial Basque Region.

Finally, we provide additional evidence for the proposed channel of these differences: migration. When performing a naive difference-in-difference analysis on people born in the Basque Region and born elsewhere in Spain, it is clear that Basque-born have become relatively more likely to migrate. At the same time, in concordance with our theoretical model, the average Basque-born individual has relatively increased his level of education. Most significantly, while the average Basque-born out-migrant has a higher level of education, it has decreased relatively compared to out-migrants from other regions, as forecast by our theoretical model.

# 2 Measuring Ethno-Linguistic Affinity Between Nations

**Abstract**

Research on ethno-linguistic ties has so far mostly focused on domestic measures of ethno-linguistic heterogeneity. Little attention has been given to the possibility that ethno-linguistic relations between countries may affect outcomes, particularly in a spatial econometric context. In this paper, I propose a way of measuring Ethno-Linguistic Affinity between nations. This new index measures the degree of similarity two randomly drawn individuals from two different populations can be expected to display. I show that this measure has a number of attractive theoretical characteristics, which make it particularly useful and continue to actually construct such a measure for all countries in Africa. Finally, using this measure of Ethno-Linguistic Affinity, I show that civil conflict in Africa is likely to spill over between contiguous ethno-linguistically similar countries.

**Keywords**: Ethno-Linguistic Heterogeneity; Spatial Econometrics; Conflict; Africa

**JEL code**: C21, F51, O10

## Introduction

Measures of Ethno-Linguistic Fractionalisation (ELF) and other types of ethno-linguistic heterogeneity, have been around for quite some time now. The most well-known contribution in this field is of course by Easterly and Levine (1997), who analyse the relationship between ethnic diversity and a range of economic indicators. They argue that a high level of ethno-linguistic fractionalisation may lead to strong rent-seeking and other growth-retarding policies. Easterly and Levine, however, like others publishing in this field, only concern themselves with the ethno-linguistic relationships within countries. In my opinion, on the other hand, ethno-linguistic (dis)similarities may also go a long way in explaining relationships between countries. In this paper, I propose a simple measure that should be able to improve many spatial macroeconometric analyses by including an ethno-linguistic component in addition to the common spatial parameter. After all, the original premise in spatial econometrics is that influence is exerted over space and that this influence reduces when the physical distance between observations becomes larger[2]. I agree with this statement, but I also believe that one should not merely consider physical distance, but include an ethno-linguistic component as well, when performing any kind of spatial macroeconometric analysis. In the current paper, after introducing this measure and showing how it can be constructed, I apply it in the field of conflict spillovers in Africa. Civil conflict is one of a number of events that is regularly analysed with a spatial component, but has so far received little attention in an ethno-linguistic spill-over framework, except for Alesina *et al*. (2006), who look at non-natural borders as a proxy for states that may share one ethnic group.

In the following section of the paper, I give some more background information on ethno-linguistic fractionalization, international ethno-linguistic relations and the spatial econometric literature on conflict spillovers. In the third section, I introduce my own measure, describe how it is set up and show its most important characteristics. In the fourth section, I show my empirical application in the field of spillovers of conflict and the final section concludes.

## Related Literature

### Ethno-Linguistic Fractionalisation Indices

As mentioned earlier, Easterly and Levine (1997) are the most well-known early adopters of a measure of Ethno-Linguistic Fractionalisation. Their measure of ELF measures the probability of two randomly drawn people from a population being part of two different ethnic groups: $1 - \sum_{j=1}^{J} \alpha_j^2$, where $j \in J$ are all different ethnic groups in a society and $\alpha_j$ is the share of population of group $j$. The data used in their analysis comes from the Soviet *Atlas Narodov Mira* from 1964, which has long been considered to be the most precise data available at the most disaggregated level. A significant disadvantage of this source, however, is the fact that it focuses strongly on linguistic differences and does not in fact take ethnic differences into consideration. As a result, the famous example of Rwanda is awarded an ELF of 0.14[3], as Hutus and Tutsis speak the same language and are thus considered to be one single group. Of course, history has taught us the flaw in

[2] This is the spatial version of saying that events from the recent past have a stronger influence than similar events from a more distant past, as one could say in time series analysis.

[3] According to Appendix 3 of Mauro (1995), Rwanda has a level of ethno-linguistic fractionalization of 0.14, which is very low compared to their neighbours Zaire/DRC (0.90), Uganda (0.90) and Tanzania (0.93), but similar to Burundi (0.04).

that assumption when up to 1 million were killed in the mid-nineties in ethnic violence between the Hutu majority and the Tutsi minority.

Easterly and Levine (1997) may be among the most famous for using a measure of ELF, but they were neither the first, nor the last to do so. Mauro (1995) is generally credited with being the first to introduce the measure to wider scientific interest in the field of economics, although it had been around for quite some time already. He uses the same ELF as Easterly and Levine do two years later, as an instrument for corruption to analyse its effect on growth in a cross-section of countries.

After Easterly and Levine, others have introduced alternative measures to deal with some of the drawbacks of the simple ELF used by earlier authors. Alesina *et al*. (2003) introduce several new measures for ethnic, linguistic and religious fractionalisation in a large set of countries. The first point they wanted to address is the problem that standard ELF focuses too strongly on linguistic groups, whereas these may not be the ties that distinguish groups the most from each other. They therefore calculate separate indices for religion, language and finally ethnicity. For the ethnicity data, the authors use an interesting approach, changing the definition of ethnicity per country. For example, they make racial distinctions in most of Latin America, whereas linguistic differences are used in Europe. The final important difference between the measures of Alesina *et al*. and previous measures of ELF is the source of the data. For their work, Alesina *et al.* use recent data, instead of the 1960s Atlas Narodov Mira. This has the advantage that the data is more detailed, more precise and possibly more trustworthy. It does, however, bring up the issue of endogeneity, where it can no longer be guaranteed that the ethno-linguistic composition of countries is independent of the outcomes that they are trying to measure (output, particularly). However, it is argued that such endogenous changes are extremely rare and that all measures of fractionalisation show very strong persistence over time.

Another alternative measure for ELF is discussed by Laitin (2000), Fearon (2003) and others, who take a more linguistic approach, based on Greenberg (1956). The use of distance in a tree diagram of languages reflects the expectation that languages that have branched out from each other more recently are expected to be more culturally similar. While the signal is recognised by Fearon to be noisy, he argues that the proposed fractionalisation measure is informative nonetheless: $1 - \sum_{i=1}^{I}\sum_{j=1}^{J}\pi_i\pi_j r_{ij}$, in which $\pi_i$ and $\pi_j$ are population shares and $r_{ij}$ is the so-called resemblance factor proposed by Greenberg. $r_{ij} \in [0, 1]$ measures how much two ethnic groups resemble each other, but the specification of this factor is still inconclusive. Greenberg, for example, wants to use the proportion of resemblances between each pair of languages on the most recent version of the glottochronology list. Unfortunately, the science of glottochronology has since lost most of its credibility and is no longer practiced on a large scale. Fearon, instead, proposes $r_{ij} = \left(\frac{l}{m}\right)^{\alpha}$, where $l$ is the number of linguistic levels shared between language $i$ and $j$, $m$ is the largest number of linguistic levels recorded in the dataset and $\alpha \in [0, 1]$. Laitin, finally, counts the number of branchings that are shared between languages according to the *Ethnologue* (Gordon, 2005) dataset.

Bossert *et al*. (2008) take all of the aforementioned indices of fractionalisation and combine them in order to create a Generalised Index of Fractionalisation (GELF). For this they use Census results that provide characteristics of the individual members of the population and they continue to calculate the level of fractionalisation. As shown in their contribution, this approach actually includes most measures of ELF and also possesses a number of desirable characteristics.

A completely different approach to Ethnic Fractionalisation that should be discussed here is proposed by Posner (2004) and only looks at politically relevant groups. He refers to his index as PREG (Politically Relevant Ethnic Groups) and it consists of a standard ELF, but instead of using all subgroups as is done in the traditional ELF literature, he takes only the politically relevant groups for each country, which are normalised to 100%. So, for example in Kenya, instead of using the 21 groups named in the *Atlas Narodov Mira* or the 64 groups mentioned by Gordon (2005), only the population sizes of the politically relevant Luo, Kalenjin and Kikuyu are used. Of course, while this may be an interesting measure, it endogenises the problem of having to decide which groups are the politically relevant ones in a country. Posner seems to have done this very carefully, but it leaves considerable room for criticism. In the final part of his paper, he replicates the results of Easterly and Levine and confirms that ethnic fractionalisation has a significant and strong negative impact on economic growth in his selection of African nations. A final problem with Posner's approach is that it does not cover issues that are caused by the difference between included and excluded groups. That is, certain political events that one is trying to research can be caused precisely by the fact that certain groups that are significant portions of the population are excluded from the political process.

Two further papers that use alternative approaches are Fearon and Laitin (2003) and Michalopoulos (2007). The first combine an array of quasi-standard ELF indices, with some unconventional measures, such as the number of distinct languages spoken by at least 1% of the population. Of course, this latter measure has typical problems of how to define distinct languages, and what makes a person to be a speaker of a particular language. Michalopoulos is worth mentioning because his results go in the opposite direction of the conventional wisdom. He argues that ethnic heterogeneity is the result of geographic heterogeneity and that, therefore, a measure of geographic heterogeneity can be used as an Instrumental Variable to analyse the influence between ELF and economic outcomes. According to his results, the strong effect found by previous authors was a spurious relationship and there is no real correlation between ELF and development.

Finally, many of the more recent authors[4] argue in favour of collecting more modern data. Often, the central thesis is that the potential endogeneity bias is of less concern than the actual problems with the data from the *Atlas Narodov Mira*. The specific data problems are manifold, but particularly the groupings, the definition of ethnic groups and the actual measurement of different group sizes have all been called into question. More recent data sources, on the other hand, are able to provide data that is more accurate and less biased in favour of any of the participating groups. Another advantage is the possibility of collecting the same data from different sources, comparing them and being able to arrive at a more precise estimate than when using a single source from 1964.

## International Component of Ethno-Linguistic ties

One thing that has received little attention in the literature so far, is the international component of ethno-linguistic ties. The relationship between nations has been subject of many studies, but in nearly every case pure geographical proximity is used as variable of interest[5]. For some particular purposes, such a contiguity-measuring variable may occassionally be augmented with a measure of economic interrelatedness, such as the size of total trade between nation dyads. However, strong arguments can be made in favour of using a measure of ethno-linguistic proximity, as augmentation of simple geographic proximity.

One can think of many instances, in which such ethno-linguistic ties exacerbate existing geographic connections. From conflict literature, there is the example of the conflict in Rwanda, which spilled over into Burundi, due to their shared ethno-linguistic ties,

[4] These include most of the aforementioned authors, as well as Roeder (2001), who compares older results with more modern data and also proposes different definitions of ethnic groups to come up with an array of indicators for ethno-linguistic heterogeneity. Annett (2001) uses data from the *World Christian Encyclopedia* to derive more precise estimates for both ethnic and religious levels of fractionalisation.

[5] Among those studies are Sambanis (2002), Murdoch and Sandler (2004) and Abreu *et al*. (2004), as well as Ward and Gleditsch (2002) and Gleditsch (2007), which are discussed in the following subsection.

whereas Uganda and Tanzania, both also bordering on Rwanda, were spared. Another example is the relatively large size of trade between Austria and Germany. This is not merely explained by the fact that these countries are contiguous, but the historical and ethno-linguistic ties between them explain why, *ceteris paribus*, people from these countries may have a preference for each other over their other neighbours. The existence of such cross-border effects, particularly in Africa, where borders have been randomly drawn by European colonisers, should not come as a surprise. It is possibly more surprising that no comprehensive measure exists to cover this issue.

One paper that does address the ethno-linguistic aspects of conflict spill-overs is by Buhaug and Gleditsch (2008), who research whether conflicts indeed spill over or whether they actually cluster in space due to the clustering of other factors that explain conflict. They conclude that transnational ethnic links are indeed a key element in conflict clustering. Unfortunately, these authors do not provide a thorough description of their measure of *ethnic linkages*, but it is one of only few studies that consider the ethno-linguistic ties in the field of conflict spillovers.

Another paper that includes an international dimension is by Alesina *et al*. (2006), who introduce an innovative way of measuring how artificial international borders are. They do this with a so-called fractal measure, according to the following procedure. For the border of a particular country, a grid is laid out on the border and the number of boxes within the grid that cross the border are counted. Afterwards, the grid size is increased and a new count is made. When this procedure has been repeated several times, the authors have a dataset containing box-sizes and box-counts for a particular border and when one then regresses the natural logarithms of these on each other as follows, coefficient $\beta$ will give an indicator of artificialness of a border:

$$\ln(box\_count) = \alpha + \beta \cdot \ln(box\_size)$$

The authors continue to combine this fractal measure with an indicator of "partitioned groups", defined as the percentage of the population of a country that belongs to a "partitioned group", where a partitioned group is defined as a group that appears in one or more adjacent countries as well[6]. Using these results, the authors then focus on using their measures as explanatory variables for economic and political success and accomplish satisfactory results. Unfortunately, Alesina *et al.* do not actually apply any of their measures on relations between countries, although they do mention that, with additional

[6] One issue Alesina *et al*. (2006) do not deal with carefully is group definitions. This could have a strongly distortive effect on their measure of "group partitioning" and should be discussed thoroughly.

work, research into international conflict could come from their line of research.

A final approach that is becoming more prominent recently is the use of genetic distances. First popularised by Cavalli Sforza *et al*. (1994), this way of measuring ethnic distances is used by a number of authors (e.g. Spolaore and Wacziarg, 2006, and Guiso *et al.*, 2007). Spolaore and Wacziarg carefully explain how they apply the genetic distance data in their paper. The measure for genetic distance supposedly measures how differently distributed the alleles of different populations are. Given that a population has a particular genetic profile, with particular distributions of the relevant alleles, it is possible to compare populations with each other and comment on their genetic level of similarity. The more different these allele distributions are, the further away these groups are from a common ancestor. The actual index is constructed as follows:

$$F = 1 - \frac{p_a q_a + p_b q_b}{2\overline{pq}}$$

where $p, q$ are frequencies of different alleles in populations $a$ and $b$ and $2\overline{pq} = 1 - \left[\left(\frac{p_a+p_b}{2}\right)^2 + \left(\frac{q_a+q_b}{2}\right)^2\right]$ [7].

While this is an interesting measure, with a strong scientific basis, it faces two problems. Firstly, populations may be genetically similar despite belonging to different ethno-cultural groups, particularly when considering cultural features that have not yet existed for a long time (such as religion). The second issue is the fact that such data is only available at a highly aggregated level. In total, only 42 population groups are available, which are supposed to capture all of Earth's population. It would seem reasonable to say that ethnic competition or cooperation takes place at a more disaggregated level.

## Causes of Civil Conflict

For now, it seems the debate on the causes of civil conflict has converged on the greed versus grievance theory of Collier and Hoeffler (2004), who argue that there can be two main sources for civil conflict. The grievance caused by an atypically unfair distribution of wealth or power or another kind or repression of a significant minority, appears to have relatively little explanatory power, whereas the greed explanation of opportunity for rebellion has much stronger results in regressions regarding the occurrence of

[7] Actually, the authors claim that $2\overline{pq} = 1 - \left(\frac{p_a+p_b}{2}\right)^2 + \left(\frac{q_a+q_b}{2}\right)^2$, which must clearly be a typo, even though they repeat the same mistake in equations 18, 19 and 22. The final results, however, are consistent.

civil conflict. However, not everyone agrees and there are still authors who continue to argue that an ethnic component may play an important role in the occurrence of conflict. Whether this is purely grievance-based remains a question, because there may be greed-based explanations related to ethnic division as well. More particularly, Fearon and Laitin (2003) investigate the greed versus grievance issue in their own way and they conclude that while grievance may be a small source, it is mostly economical reasons that cause civil war and not ethnic heterogeneity, in any of the ways they measure it. However, while Fearon and Laitin look at several aspects of ethno-linguistic heterogeneity, they do not approach one source of conflict in ethnic relations. This concerns the conflict between insiders and outsiders, which is an issue that relies both on heterogeneity, but also on the way heterogeneity is represented in the political system. The untested theory in this case is whether countries of which the governments are less representative of the ethno-linguistic heterogeneity, are more likely to suffer conflict.

Montalvo and Reynal-Querol (2005) put forward another strong critique of previous analyses of the relationship between ethnic (or religious) heterogeneity and conflict. They convincingly argue that it is not ethno-linguistic or religious fractionalisation that matters for the probability of conflict, but polarisation. In their excellent contribution, the authors show that their proposed index measures polarisation properly and is also compatible with a discrete version of the generalised polarisation index proposed by Esteban and Ray (1994):

$$EP = 4\sum_{j=1}^{J}\pi_j^2\left(1-\pi_j\right)$$

In their paper, Montalvo and Reynal-Querol show the excellent properties of this simple index and then continue to show that, contrary to previous results, ethnic polarisation is relevant in predicting civil conflict. Economic variables remain important, but ethnic polarisation (and to some degree: religious polarisation) plays an important role as well.

Of course, there are many other variables that are generally used to analyse conflict probabilities as well. These include the percentage of rough terrain, population density, population size, democratic freedoms and dependence on primary exports (particularly oil). I return to this list in section 4, where I run my own regressions regarding civil conflict.

One last feature of civil conflict that is relevant in the context of this paper, however, is the existence of spill-over theories. Some papers have included different kinds of geographical features in their conflict analyses, including a few that have followed the same line of reasoning that I apply in section 4, in that (civil) conflict is more or less likely to spill over from one country to another. Among the most relevant references in this case are Ward and Gleditsch (2002) and Gleditsch **(**2007**)**. In their excellent contribution, Ward and Gleditsch (2002) use Markov Chain Monte Carlo estimations to show that conflict spillovers occur and they are able to correctly forecast a significant number of conflicts.

There are several reasons why international ethno-linguistic linkages may affect the initiation of conflict. These include a revisitation of the grievance literature, which in an international context would argue that when majority group A slaughters minority group B in country 1, majority group B in country 2 may want to take revenge on minority A. Additionally, a conflict spillover could also be the direct result of the export of combatants, if refugees in a neighbouring country choose to continue their battles there. Another potential source of spillover is found when neighbouring populations are inspired by a conflict. For example, with 2 neighbouring countries in which majority group A represses minority group B, a succesful uprising by B in country 1 may lead the people in country 2 to be equally inspired to rise against their oppressors. A final theory on how ethno-linguistic linkages could play a role in conflict is related to precariously stable nations. Imagine a country consisting of two fairly balanced ethnic groups who share power in a reasonable way. Now, due to neighbouring conflict, one of the groups becomes more dominant, which can cause them to use the opportunity to repress the other group, which can lead to conflict.

Gleditsch (2007) argues that it is surprising how underlit the transnational dimensions of civil conflict actually are. He suggests there are several ways through which transnational links may influence civil conflict outcomes. Shared ethnic links, spillovers of autocratic tendencies and economic ties. When regressing both domestic and transnational features on conflict, and using Maximum Pseudo-Likelihood (MPL) techniques to approximate the likelihood function, he concludes that his measure of ethnic linkages is indeed significant.

## A Measure of Ethno-Linguistic Affinity

In this paper, I propose to construct an index that I shall refer to as a measure of Ethno-Linguistic Affinity (ELA) between nations. Such an index can be used to augment existing distance measures to include both geographic and ethno-linguistic ties, which is important in order to achieve more accurate results regarding the existence of spill-over effects. In my opinion, there is a strong argument to be made in favour of saying that ethno-linguistic ties between nations are a factor that could strongly improve such research. Of course, there are natural phenomena which are driven purely by geographical proximity. Spatial correlations may be found due to similar weather patterns (consider Miguel *et al*., 2004, for example) or due to regional resource abundance (consider the Middle East, for example. In other cases, neighbouring states may be influenced due to direct spillovers. In this context, think of the increased economic growth in Northern Ireland resulting from increased demand from the Republic of Ireland during the 1990s. Another example could be the recent oil-driven boom in Russia. Culturally close Belarus has benefitted from this economic improvement, while culturally distant Mongolia, despite its long Russian border, has not.

Obviously, it could be that economic ties are a determining feature in the relationship between countries. However, this is not necessarily the case. While in the previous examples economic ties are a relevant channel, for spillovers of anything else than growth (e.g. conflict), other channels may play a role too. But even for purely economic spillovers, there could be other channels than simply trade. Ethno-linguistic ties could also play a role in how strongly one country responds to events in their neighbouring countries. Finally, a significant problem with using economic ties to analyse spillovers is the endogeneity of the measure. After all, when trying to analyse when e.g. economic growth in one country spills over into another country, the use of economic channels confuses the analysis, as the existence of the channel may be the source of growth in the first place. It is also important to remember that such an analysis would not answer the question why the economic ties are there in the first place. I am arguing that, while economic relations may play a role in all kinds of spill-over effects, these economic ties are the result of ethno-linguistic ties between nations. Therefore, measuring the ethno-linguistic ties between nations and combining that with geographic spillover analyses makes more sense, because it captures both spillovers that stem directly from the ethno-linguistic ties and the spillovers that happen due to ethno-linguistically induced economic ties.

Of course, this leads to the problem of measuring ethno-linguistic ties. One thing that one could do is related to what Alesina *et al*. (2006) did and consists of looking at dyads of countries and measuring the percentage of the population that belongs to ethnic groups existing in both countries. However, this imports many of the problems that haunt the ELF literature, particularly group definitions. When using this method, it is very important to decide on which level of disaggregation the ethnic groups are measured. Considering northern Africa, the measure will give very different results when e.g. Berbers are considered one group or whether all individual Berber clans are considered separately. Additionally, the strict boundaries between ethnic groups are simply unrealistic, both in theory and in practice, for measuring purposes.

Instead, I propose an alternative way of measuring ELA. The first step in the construction of this measure is the recognition that ethnic identities consist of a number of different so-called *identity characteristics*. One could argue that different historical periods and different regions of Earth may require a different set of identity characteristics and I will therefore not define them yet. However, the kind of characteristics that one could think of are race, national origin, language, religion, clan identification, et cetera. An important feature of these identity characteristics is their cumulative nature: the more characteristics shared between two individuals, the more ethno-linguistically similar they are. Assuming that one is able to come up with a satisfactory set of identity characteristics, it is easy to see that it should be possible to classify all ethnic groups within a population according to them. Particularly, when using a very disaggregated ethnic dataset, it is possible to strictly identify each of the different characteristics that make up a particular ethnic group. This solves another problem from the ELF literature: how an ethnic group is defined. With this method, it can be recognised that two groups are highly similar, while still recognising their individuality. In fact, as I will mention later, using this same technique it is also possible to set up a measure of within-nation ethno-linguistic fractionalisation that does not have the problem of having to choose a level of aggregation of the data, and combines different features that one might deem important. Of course, such a measure is automatically closely related to Fearon's (2003) measure.

After having constructed a dataset of the region of interest that consists of ethnic groups at the most disaggregated level possible with values for each of the different identity characteristics, one can construct a measure that incorporates both the sizes of the different ethnic groups and how different these groups really are. After all, groups $i$ and $j$ that share all-but-one of their characteristics are more likely to feel affinity towards each other than groups $i$ and $k$, who share only one of these characteristics. The measure I am proposing to use is the following

$$ELA = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\alpha_i \cdot \beta_j \cdot c_{ij}\right)$$

where $\alpha_i$ is the share of population that ethnic group $i \in I$ has in country A and $\beta_j$ is the share of population that ethnic group $j \in J$ has in country B. $c_{ij}$, finally, is the percentage of identity characteristics that are shared between groups $i$ and $j$. This parameter can be anywhere between 0, if the two groups have nothing to do with each other, and 1, if they are actually the same group but live in different countries. This measure is closely related to the way Greenberg (1956), as reproduced by Fearon (2003), proposed to construct an alternative measure of ELF, except that it involves the relationship between countries instead of measuring just within one country and that I have approached the definition of their *resemblance factor* in a different way. In fact, it is easily possible to apply the same type of resemblance factor they use, instead of my identity characteristics. However, I feel that linguistic distances alone do not appropriately capture the entire arena of ethno-linguistic ties that one can describe with my proposed identity characteristics. On the other hand, it would also be easy, and feasible, to use my identity characteristics as their resemblance factor and come up with an alternative measure of ELF[8].

An advantage of this measure is its clear interpretation, similar to that of the ELF. Remember the interpretation of the ELF is *the probability that two individuals randomly drawn from a population are of the same ethnic group*[9]. This measure of ELA, on the other hand, measures *the percentage of shared identity characteristics of two individuals randomly drawn from two different populations*. In other words, how much affinity can a random person from country $A$ be expected to have with a random person from country $B$.

In a practical application, however, it is probably rare to expect that Ethno-Linguistic Affinity is the only channel through which spillovers take place. In most conceivable examples, a combination of the Ethno-Linguistic and Geographic channels should be expected. A combination of these two channels is very easy, when using standard spatial econometric techniques. When setting up a contiguity matrix, one can simply multiply each of the observations for geographic distance with the corresponding measure of ELA, before performing the required row-normalisation[10]. Like in other spatial econometric analyses, the kind of geographic distance measure is still open to debate, but this technique works, independent of whether center-point distances, distances of closest approach, border-lengths or another measure of contiguity are used.

---

[8] Such a measure would look as follows: $ELF = \sum_{i=1}^{I}\sum_{j=1}^{J}(\pi_i \cdot \pi_j \cdot c_{ij})$, with $\pi_i$ the percentage of group $i$ in the total population and $c_{ij}$ still the shared percentage of identity characteristics. Such a measure would take the linguistic focus of the measure Greenberg (1956) and Fearon (2003) propose and yield a measure that is not as perceptible to definition changes.

[9] Actually, ELF usually measures the probability that two individuals are from different ethnic groups, but to show the similarity between the measures, the current interpretation is more convenient.

[10] Appendix A shows how such contiguity matrices are set up.

Another thing to remember is that, so far, this measure has a range of potential applications and gives the researcher a lot of room for adjusting it to a suitable situation. The set of *identity characteristics* is, so far, undefined and can be chosen in order to accommodate the particular issue that is being researched. Channels of Ethno-Linguistic Affinity can be expected to differ strongly, depending on time and space. Examples of characteristics that may be relevant in some regions, but not in others include clan affilition (in Africa), caste (in India) or ancestry (in North America). This implies that it is important to decide which are the identity characteristics that fully describe the type of ethno-linguistic group association one is trying to measure.

## Features of the ELA measure

Looking at the way the measure has been constructed, it appears to possess many appealing characteristics. First of all, the range of the measure is linear and clearly defined: $ELA \in [0, 1]$, where $0$ means that the populations of two nations have absolutely nothing in common and $1$ means that the two countries have completely homogeneous native populations, who share all their identity characteristics. There are two ways in which a country dyad can have a lower level of ELA. First, the populations can become more different. For example, two completely homogeneous nations, of which the two population groups share only half the identity characteristics is going to yield an ELA of $0.5$. After all, without any uncertainty, two randomly drawn individuals will always share 50% of their characteristics[11]. The second way is when, instead of two equal homogeneous populations, the two countries both consist of the same two, equally-sized, ethnic groups that do not share any identity characteristics with each other. This would also lead to an ELA of $0.5$, because there is a 50% probability of drawing two completely equal individuals and 50% probability of drawing two completely different individuals. The expected value is therefore $0.5$.

Another desirable characteristic of the measure is its divisibility. After all, the current measure is simply a sum of the separate distributions of the different identity characteristics:

$$ELA = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\alpha_i \cdot \beta_j \cdot c_{ij}\right) = \frac{1}{C}\sum_{c=1}^{C}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\alpha_i \cdot \beta_j \cdot (1 \mid c_i = c_j)\right)$$

[11] In fact, the condition that both countries have a completely homogeneous population is unnecessary, as long as $c_{ij} = 0.5 \; \forall i, j$, the level of Ethno-Linguistic Affinity between the two countries will always be 0.5.

where $c \in C$ are the different identity characteristics and $c_i$ is the value of identity characteristic $c$ for population group $i$.

A final attractive feature of this measure is the fact that the value of ELA does not change when a particular ethnic group is subdivided incorrectly. Measuring ethnic groups at a subdivision that is more detailed than strictly necessary will not change the value of ELA. After all, when several small groups share the same identity characteristics, these are summed up again when calculating the actual measure[12]:

$$\begin{aligned}
ELA &= \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\alpha_i \cdot \beta_j \cdot c_{ij}\right) = \sum_{i=1}^{I}\left[\left(\sum_{j=1}^{J-1}\left(\alpha_i \cdot \beta_j \cdot c_{ij}\right)\right) + \left(\alpha_i \cdot \beta_J \cdot c_{iJ}\right)\right] = \\
&\sum_{i=1}^{I}\left[\left(\sum_{j=1}^{J-1}\left(\alpha_i \cdot \beta_j \cdot c_{ij}\right)\right) + \left(\alpha_i \cdot \beta_{J_1} \cdot c_{iJ_1}\right) + \left(\alpha_i \cdot \beta_{J_2} \cdot c_{iJ_2}\right)\right] \\
\text{if } \beta_{J_1} + \beta_{J_2} &= \beta_J \text{ and } c_{iJ} = c_{iJ_1} = c_{iJ_2}
\end{aligned}$$

In fact, it is reasonable to say that for each identity characteristic the distribution over different groups is actually irrelevant. More particularly, the sub-measure for a single identity characteristic, where $k \in K$ are the different values a particular identity characteristic can take can be summarised as following:

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\alpha_i \cdot \beta_j \cdot (1 \mid c_i = c_j)\right) = \sum_{k=1}^{K}\left(\min\left(\alpha_k, \beta_k\right)\right)$$

where $\alpha_k$ and $\beta_k$ are the population shares that $k$ has in nations A and B respectively.

## Practical Construction of Measure

In this subsection, I set up a measure of Ethno-Linguistic Affinity between nations in Africa, which is used to analyse the spill-over effects of conflict in the following section. To construct my measure, I utilise an unusual source that does not seem to have been used often before: The *Joshua Project* (2007). This was a project originally started in 1995 and is currently an official ministry of the *U.S. Center for World Mission*, an evangelical organisation aiming to spread the word of their religion to the so-called

[12] In fact, like Bossert *et al.* (2008) show in the case of a measure of ethno-linguistic fractionalisation, the optimal result is achieved when using actual individuals instead of groups. However, it is imply unfeasible to have such detailed data available for any sizeable group of countries.

"unreached peoples of the Earth". While this is an unorthodox source, one can make strong arguments in favour of using this particular one. The data provided by the *Joshua Project* is extremely detailed and seems to combine many of the sources used in other papers[13,14], with an extensive local network that is able to provide more detail from a local perspective. Often, religious data may be questionable in its veracity, but the stated goal of the *Joshua Project* shows why this data is worth using: "The mission of *Joshua Project* is to help bring definition to the unfinished task of the Great Commission by identifying and highlighting the people groups of the world that have the least exposure to the Gospel and the least Christian presence in their midst"[15]. The religious fervency with which this organisation collects data works in our advantage. After all, no religion would want to underestimate their own follower base, but this project especially is trying to analyse which particular groups need their "help" the most, and therefore, it is also imperative not to overestimate their own following either. In fact, the *Joshua Project* is clearly best-served with true and correct data. Of course, one should not trust blindly, and where possible, I have consulted alternative sources to see whether the data provided by the *Joshua Project* was compatible and by and large, this did seem to be the case. Nowadays, the most popular source is the *World Christian Encyclopedia* (Barrett *et al.*, 2001) and in order to check the compatibility of that source and the Joshua Project, I have tried to match the entries from each of these sources. This process is not very easy, because different names are used for the same population groups and the level of detail differs per source as well. However, despite these problems and despite the difference in the time frame of the different sources, the correlation coefficient between the different entries is approximately 0.96. This re-enforces my premise that the *Joshua Project* is a valid data source.

A large advantage of the *Joshua Project* data is its amazing level of detail. Ethnic groups are split into micro-groups, as a result of which you do get a proper overview of all the information available. Most other sources (and other papers) use only groups that contain at least 1% of the population, but with my measure of ELA, this would not be convenient. After all, when calculating a traditional ELF-index, ethnic shares are multiplied with themselves and a group that is smaller than 1% of the population exerts influence of less than $0.01 \times 0.01 = 0.0001$ on the total ELF. However, with my index of ELA, in the most extreme case, where the 1%-group of one nation is 100% compatible with

[13] Including Ethnologue, the World Christian Encyclopedia, the CIA World Factbook, PeopleGroups.org and Harvest Information System.
[14] Of course, this source is also usable when analysing other world regions. Additionally, if one were to set up a domestic version of this measure, as suggested in section 3, any data required are available in there as well.
[15] From www.joshuaproject.net

100% of the population of a neighbouring state, the total influence would be significant at $0.01 \times 1 = 0.01$[16]. For Africa, the *Joshua Project* reports results for 3704 country-groups[17] and it is this level of detail that outweighs the problems this source may inherently contain. Of course, this does not answer any of the standard questions regarding endogeneity. When researching the influence of ethno-linguistic composition on some macro effect (particularly civil conflict), one should use the ex ante ethno-linguistic composition, as the ethno-linguistic composition may have been endogenously determined by the occurrence of conflict or anything else you are trying to measure. However, it has been argued in the past that due to the strong level of persistence among ethnic composition, one does not need to worry much about this problem. Roeder (2001) shows that, particularly in Africa, the ethnic composition persistence is indeed very high and I use this as a basic assumption to be able to continue with this dataset.

When an appropriate dataset is found, the foremost issue that comes to mind when using the previously proposed way of setting up a measure of ELA, is the recognition of the relevant identity characteristics. Given that the aim of this exercise is to explain conflict spillovers and the geographic area of interest is Africa, this already points in the direction of the type of characteristics that one should be looking for. They are largely determined by ethnic characteristics, which are unfortunately typically hard to classify. A first measure, however, is the self-identified (internationalised) ethnic group. This is the most basic level of ethnic affiliation and is directly connected to the second identity characteristic, original language spoken by an ethnic group. While these two characteristics are likely to be strongly correlated, they capture in fact two different aspects, because different groups may speak the same language and in extraordinary cases, the same group may be speaking different languages in different nations.

These first two characteristics are at a very micro level, but it is important to try and capture the interconnections of these groups at a slightly higher level as well. For this, I use macro-measures for both of the first two identity characteristics. In the case of linguistic ties, I follow Greenberg's (1956) theory that languages that have split more recently belong to ethnic groups that are more closely related and have therefore used the *Ethnologue* (Gordon, 2005) and *Rosetta Project* (2007) databases to construct separate linguistic groups at the level of a sub-family. The subfamily of Atlantic Congo (family: Niger-Congo), however, since it is so prevalent in Africa, turned out to contain a majority of the country-groups and of the population involved, so I have split this subfamily into

---

16 Of course, this example is extreme but for example the Ukrainians make up some 72% of the population in Ukraine. In nearby countries like Georgia and Kyrgyzstan, Ukrainians indeed make up around 1% of the population, so the influence of this relationship on the total level of affinity is still quite significant.

17 The whole world contains a total of 15,965 country-groups.

1.Summary statistics for different identity characteristics

| | categories | average | | largest category | |
|---|---|---|---|---|---|
| | Nr. | groups (nr.) | ppl (mln) | groups (nr.) | ppl (mln) |
| Language | 2079 | 1.8 | 0.5 | 83 | 45.7 |
| Linguistic group | 65 | 57.0 | 14.4 | 906 | 259.5 |
| Ethnic group | 2435 | 1.5 | 0.4 | 53 | 44.4 |
| People Cluster | 98 | 37.8 | 9.5 | 319 | 65.9 |
| Religion | 22 | 168.4 | 42.4 | 988 | 378.4 |
| Total | 3704 | 1 | 0.3 | 1 | 43.4 |

*Note: These summary statistics include the size of the average and largest categories within an identity characteristic. For the largest category, the number of groups and the number of people are not necessarily in the same category (for example, while the largest number of groups can be found in the Benue people cluster, it is the Egyptian people cluster that contains the largest number of people).*

smaller sections[18], following Gordon and the *Rosetta Project*. In Appendix B, there is an overview of all the language groups and the percentages of the population beloninging to them[19].

The ethnic equivalent of this last characteristic is a division made according to "People Cluster". This term is posited by Johnstone (2007) and defined as "[a] smaller grouping of peoples within an affinity bloc, often with a common name or identity, but separated from one another by political boundaries, language or migration patterns", where an affinity bloc is defined as "[a] large grouping of peoples related by language, history, and culture, and usually indigenous to a geographical location". Johnstone's objective in the construction of the measure of People Clusters is a simplification of the task of evangelisation. While not his original aim, he does provide a framework for the logical clustering of all these ethnic groups that makes the list of available groups significantly smaller. As can be seen in Appendix C, the total number of People Clusters in Africa is 98 (after merging some non-native African groups that were too small to exist on their own) and this measure truly seems to capture an appropriate subdivision of all different ethnicities in Africa.

The final characteristic I use is one that moves away from evolutionary development over

18 To be precise, the Atlantic Congo subfamily has been split in Atlantic, Ijoid and Volta-Congo, where the last was split into Dogon, Kru, Kwa, Northern Languages and Benue Congo. Finally, Benue Congo was split in West Benue Congo, Cross-River, Platoid, Bandoid (non-Bantu) and Narrow Bantu.

19 Another interesting linguistic identity characteristic that is worth considering is the usage of a particular language as a Lingua Franca. Unfortunately, both the definition of a Lingua Franca and the collection of local data are not at an advanced stage yet.

time and applies another, relatively recent, phenomenon: religion. In the area of conflict, religion is known to be a significant divider between different sides and a unifier among those who follow the same religion. Therefore, religion is used as one of the five identity characteristics. The *Joshua Project* provides data on what the main religion of an ethnic group is for all groups, but unfortunately the subdivision is not always provided. I have used generally accessible sources, such as the *Encyclopedia Brittanica* to fill in some of the blanks of ethnicities that did not yet have a subgrouping. As a result, as can be seen in Appendix D, about 72% of all ethno-linguistic groups, covering more than 90% of the population, have been given a religious subgrouping. The missing groups do have the general religious affinity (i.e. Christianity or Ethnic Religions), and this data has been used to replace the missing values. In fact, a strategy has been followed in which an observation of the affinity between subgroups of religions has been replaced by the affinity between actual religions whenever one of the two groups did not have an observation for the religious subgrouping. This clouds the actual estimation, but due to the fact that the missing groups are only small in number and are the relatively smaller groups, I think the estimation is still very reasonable. Table 1 contains the most important summary statistics for all five identity characteristics. *Average* refers to the number of groups and the number of people in the average category of an identity characteristic. *Largest category* refers to the largest category for each.

Having so constructed a profile of ethno-linguistic identity characteristics, it is thus possible to construct the measure of ELA as proposed in the previous subsection. Doing this for the African data that are available to me now, generates a matrix of $53 \times 53 = 2809$ dyadic relations and a corresponding measure of Ethno-Linguistic Affinity between them. Table 2 reports the summary statistics for the dyadic relations. As can be seen, the spread is quite large. Whereas Burundi's maximum is reached at 0.591, implying that a randomly drawn individual from Burundi shares 59.1% of its characteristics with a randomly drawn individual from Rwanda, the Central African Republic's highest value[20] is only 0.115. On average, two randomly drawn individuals from two different countries in Africa share 8.1% of their identity characteristics and two individuals from two neighbouring states share 19.8% of their characteristics. Remembering that I use 5 different identity characteristics, it can be said that two individuals from two neighbouring countries share on average approximately one identity characteristic. ELA is related to distance, but not as strongly as one might expect. The correlation coefficient between the logarithmic distance in kilometers between centre points and ELA is -0.42.

[20] Surprisingly, the Central African Republic's highest value of ELA is achieved in its relationship with non-contiguous Burkina Faso.

2.Summary statistics for ELA dyads

| | Max | Average | Min |
|---|---|---|---|
| Highest ELA | 0.591 | 0.315 | 0.115 |
| | BUR-RWA | - | CAR-BFA |
| Avg ELA | 0.134 | 0.081 | 0.031 |
| | COM | - | MAD |
| Avg ELA (contiguous) | 0.439 | 0.198 | 0.067 |
| | TUN | - | CAR |
| Lowest ELA | 0.025 | 0.003 | $2.3 \times 10^{-8}$ |
| | TZA-ETH | - | ERI-LES |
| Lowest ELA (contiguous) | 0.391 | 0.140 | 0.005 |
| | TUN-LIB | - | DRC-SUD |

*Note: Max, average and Min values are reported for the highest ELA, the ELA country averages, ELA country averages including contiguous states only, lowest ELA and the lowest ELA including contiguous states only. The table also reports between/in which nations these extremes are found. For the neighbour-only values, only directly contiguous neighbours are included and island nations are left out of the analysis completely.*

Unfortunately, some caution is in order for the current measure. Two major problems should be discussed, although I think they can be dismissed in the end. First, there is the *ex post* definition of the individual groups. As discussed before, previous authors have dismissed this problem as minor but I am afraid that the level of detail of the data used makes them more susceptible to problems. Additionally, the fact that the data are mostly very recent should also create worries, because the cumulative amount of dislocated people due to civil war has increased significantly over time. Roeder (2001) compares ELF's from different time periods, but his most recent one employs data from 1985. This excludes the increased number of violent civil conflicts from the nineties, most importantly the Rwanda and Burundi ethnic conflicts. Unfortunately, there is little to be done about this and I simply have to follow previous authors and their argument that ethnic heterogeneity persistence really is very high.

The second problem is the actual data source. While I argued earlier that there are strong arguments in favour of using this particular source, potential contamination due to its religious purpose cannot be excluded. Again, as explained earlier, I have made all possible effort to make sure that this contamination is as limited as possible, but it would be interesting to replicate the results with data from an alternative source. Unfortunately this is not feasible in the short run, as no comprehensive alternative data source exists that contains as much information in such detail as this particular one.

## Practical Application: Conflict Spillovers

In this section, I set up a practical application of my index of ELA between nations. Several authors, including the ones I have quoted earlier, have published papers which try to explain the occurrence of civil conflict. A smaller number of authors have used spill-over effects as one of the mechanisms involved in this and I think it is very important to do so. I am therefore proposing to use a combination of ELA and a measure of geographic distance, in addition to a number of other variables that are known to predict civil conflict.

Spatial econometrics makes use of a number of specially developed techniques, as explained so well in Anselin (1988). More recently, Beck *et al* (2006) is an excellent contribution to the estimation issues when using spatial econometrics in a political economics context. Their explanation of the particularities regarding the interpretation of coefficients and the estimation techniques is illuminating.

### Model

Collier and Hoeffler (2004), Montalvo and Reynal-Querol (2005) and Gleditsch (2007), all use logit models to analyse the impact of different factors on the incidence of conflict. I use the same technique, with the exception of using conflict initiation instead of conflict incidence as a dependent variable. As a source of the conflict data, I use the PRIO database (Gleditsch *et al*., 2002). Among the different standard explanatory variables used by these authors and that I use as well, are the natural logarithms of GDP and population (Heston *et al*., 2006), the years since the most recent conflict in a country (Gleditsch *et al*., 2002)[21], a measure of political freedom (Center for Global Policy, 2008)[22], a measure of ethnic polarization (Montalvo and Reynal-Querol, 2005) and a

[21] Following Gleditsch (2007), instead of just the number of years since the last conflict, an exponential function is included: $e^{\left[-\frac{y}{\alpha}\right]}$, in which $y$ is the number of conflict-free years and $a$ takes the experimentally determined value of 4. For the number of years since the last conflict, only years are included after 1950 and after independence.

[22] Following Gleditsch (2007), footnote 14, I do not directly use the Polity2 index. As Gleditsch warns, the makers of the index replace all missing values with a value 0. This is a dubious choice, since missing values are generally caused by an extreme flux in the political variables. Instead, Gleditsch awards the lowest possible value of -10 to these observations with 'irregular policy values'.

measure for mountainous terrain (Gerrard, 2000)[23]. The entire sample consists of 53 countries over 45 years (1960-2004). However, since some countries were not yet independent during some time of this sample, the maximum number of observations is 2134.

The most important part of the analysis, however, is obviously the way I make use of my measure of ELA. It is important to recognise that even if conflict is likely to follow ethnolinguistic patterns when spillovers take place, there has to be a geographical proximity factor involved as well. I therefore combine ELA with a measure of geographic distance. In fact, the measure of geographic proximity I use is border length between nations (CIA, 2007)[24]. The possibility of conflict spilling over from one country to a noncontiguous one seems dismissable and including that would only increase the possibility of spurious correlations[25]. So, the contiguity matrix $W$ consists of a square matrix with all nations along the horizontal and vertical axes and with the matrix elements $e_{ij}$ describing the relation between $i$ and $j$, normalised over rows, and is defined as follows:

$$e_{ij} = \frac{ELA_{ij} \circ \delta_{ij}}{\sum_{i=1}^{N} (ELA_{ij} \circ \delta_{ij})}$$

where $\delta_{ij}$ is the geographical distance measure in use[26]. A potential criticism is that this assumes equality between all borders and that spillovers should be more likely in the cases where borders clearly divide particular ethnic groups, as opposed to distant borders across mostly impassable terrain (e.g. the southern borders of Algeria). However, such prohibitive geographical features that limit potential spillovers (sea, desert, mountains) also lead to a stronger separation of the populations on opposite sides of the division. As a result, the level of Ethno-Linguistic Affinity can be expected to be low in such cases

---

[23] For this variable, I follow Collier and Hoeffler (2004), who use this same source because it gives a good estimate for mountainous regions that give an opportunity for rebels to hide. The measure combines elevation, relative relief and area in order to identify mountainous areas.

[24] As often in spatial econometrics, islands present a problem. I deal with the issue on a case-by-case basis and employ several formulas. The assumed border length influence of *i* on *j*, when *i* is either a coastal nation or an island and *j* is an island is $\delta_{ij} = 100 \cdot \frac{coastline_i}{distance_{ij}}$; the border length influence of an island *i* on coastal nation *j* is $\delta_{ij} = coastline_i \cdot \frac{coastline_j}{\sum borderlength_j} \cdot \frac{distance_{ij} \cdot coastline_j}{\sum_k (distance_{jk} \cdot coastline_k)}$, where $k$ stands for all the islands that are within reach of $j$.

[25] As Beck *et al*. (2006, p. 28) note, "The assumption that these connectivities are known a priori is both a strong assumption and a cricitical one for the methods of spatial econometrics to work". For this reason, it is important to reject other mechanisms (such as distance between center points or distances of closest approach) on theoretical grounds.

[26] For the unfamiliar reader, appendix A contains an extensive explanation of the practical application of contiguity matrices in a spatial econometric context.

and this is unlikely to be countered by the border lengths. In the case of Algeria, the presence of the Sahara to separate the northern Algerian population centres from their southern neighbours in Mali and Niger, leads to a low level of Ethno-Linguistic Affinity between Algerians and their souther neighbours and consequently a lower spill-over probability.

Putting all these factors together in a logit model results in a complete regression that looks as follows:

$$\Pr(y_{i,t}=1)=\frac{e^{\eta_{i,t}}}{1+e^{\eta_{i,t}}}$$

where

$$\begin{aligned}\eta_{i,t} \quad = \quad & \beta_0+\beta_1\ln(gdp_{i,t-1})+\beta_2\ln(pop_{i,t-1})+\beta_3 peace_{i,t}+ \\ & \beta_4 mount_i+\beta_5 dem_{i,t}+\beta_6 W(conf_t)+\varepsilon_{it}\end{aligned}$$

in which $y_{i,t}$ is a variable that takes value 1 if a new conflict was initiated in country $i$ during year $t$, $gdp_{i,t-1}$ is the one-period lagged level of GDP, $pop_{i,t-1}$ is the one-period lagged population size, $peace_{i,t}$ is a measure for the years of peace at the start of the year, $dem_{i,t}$ is the adjusted Polity2 score for an observation-year and $conf_{i,t}$ is a dummy that takes value 1 if conflict is taking place in a country during period $t$. Estimating this seemingly easy regression is not completely trivial, however, as $y_{i,t} \in conf_t$ and as a result, the value change in the dependent variable influences that particular independent variable. This is one of the things that Beck *et al*. (2006) mean when they say that caution has to be exercised when interpreting the coefficients in a spatial econometric model. There is, however, a solution. According to Ward and Gleditsch (2002) and Gleditsch (2007), one can use either Markov Chain Monte Carlo simulation or Maximum Pseudo-Likelihood (MPL) methods. The results, however, are very similar and as the MPL method is easier to apply I use this in the estimations.

## Results

Column 1 of table 3 contains a baseline model for conflict initiation, in which there are no spillovers. As can be seen, lagged GDP, lagged population and the years of peace since the last conflict are all significant and have the expected sign. $mountain$ and $polity2$ have the expected sign, but are not significant. The table then continues to show the model described in the previous subsection, of which the results are shown in column

2. Again, lagged GDP has a negative impact on the probability of new conflict initiation and lagged population a positive one. Due to the way the number of peace years has been defined, the positive and significant coefficient of $peace$ implies that a country that has been in peace for a longer time has a lower probability of conflict outbreak. $mountain$, which measures the inaccessibility of terrain, is positive, as expected, but not significant. Finally, the $polity2$ score has an insignificant negative impact. The most interesting and relevant variable, however, is of course $W \cdot conf$. It can be seen immediately that the spillover of conflict along the combination of a geographic and my proposed ethno-linguistic channel is positive and significant, which means that a country whose ethno-linguistically close neighbours are suffering from conflict is more likely to suffer conflict initiation as well. In order to check whether the claim is warranted that the ethno-linguistic channel plays an important role in this, column 3 of table 3 shows the same result, but with $W$ defined using only border-lengths. As can be seen, the significant relation between neighbouring conflict and home-country conflict disappears. Finally, in column 4, the result is shown when only ELA is used in the contiguity matrix. In this case, the result also disappears, which is not surprising, because it includes linkages that are too far-sought to influence conflict spillovers (such as the strong ethno-linguistic ties between the north-east and north-west of Africa).

In table 4, a number of variations are shown. The first three columns contain results for the same regression, but with a measure of ethnic polarisation included (from Montalvo and Reynal-Querol, 2005). The strength of the results is slightly reduced, but overall the conclusion remains the same. I believe, however, that ethnic polarisation should not be included in these regressions, due to the interference it has with the measurement of Ethno-Linguistic Affinity, as these variables are both measuring along the same dimension.

The final two columns (4 and 5) drop the insignificant variables of $polity2$ and $mountain$ respectively. This is not done so much on theoretical grounds, but based on the fact that these variables are the bottlenecks for the number of observations. Dropping either of these variables increases the number of observations significantly, and the results are unaffected. The regressions that exclude either the ELA element or the geographic element (not shown) also yield the same results as before. A final check (not shown) leaves out both insignificant variables and the results still remain the same (with N=2041).

Collier and Hoeffler (2004) also argue that a measure for exports of primary commodities should be used as an explanatory variable. According to them, countries that have large exports of primary commodities are more likely to have rebellion due to the oppor-

3.Regression results of logit MPL estimations

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| spill-overs: | baseline | ELA*border | border | ELA |
| $C$ | -2.747** | -3.182** | -2.934** | -3.085** |
| | 1.294 | 1.294 | 1.287 | 1.273 |
| $\ln(gdp_{t-1})$ | -0.335** | -0.319** | -0.327** | -0.340** |
| | 0.145 | 0.147 | 0.146 | 0.144 |
| $\ln(pop_{t-1})$ | 0.224*** | 0.239*** | 0.225*** | 0.212** |
| | 0.084 | 0.081 | 0.083 | 0.084 |
| $peace$ | 0.668*** | 0.667*** | 0.669*** | 0.695*** |
| | 0.251 | 0.251 | 0.251 | 0.253 |
| $mountain$ | 0.460 | 0.268 | 0.353 | 0.435 |
| | 0.410 | 0.440 | 0.433 | 0.409 |
| $polity2$ | -0.014 | -0.013 | -0.014 | -0.019 |
| | 0.020 | 0.020 | 0.020 | 0.020 |
| $W \cdot conf$ | | 0.752** | 0.496 | 2.224 |
| | | 0.347 | 0.356 | 1.432 |
| $N$ | 1837 | 1837 | 1837 | 1837 |
| $LR - \chi^2$ | 32.68 | 36.30 | 34.36 | 34.73 |
| $df$ | 5 | 6 | 6 | 6 |

*Note: Results of the most important regressions, using robust Maximum Pseudo-Likelihood estimations in a logit model. Variables defined in the text. *, ** and *** imply significance at 10%, 5% and 1% respectively.*

4.Further regression results of logit MPL estimations, using alternative specifications

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | ELA*dist | dist | ELA | ELA*dist | ELA*dist |
| $C$ | -3.440*** | -3.173** | -3.353*** | -3.173** | -2.923** |
| | 1.308 | 1.299 | 1.287 | 1.264 | 1.247 |
| $\ln(gdp_{t-1})$ | -0.382** | -0.399** | -0.408** | -0.368*** | -0.347** |
| | 0.158 | 0.158 | 0.158 | 0.141 | 0.143 |
| $\ln(pop_{t-1})$ | 0.258*** | 0.245*** | 0.236*** | 0.286*** | 0.230*** |
| | 0.085 | 0.086 | 0.089 | 0.073 | 0.077 |
| $peace$ | 0.522* | 0.520* | 0.548** | 0.518** | 0.700*** |
| | 0.275 | 0.274 | 0.278 | 0.237 | 0.239 |
| $mountain$ | 0.292 | 0.366 | 0.412 | 0.259 | |
| | 0.447 | 0.442 | 0.424 | 0.439 | |
| $polity2$ | -0.009 | -0.009 | -0.013 | | -0.016 |
| | 0.020 | 0.020 | 0.020 | | 0.020 |
| $polar$ | 1.154** | 1.161** | 1.198** | | |
| | 0.568 | 0.574 | 0.578 | | |
| $W \cdot conf$ | 0.643* | 0.340 | 1.968 | 0.742** | 0.973*** |
| | 0.358 | 0.364 | 1.482 | 0.345 | 0.320 |
| $N$ | 1753 | 1753 | 1753 | 1950 | 1894 |
| $LR - \chi^2$ | 40.68 | 39.28 | 39.82 | 35.33 | 39.18 |
| *df* | 7 | 7 | 7 | 5 | 5 |

*Note: Results of the most important regressions, using robust Maximum Pseudo-Likelihood estimations in a logit model. Variables defined in the text. *, ** and *** imply significance at 10%, 5% and 1% respectively.*

tunity of rebel financing it presents. The authors also show that in their dataset they get significant results implying that the proposed mechanism is indeed at work. However, Fearon (2005) argues that the relationship between primary commodity exports and conflict is not nearly as clear. His argument is that these exports actually provide an easy source of finance for the government, which may lead to less stable institutions, but also potentially to a government that is better able to fight off a rebel uprising. In order to see whether there is any effect, I have also added a measure for primary commodity exports to my model[27], but the influence is insignificant. The spill-over effect remains strong and other explanatory variables are also unaffected.

The interpretation of the results is one thing to look at carefully, however. Due to the logit structure of the analysis, one must be careful in interpreting the coefficients. In fact, it is most convenient to report the estimates for the influence the different variables have, keeping the others constant at their mean. Taking the original model, as shown in column 2 of table 3 and keeping all variables at their mean values, the ceteris paribus addition of one standard deviation of conflict among neighbours increases the probability of conflict by 1.0 percentage point. To compare, increasing lagged GDP, lagged population or $peace$ by one standard deviation, leads to changes of -1.2, 1.5 and -1.1 percentage points respectively. So changes in those factors impact the probability of conflict initiation to a similar degree as changes in neighbouring conflict. The implied changes appear to be quite small, but it should be taken into account that the probability of conflict initiation is small to begin with at 4.8%. Therefore, a change by 1.0 percentage points, implies an increased probability of 21.3%, which cannot be considered small.

Finally, it is good to have another look at the measure of ELA. While it would be desirable to do a factor analysis of some kind to analyse whether the set of identity characterics used now is appropriate to use, this is unfortunately not possible. Any analysis of such kind is based on the assumption that the final measure is a linear sum of the different components. This is not the case here, due to the fact that the actual measure used in the regression is the result of the row-normalisation of a multiplication of ELA and border length. This makes different components non-linear. However, as an alternative measure, it is possible to re-create alternative ELAs. In table 5, for each of the columns, a different identity characteristic is dropped. At that moment, a new ELA is calculated on basis of only four identity characteristics, which is then combined with the distance measure and finally used in these regressions. The results show that little

[27] I employ the dataset that Fearon (2005) uses, which is a dataset created on basis of Collier and Hoeffler's (2005) data. The latter use 5-year periods for their analyses, whereas the first uses yearly data, just like me. Unfortunately, the dataset finishes is 1999, which causes a fairly large drop in the number of available observations (N=1630).

5.Logit regressions, where each column drops one of the identity characteristics

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| *missing char* | GRP | LAN | LNG | CLUS | REL |
| $C$ | -3.188** | -3.183** | -3.097** | -3.210** | -3.138** |
| | 1.294 | 1.294 | 1.292 | 1.296 | 1.293 |
| $\ln(gdp_{t-1})$ | -0.318** | -0.319** | -0.319** | -0.316** | -0.324** |
| | 0.147 | 0.147 | 0.146 | 0.146 | 0.147 |
| $\ln(pop_{t-1})$ | 0.240*** | 0.239*** | 0.233*** | 0.239*** | 0.241*** |
| | 0.081 | 0.081 | 0.082 | 0.081 | 0.081 |
| $peace$ | 0.666*** | 0.666*** | 0.675*** | 0.663*** | 0.675*** |
| | 0.251 | 0.251 | 0.251 | 0.251 | 0.251 |
| $mountain$ | 0.268 | 0.268 | 0.288 | 0.261 | 0.287 |
| | 0.440 | 0.440 | 0.437 | 0.442 | 0.439 |
| $polity2$ | -0.013 | -0.013 | -0.014 | -0.013 | -0.013 |
| | 0.020 | 0.020 | 0.020 | 0.020 | 0.020 |
| $W \cdot conf$ | 0.757** | 0.754** | 0.652* | 0.779** | 0.684** |
| | 0.346 | 0.346 | 0.334 | 0.349 | 0.343 |
| $N$ | 1837 | 1837 | 1837 | 1837 | 1837 |
| $LR - \chi^2$ | 36.38 | 36.34 | 35.36 | 36.76 | 35.49 |
| *df* | 6 | 6 | 6 | 6 | 6 |

*Note: Results of logit regressions with alternative measures of ELA. Each column drops one of the identity characteristics for its construction of ELA. The missing characteristics are GRP=Ethnic Group; LAN=Language; LNG=Linguistic Group; CLUS=People Cluster; REL=Religious subgroup and other variables defined in the text. *, ** and *** imply significance at 10%, 5% and 1% respectively.*

changes when dropping any of the variables. Only the Linguistic Subgroup has a strong effect on the point estimate and causes a reduction in the significance level (to $z = 1.95$), which shows its importance. In the other direction, dropping the People Cluster variable increases the point estimate for that parameter by a small amount. This can safely be ignored.

## Conclusion

In this paper, an index of Ethno-Linguistic Affinity between nations is set up. Using relatively simple tools, it appears to be easy to set up such an index that is able to avoid many of the caveats that haunt ethno-linguistic indices in general. When dissecting ethnicities into separate identity characteristics and considering the (dis)similarity on different levels it turns out to be possible to set up a measure that successfully exploits the varying sizes of differences between ethnicities along the lines of these characteristics.

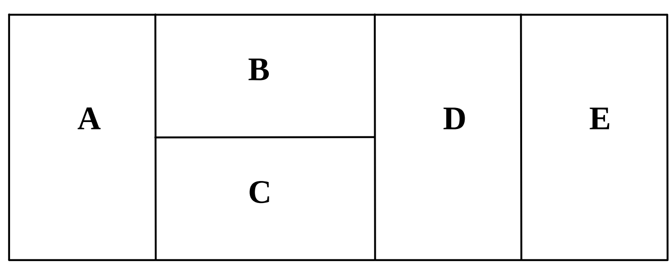
Many spatial-econometric analyses use geographic measures of distance between nations as core variables, but in this paper I argue that in many of these cases it is not merely physical proximity that has the strongest influence, but the ethno-linguistic (dis)similarity in dyads of nations. Examples from the spatial-econometric literature include the spill-over effects of trade, conflict and economic growth, all of which might benefit from the inclusion of an ethno-linguistic component. The most interesting and promising field, however, is the spillover of institutions. It can easily be argued that such spillovers are among the most likely to spill over along ethno-cultural lines, particularly if one is able to include particularly appropriate identity characteristics. So far, democracy has been the only kind of institution for which there is a fairly substantial body of literature, whereas other kinds of institutions, including trade and social institutions could also be considered for spill-over effects. In all of these, though, the inclusion of an ethno-linguistic component is an idea worth considering.

However, a large challenge for the practical application of the proposed measure concerns data collection. Clearly, this measure benefits from the most detailed level of data collection, but it may be difficult to collect such detailed data and guarantee its accuracy and its completeness. The latter is pivotal to being able to actually use this measure of ELA, so it should not be underestimated. This is also immediately related to the largest drawback of the application worked out in the current paper. The data source is unorthodox, which may lead some to dismiss it. However, I argue that the creators of this database had strong incentives to make it as accurate as possible and that it can therefore be used.

Finally, an example of an application of this measure of ELA is included. Ordinary geographic distances dismiss the existence of conflict spillovers in Africa, but when including my measure of ELA, the results change drastically. Conflict can clearly be seen to be more likely to initiate in countries that have an ethno-linguistically similar neighbouring country suffering from conflict.

# A Shape of Contiguity Matrix W

In this appendix, an example is given of the practical application of the contiguity matrix. A contiguity matrix is an $N \times N$ matrix of which the elements indicate the weights of all $i \in N$ on all $j \in N$. The elements of a contiguity matrix are the row-normalised distances between the different $i, j$. They are calculated as follows: $e_{ij} = \frac{\delta_{ij}}{\sum\limits_{i} \delta_{ij}}$, where $\delta_{ij}$ is an operator for defining the distance between countries $i$ and $j$. This distance operator can be defined in many ways, including simple contiguity, border lengths, center point distances, distance of closest approach, ethnic similarity, trade intensity or a combination of these. When the contiguity matrix has been constructed, a multiplication takes place with the variable of interest $y_i$. That way, one reaches the spatial equivalent of the AR term in time-series analysis which gives a *distance measure*-weighed average of $y_{neighbours}$. For clarification, I now show how to construct a contiguity matrix for five hypothetical nations on a hypothetical continent with the following shape:




## Border-length contiguity

When border-length contiguity is used, the transition matrix, which is the first step in creating a contiguity matrix and contains the elements $\delta_{ij}$, looks as follows:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{bmatrix}, \text{ so row-summing, } \sum = \begin{bmatrix} 2 \\ 4 \\ 4 \\ 4 \\ 2 \end{bmatrix}, \text{ which makes it possible}$$

to arrive at the actual border-length contiguity matrix, by dividing the transition matrix with the sum of the rows: $W_{bor} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$.

The contiguity matrix now shows the size of the influence of the different neighbours on the countries. The way this contiguity matrix is applied in practice is through multiplication with the variable of interest $Y$:

$$W_{bor} \cdot Y = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} y_A \\ y_B \\ y_C \\ y_D \\ y_E \end{bmatrix} = \begin{bmatrix} \frac{1}{2}y_B + \frac{1}{2}y_C \\ \frac{1}{4}y_A + \frac{1}{2}y_C + \frac{1}{4}y_D \\ \frac{1}{4}y_A + \frac{1}{2}y_B + \frac{1}{4}y_D \\ \frac{1}{4}y_A + \frac{1}{4}y_C + \frac{1}{2}y_E \\ y_D \end{bmatrix}$$

This result can then be used in the actual regression as the spatially weighed value of $y_{neighbvours}$.

## Ethno-Linguistic Affinity

In the case of Ethno-Linguistic Affinity, an example could be a matrix of ELA in which the different populations have these hypothesized levels of affinity for each other:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} - & 0.4 & 0.1 & 0.05 & 0.2 \\ 0.4 & - & 0.3 & 0.2 & 0.1 \\ 0.1 & 0.3 & - & 0.05 & 0.05 \\ 0.05 & 0.2 & 0.05 & - & 0.5 \\ 0.2 & 0.1 & 0.05 & 0.5 & - \end{bmatrix}, \text{ so row-summing, } \sum = \begin{bmatrix} 0.75 \\ 1 \\ 0.5 \\ 0.8 \\ 0.85 \end{bmatrix}, \text{ which}$$

leads to the following contiguity matrix:

$$W_{ELA} = \begin{bmatrix} 0 & \frac{8}{15} & \frac{2}{15} & \frac{1}{15} & \frac{4}{15} \\ \frac{4}{10} & 0 & \frac{3}{10} & \frac{2}{10} & \frac{1}{10} \\ \frac{2}{10} & \frac{6}{10} & 0 & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{16} & \frac{4}{16} & \frac{1}{16} & 0 & \frac{10}{16} \\ \frac{4}{17} & \frac{2}{17} & \frac{1}{17} & \frac{10}{17} & 0 \end{bmatrix}.$$

## Combining ELA and border-lengths

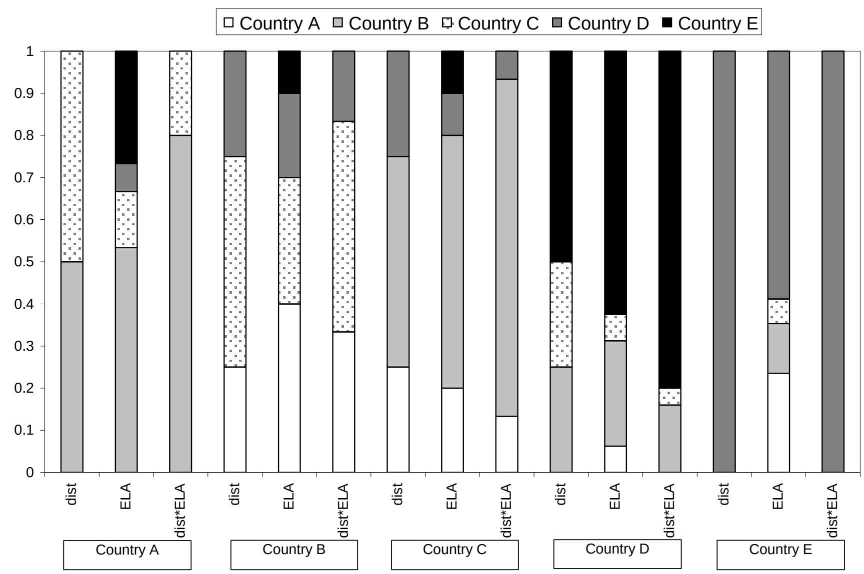
Using the previous measures of border-lengths and Ethno-Linguistic Affinity, multiplying them element-wise would yield the following transition matrix:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} - & 0.4 & 0.1 & 0 & 0 \\ 0.4 & - & 0.6 & 0.2 & 0 \\ 0.1 & 0.6 & - & 0.05 & 0 \\ 0 & 0.2 & 0.05 & - & 1 \\ 0 & 0 & 0 & 1 & - \end{bmatrix}, \text{ row-summing, } \sum = \begin{bmatrix} 0.5 \\ 1.2 \\ 0.75 \\ 1.25 \\ 1 \end{bmatrix}, \text{ and}$$

$$W_{ELA \circ dist} = \begin{bmatrix} 0 & \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ \frac{2}{6} & 0 & \frac{3}{6} & \frac{1}{6} & 0 \\ \frac{2}{15} & \frac{12}{15} & 0 & \frac{1}{15} & 0 \\ 0 & \frac{4}{25} & \frac{1}{25} & 0 & \frac{20}{25} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

## Comparing contiguity matrices

Comparing these possible contiguity matrices, one can see the level of influence of each nation changes per contiguity matrix. The following figure shows the different levels of influence exercised on each of the nations.




*Note: The different levels of influence the hypothetical nations will exercise on the others for the contiguity matrices calculated in appendix* A.

# B List of Linguistic Groups

List with different Linguistic groups:

| Nr. | Language Groups | Nr. country-groups | People (mln) | %Population |
|---|---|---|---|---|
| | **Niger-Congo** | | | |
| | *Atlantic Congo* | | | |
| 1 | Atlantic | 144 | 34.9 | 3.7% |
| 2 | Ijoid | 11 | 25.0 | 0.3% |
| | *Volta-Congo* | | | |
| 3 | Dogon | 12 | 0.7 | 0.1% |
| 4 | Kru | 52 | 2.9 | 0.3% |
| 5 | Kwa | 130 | 30.4 | 3.3% |
| 6 | Northern Languages | 355 | 37.1 | 4.0% |
| | *Benue Congo* | | | |
| 7 | West Benue-Congo | 94 | 62.6 | 6.7% |
| 8 | Cross-River | 100 | 4.7 | 0.5% |
| 9 | Platoid | 69 | 8.9 | 1.0% |
| 10 | Bandoid, non-Bantu | 170 | 9.6 | 1.0% |
| 11 | Narrow Bantu | 906 | 273.9 | 29.4% |
| 12 | Kordofanian | 27 | 0.6 | 0.1% |
| 13 | Mande | 134 | 25.5 | 2.7% |
| | **Indo-European** | | | |
| 14 | Italic | 86 | 3.3 | 0.3% |
| 15 | Slavic | 12 | 0.1 | 0.0% |
| 16 | Indo-Iranian | 60 | 3.1 | 0.3% |
| 17 | Greek | 18 | 0.2 | 0.0% |
| 18 | Germanic | 111 | 10.0 | 1.1% |
| 19 | Albanian | 1 | 0.0 | 0.0% |
| 20 | Armenian | 3 | 0.0 | 0.0% |

| Nr. | Language Groups | Nr. country-groups | Speakers (millions) | %Population |
|---|---|---|---|---|
| **Nilo-Saharan** | | | | |
| 21 | Berta | 2 | 0.2 | 0.0% |
| 22 | Central Sudanic | 102 | 9.8 | 1.1% |
| 23 | Easten Sudanic | 139 | 28.1 | 3.0% |
| 24 | Fur | 5 | 1.0 | 0.1% |
| 25 | Komuz | 9 | 0.3 | 0.0% |
| 26 | Kunuma | 3 | 0.2 | 0.0% |
| 27 | Maban | 11 | 0.9 | 0.1% |
| 28 | Saharan | 27 | 8.3 | 0.9% |
| 29 | Songhai | 20 | 5.8 | 0.6% |
| 30 | Unclassified | 8 | 0.2 | 0.0% |
| **Afro-Asiatic** | | | | |
| 31 | Berber | 51 | 19.2 | 2.1% |
| 32 | Chadic | 237 | 46.5 | 5.0% |
| 33 | Cushitic | 91 | 51.3 | 5.5% |
| 34 | Omotic | 31 | 5.2 | 0.6% |
| 35 | Semitic | 236 | 210.7 | 22.6% |
| 36 | Unclassified | 1 | 0.0 | 0.0% |
| **Khoisan** | | | | |
| 37 | Hatsa | 1 | 0.0 | 0.0% |
| 38 | Sandawa | 1 | 0.0 | 0.0% |
| 39 | Southern African | 39 | 0.5 | 0.1% |
| **Creole** | | | | |
| 40 | Afrikaans-based | 1 | 0.0 | 0.0% |
| 41 | Arabic-based | 2 | 0.0 | 0.0% |
| 42 | English-based | 11 | 2.9 | 0.3% |
| 43 | French-based | 10 | 0.8 | 0.1% |
| 44 | Kongo-based | 3 | 7.7 | 0.8% |
| 45 | Ngbandi-based | 6 | 0.6 | 0.1% |
| 46 | Portuguese-based | 13 | 0.9 | 0.1% |
| 47 | Swahili-based | 1 | 0.0 | 0.0% |

| Nr. | Language Groups | Nr. country-groups | Speakers (millions) | %Population |
|---|---|---|---|---|
| **Sino-Tibetan** | | | | |
| 48 | Chinese | 18 | 0.2 | 0.0% |
| **Isolates** | | | | |
| 49 | Korean | 4 | 0.0 | 0.0% |
| 50 | Centúúm | 1 | 0.0 | 0.0% |
| **Pidgin** | | | | |
| 51 | English-based | 1 | 0.1 | 0.0% |
| 52 | Zulu-based | 2 | 0.0 | 0.0% |
| **Austronesian** | | | | |
| 53 | Malayo-Polynesian | 44 | 18.8 | 2.0% |
| **Dravidian** | | | | |
| 54 | South-Central | 2 | 0.1 | 0.0% |
| 55 | Southern | 5 | 0.5 | 0.1% |
| **Altaic** | | | | |
| 56 | Turkic | 1 | 0.0 | 0.0% |
| **North Caucasian** | | | | |
| 57 | Northwest | 1 | 0.0 | 0.0% |
| **Japanese** | | | | |
| 58 | Japanese | 1 | 0.0 | 0.0% |
| **Mixed** | | | | |
| 59 | Makhua-Nyanja | 1 | 0.3 | 0.0% |
| 60 | Songhay-Berber | 1 | 0.0 | 0.0% |
| 61 | Bantu-Cushitic | 1 | 0.0 | 0.0% |
| **Other** | | | | |
| 97 | Sign Language | 22 | 0.2 | 0.0% |
| 98 | Unclassified | 8 | 0.0 | 0.0% |
| 99 | Language Unknown | 34 | 0.0 | 0.0% |
| Total | | 3704 | 933.0 | 100% |

# C List of People Clusters

This list of People Clusters comes from the Johnstone (2007), but with some small adjustments. Some of the non-native African peoples have been merged into slightly larger groups in order to reduce the number of groups.

| Nr. | People Cluster | Nr. country-groups | People (mln) | %Population |
|---|---|---|---|---|
| 1 | Adamawa-Ubangi | 165 | 9.4 | 1.0% |
| 2 | Afar | 6 | 2.1 | 0.2% |
| 3 | Anglo-Saxon[28] | 65 | 1.5 | 0.2% |
| 4 | Arab, Arabian | 20 | 2.5 | 0.3% |
| 5 | Arab, Hassaniya | 20 | 7.2 | 0.8% |
| 6 | Arab, Levant | 25 | 1.5 | 0.2% |
| 7 | Arab, Libyan | 7 | 3.7 | 0.4% |
| 8 | Arab, Maghreb | 14 | 54.6 | 5.8% |
| 9 | Arab, Shuwa | 13 | 2.3 | 0.2% |
| 10 | Arab, Sudan | 54 | 24.9 | 2.7% |
| 11 | Arab, Yemeni | 8 | 0.5 | 0.1% |
| 12 | Atlantic | 65 | 5.9 | 0.6% |
| 13 | Atlantic-Jola | 20 | 0.6 | 0.1% |
| 14 | Atlantic-Wolof | 9 | 5.2 | 0.6% |
| 15 | Baloch | 1 | 0.0 | 0.0% |
| 16 | Bantu, Cameroon-Bamileke | 65 | 4.0 | 0.4% |
| 17 | Bantu, Central-Congo | 155 | 20.9 | 2.2% |
| 18 | Bantu-Central-East | 11 | 2.9 | 0.3% |
| 19 | Bantu, Central-Lakes | 88 | 51.1 | 5.5% |
| 20 | Bantu, Central-Luba | 15 | 12.9 | 1.4% |
| 21 | Bantu, Central-South | 130 | 21.1 | 2.3% |

[28] Combination of Anglo-Celt, Anglo-American and Caucasian Peoples, Generic

| Nr. | People Cluster | Nr. country-groups | People (millions) | %Population |
|---|---|---|---|---|
| 22 | Bantu, Central-Southeast | 11 | 7.2 | 0.8% |
| 23 | Bantu, Central-Southwest | 26 | 7.6 | 0.8% |
| 24 | Bantu, Central-Tanzania | 55 | 18.6 | 2.0% |
| 25 | Bantu, Chewa-Sena | 34 | 16.7 | 1.8% |
| 26 | Bantu, East-Coastal | 26 | 3.9 | 0.4% |
| 27 | Bantu, Gikuyu-Kamba | 16 | 13.8 | 1.5% |
| 28 | Bantu, Kongo | 10 | 11.8 | 1.3% |
| 29 | Bantu, Makua-Yao | 40 | 21.3 | 2.3% |
| 30 | Bantu, Nguni | 25 | 25.2 | 2.7% |
| 31 | Bantu, Northwest | 96 | 6.6 | 0.7% |
| 32 | Bantu, Shona | 26 | 12.5 | 1.3% |
| 33 | Bantu, Soth-Tswana | 44 | 14.0 | 1.5% |
| 34 | Bantu-Southeastern | 2 | 1.1 | 0.1% |
| 35 | Bantu, Swahili | 33 | 3.4 | 0.4% |
| 36 | Bedouin, Arabian | 1 | 0.9 | 0.1% |
| 37 | Bedouin, Saharan | 17 | 3.2 | 0.3% |
| 38 | Beja | 5 | 2.9 | 0.3% |
| 39 | Benue | 319 | 26.2 | 2.8% |
| 40 | Berber-Kabyle | 2 | 3.2 | 0.3% |
| 41 | Berber-Riff | 3 | 2.5 | 0.3% |
| 42 | Berber-Saharan | 25 | 1.2 | 0.1% |
| 43 | Berber-Shawiya | 3 | 1.9 | 0.2% |
| 44 | Berber-Shilha | 12 | 8.6 | 0.9% |
| 45 | Chadic | 215 | 9.6 | 1.0% |
| 46 | Chinese | 20 | 0.2 | 0.0% |
| 47 | Deaf | 53 | 0.2 | 0.0% |

| Nr. | People Cluster | Nr. country-groups | People (millions) | %Population |
|---|---|---|---|---|
| 48 | Eastern European[29] | 38 | 0.4 | 0.0% |
| 49 | Egyptian | 5 | 65.9 | 7.1% |
| 50 | Ethiopian | 30 | 42.3 | 4.5% |
| 51 | French | 46 | 0.8 | 0.1% |
| 52 | Fulani/Fulbe | 48 | 29.8 | 3.2% |
| 53 | Germanic | 21 | 4.1 | 0.4% |
| 54 | Ger-Naba of Chad | 7 | 0.5 | 0.1% |
| 55 | Guinean | 155 | 33.0 | 3.5% |
| 56 | Gujarati | 16 | 1.1 | 0.1% |
| 57 | Gur | 164 | 28.5 | 3.1% |
| 58 | Gypsy | 5 | 1.2 | 0.1% |
| 59 | Hausa | 20 | 29.7 | 3.2% |
| 60 | Hindi | 18 | 0.5 | 0.1% |
| 61 | Hispanic-Iberian[30] | 20 | 1.0 | 0.1% |
| 62 | Igbo | 14 | 20.5 | 2.2% |
| 63 | Ijaw | 11 | 2.5 | 0.3% |
| 64 | Italian | 9 | 0.2 | 0.0% |
| 65 | Japanese-Korean[31] | 5 | 0.0 | 0.0% |
| 66 | Jews | 19 | 0.1 | 0.0% |
| 67 | Kanuri-Saharan | 30 | 8.7 | 0.9% |
| 68 | Khoisan | 65 | 1.8 | 0.2% |
| 69 | Kru | 54 | 2.9 | 0.3% |
| 70 | Malagasy | 41 | 18.8 | 2.0% |
| 71 | Malinke | 48 | 8.5 | 0.9% |
| 72 | Malinke-Bambara | 13 | 5.0 | 0.5% |
| 73 | Malinke-Jula | 11 | 1.3 | 0.1% |

[29] Combination of Albanian, Maltese, Romanian, Armenian, Greek, Caucasus, Slav, Eastern and Slav, Southern.
[30] Combination of Hispanic, Spanish and Portuguese-European.
[31] Combination of Japanese and Korean.

| Nr. | People Cluster | Nr. country-groups | People (millions) | %Population |
|---|---|---|---|---|
| 74 | Mande | 59 | 8.7 | 0.9% |
| 75 | Nilotic | 121 | 27.4 | 2.9% |
| 76 | Nuba Mountains | 48 | 1.0 | 0.1% |
| 77 | Nubian | 8 | 1.2 | 0.1% |
| 78 | Nupe | 9 | 2.8 | 0.3% |
| 79 | Omotic | 56 | 11.2 | 1.2% |
| 80 | Oromo | 23 | 26.1 | 2.8% |
| 81 | Other Indian[32] | 24 | 1.4 | 0.2% |
| 82 | Other Sub-Saharan African | 63 | 9.9 | 1.1% |
| 83 | Ouaddai-Fur | 38 | 2.9 | 0.3% |
| 84 | Pygmy | 33 | 0.7 | 0.1% |
| 85 | Sara-Bagirmi | 34 | 2.6 | 0.3% |
| 86 | Sinhala | 2 | 0.0 | 0.0% |
| 87 | Somali | 19 | 13.3 | 1.4% |
| 88 | Songhai | 21 | 5.8 | 0.6% |
| 89 | Soninke | 13 | 2.3 | 0.2% |
| 90 | South-East Asians[33] | 6 | 0.3 | 0.0% |
| 91 | Sudanic | 79 | 9.6 | 1.0% |
| 92 | Susu | 5 | 1.3 | 0.1% |
| 93 | Tamil | 4 | 0.5 | 0.1% |
| 94 | Tuareg | 15 | 2.3 | 0.2% |
| 95 | Turkish | 1 | 0.0 | 0.0% |
| 96 | Undefined | 1 | 0.0 | 0.0% |
| 97 | Urdu Muslim | 2 | 0.3 | 0.0% |
| 98 | Yoruba | 32 | 33.3 | 3.6% |
| Total | | 3704 | 933.0 | 100% |

32 Combination of Bengali, Bihari, Malayali, Marathi-Konkani, Punjabi, Sindhi, Telugu and Other South-Asian

33 Combination of Filipino, Central, Filipino, Muslim, Malay and Other South-East Asian.

# D List of Religions

List with different religions (and their greater affinity group):

| Nr. | Religion Subgroup | Overall Group | Nr. country-groups | %Population |
|---|---|---|---|---|
| 1 | Ancestor Worship | Ethnic Religion | 103 | 2.6% |
| 2 | Anglican | Christianity | 36 | 0.1% |
| 3 | Animism | Ethnic Religion | 172 | 3.5% |
| 4 | Chinese Folk | Ethnic Religion | 1 | 0.0% |
| 5 | Independent Christian | Christianity | 56 | 4.5% |
| 6 | Judaism | Ethnic Religion | 19 | 0.0% |
| 7 | Orthodox | Christianity | 39 | 4.0% |
| 8 | Other/Marginal | Christianity | 29 | 0.3% |
| 9 | Protestant | Christianity | 384 | 14.8% |
| 10 | Roman Catholic | Christianity | 781 | 18.5% |
| 11 | Sikhism | Other/Small | 3 | 0.0% |
| 12 | Sunni | Islam | 988 | 40.6% |
| 13 | Syncretized | Islam | 23 | 0.1% |
| 14 | Theravada | Budhism | 2 | 0.0% |
| 15 | Traditional | Ehtnic Religion | 8 | 0.1% |
| 16 | Hindu | Hinduism | 48 | 0.3% |
| 17 | Atheist | Non-Religious | 22 | 0.0% |
| 18 | Mahayana | Budhism | 1 | 0.0% |
| 19 | Syncretized | Ethnic Religion | 17 | 0.5% |
| 99 | Unknown | Unknown | 12 | 0.0% |
| Country-groups without subdivision for religious group: | | | | |
| | | Christianity | 254 | 3.9% |
| | | Ethnic Religion | 715 | 5.8% |
| Total | | | 3704 | 100% |

# 3 The Spill-Over Effects of Conflict on Economic Growth in Neighbouring Countries in Africa

**Abstract**

In this chapter, the influence of conflict on the economies of neighbouring countries is discussed. The results from previous papers show a strong negative effect for an entire area around a country suffering from conflict, but this paper reaches a different conclusion, by using more recent data and adjusting the methodology previously employed. Additionally, a new type of contiguity matrix is constructed and used in the actual analysis. The final analysis consists of a large number of regressions and concludes that conflict actually has two opposing effects. Firstly, like conflict countries themselves, directly contiguous countries actually suffer from the negative effects of proximate conflict. Secondly, however, there is also a positive spillover of conflict which affects non-contiguous countries and this effect is larger for countries that are closer to the conflict country. The results from the chapter predominantly hold for the most violent kind of conflict.

**Keywords**: Conflict; Economic Growth; Spatial Econometrics; Africa

**JEL code**: C21, F51, O11

## Introduction

Many have written about the influence of conflict on economic growth. The conclusion is clear-cut: conflict is bad for growth in the short term[34] (e.g. Collier, 1999 and Koubi, 2005). For long-run analyses, the phoenix effect[35] (Organski and Kugler, 1980) is occasionally cited as a reason why little evidence has been found that conflict negatively

[34] It has long been hypothesised that the causality between growth and conflict runs both ways. In their excellent contribution, Miguel *et al.* (2004) use an Instrumental Variable approach to show that negative growth shocks indeed increases the probability of conflict for a sample of African countries.

[35] The phoenix effect is named after the proverbial phoenix rising from the ashes.

affects growth. According to the theory of the phoenix factor, post-conflict nations actually grow faster than their peaceful counterparts. This appears to be a simple example of the conditional convergence theorem (Barro and Sala-i-Martin, 1992), which states that countries with a lower GDP, ceteris paribus, grow faster. After all, a country that has lost considerably from an episode of conflict has already proven to possess a certain amount of growth potential: it merely needs to recover its lost resources and enjoy the conditional convergence.

Outside the field of economics, political scientists and others have performed substantial research on the actual spill-over effects of conflict itself. They have addressed whether conflict is likely to spill over and what the external conditions are that make conflict more likely to infect proximate states (e.g. Sambanis, 2002). It seems reasonable to say that particular types of conflict are more likely to spill over, while other types of conflict are less likely to do so. In Africa, particularly, due to its colonial history, tribal conflicts are relatively common. These tend to spill over easily, because particular tribes may be distributed over two or more countries. On the other hand, civil war that is caused by specific local conditions is much less likely to affect the probability of conflict in neighbouring countries.

The relationship between conflict in one country and economic growth in nearby countries, however, has not often been researched[36]. One paper that addresses the influence of regional political instability, defined as coups and revolutions, finds a significantly negative effect for neighbouring countries (Ades and Chua, 1997). The only papers I am aware of that address the spill-over effects of actual conflict on growth in contiguous countries are by Murdoch and Sandler (2002a, 2002b, 2004), who have researched exactly this topic. The results of Murdoch and Sandler are discussed in the following section, in which I also argue that an alternative specification of the basic model may yield improved results. In the section that follows, a model is presented that offsets some of the shortcomings of Murdoch and Sandler. The data used in the analysis are presented in the fourth section and the results of the analysis can be found in the fifth section. The sixth and final section wraps up the findings and discusses some of the implications that result from the conclusions.

[36] This does not imply that the role of spatial effects itself is not often analysed. Research on the relationship between growth and geography is done thorougly in many different settings. Abreu *et al*. (2004) give a clear overview of the different econometric techniques used in the field

## Literature overview

As mentioned earlier, Murdoch and Sandler have published a number of papers on this topic. All of their papers basically use the same model, which is an augmented Solow growth model, as was described by Mankiw *et al*. (1992). These latter authors argued that Solow's (1956) model was correct in its intentions, but disregarded the possibility that equilibrium growth resulted from increased human capital and only focused on physical capital and labour growth rates instead. Using the common production function with diminishing returns to scale for all inputs, Mankiw *et al*. set up a new model to be used for growth analysis.

Murdoch and Sandler use this basic model and add a number of externalities which may also be able to influence the growth rate. In their papers, they argue that domestic and adjacent conflict are two sources that may have a negative influence on growth rates. While the authors assert that some of the problems caused by conflict work through the classical channels, by influencing the levels of capital and labour growth, they add that there might be other, unobserved, channels through which conflict influences growth.

The different papers vary in time periods, the sample of countries and contiguity definitions and a clear evolution in the papers is visible. For each paper, the authors choose to analyse the data in both a long term style, creating a cross-section dataset, and a short term one, where a panel dataset can be used. In the latter case, the unit of observation is a 5-year period and most of the discussions of the results focus on these analyses. The long term regressions yield little in terms of concrete results, which is attributed to the phoenix effect, as mentioned earlier. With regards to the country sample, two of Murdoch and Sandler's papers (2002a and 2004) use worldwide samples, whereas the third paper (2002b) distinguishes different geographical regions and reports separate results for each of these regions. This paper is particularly interesting for the current analysis, because it is possible to make a direct comparison between the results they find for Africa and the results from the current paper.

The definition of contiguity is one of the other things where the different papers distinguish themselves from each other. The first paper uses only direct contiguity, in which countries are required to actually share a border in order for them to be considered contiguous. The second paper, on the other hand, uses the Gleditsch and Ward (2001) dataset on minimal distances between nations to construct new types of contiguity matrices. While varying the maximum distance from 0 to 800 km in steps of 50 km, they use dummy variables to indicate whether a country is within a particular range. This provides the authors with 17 different contiguity matrices for each region and they report the results with the highest explanatory power. The third paper uses a similar method, but only for the minimum distances of 100, 300 and 800 km.

In each paper, Murdoch and Sandler come to the conclusion that in the short run, civil conflict indeed has a negative effect on economic growth in neighbouring countries. In the third paper, Murdoch and Sandler conclude that the effect of conflict is felt over an 800 km minimal distance, while the 2002b paper concludes that different regions have different relevant minimal distances[37].

The papers by Murdoch and Sandler have significant merit, as these are the first publications that address the issue of neighbouring countries suffering from the spill-over effects of conflict. However, a number of issues should be addressed in order to improve the results of these papers. First of all, there are some practical issues to be considered with regards to the data used by the authors. Some of the underlying datasets have recently been updated and extended beyond 1995[38], some contain significant errors and, occasionally, the coding of the data by Murdoch and Sandler seems to have been somewhat imprecise[39].

In addition to the aforementioned data concerns, I would like to point out two theoretical improvements, which may strengthen the original results. The first point that is addressed in this paper is the rigidity of Murdoch and Sandler's theoretical model with respect to the direction of the spill-over effects. As a result of their model selection, the spill-over effect is required to be unidimensional, while the following section of this paper will put forward a number of arguments in favour of a model in which two different types of contiguity matrices are to be used in a single regression. This adaptation leads to the possibility that some neighbours are influenced positively, whereas others are influenced negatively. This also leads to the second potential improvement of the theoretical model. The employment of a minimal distance dataset is an excellent idea, but by deciding to use dummy variables a lot of relevant information is lost. In order to encompass all possible information, a new type of contiguity matrix is introduced in order to deal with exactly this drawback.

[37] For Africa, 100 km minimal distance is the optimal distance, for Latin America it is 300 km or 0 km, depending on the definition of conflict, for Asia it is 600 km or 500 km and for the pooled Latin American and Asian sample the optimal minimal distance is 700 km.
[38] Particularly, Murdoch and Sandler use the Penn World Tables (PWT) 5.7, but it is now possible to use the more recent version PWT 6.2 (Heston *et al*., 2006) or the World Development Indicators (Worldbank, 2007).
[39] When interpreting the dataset used for Murdoch and Sandler (2004), several examples can be found of data that are internally inconsistent.

## Model

The basic model used in this paper is the same as the one used by Murdoch and Sandler in their papers. In summary, this is a basic Solow (1956) model, augmented in order to include human capital (Mankiw *et al.*, 1992). As a result, the inputs for the Cobb Douglas production function are capital ($K$), labour ($L$) and human capital ($H$) and the function looks as follows:

$$Y(t) = K(t)^{\alpha} H(t)^{\beta} \left[A(t)L(t)\right]^{1-\alpha-\beta} \tag{3.1}$$

In this model $\alpha$ and $\beta$ are the elasticities of output with respect to capital and human capital respectively and their sum is greater than 0, but less than 1. $A(t)$ represents the exogenously determined effectiveness of labour. The production function in equation  displays constant returns to scale and diminishing marginal returns for all inputs and satisfies the Inada conditions. Thanks to the constant returns to scale, it is possible to express production as production per effective worker, by dividing the left- and right-hand sides of equation  by $AL$:

$$y(t) = k(t)^{\alpha} h(t)^{\beta} \tag{3.2}$$

where $y = \frac{Y}{AL}$, $k = \frac{K}{AL}$ and $h = \frac{H}{AL}$.

The next step is to transform equation  into an equation that shows how changes in the parameters influence production. These so-called transition equations are derived in Mankiw *et al.* (1992) and look as follows:

$$\dot{k}(t) = s_k y(t) - (n + g + \delta)k(t) \tag{3.3}$$

$$\dot{h}(t) = s_h y(t) - (n + g + \delta)h(t) \tag{3.4}$$

where a $(\cdot)$ on top of a variable signifies the change in that variable. Furthermore, $s$ is the savings rate for physical and human capital ($k$ and $h$ respectively), $n$ is the rate of growth of the labour stock, $g$ is the exogenously given rate of technological progress and $\delta$ is the depreciation rate of capital. Theoretically, it is now possible to determine the long-run growth rate of an economy, as has been done by Mankiw *et al.* (1992) and others:

$$gr = b_0 + b_1 \ln(s_k) + b_2 \ln(s_h) - b_3 \ln(n + g + \delta) - b_4 \ln y(0) \tag{3.5}$$

In equation , ln refers to the natural logarithm, $y(0)$ is the original level of income per capita and $b_i$ are the coefficients that are theoretically supposed to capture the output elasticities and other influences.

Equation  shows the different ways economic growth is influenced. The investments in physical and human capital both exercise a positive influence on growth, while population growth, exogenous technological progress and capital depreciation decrease economic growth by diluting the amount of capital available per effective worker. Finally, the initial level of income per capita's influence on the growth rate is found in the conditional convergence theorem (Barro, 1992). Countries that start out with a higher level of income have more difficulty to grow, as a result of the diminishing marginal returns on all inputs.

Empirically, equation  can be parameterised in the following way:

$$gr = \beta_0 + \beta_1 \ln(y0) + \beta_2 \ln(inv) + \beta_3 \ln(sch) + \beta_4 \ln(n + g + \delta) \tag{3.6}$$

where $gr$ is the growth of income per capita, $y0$ is the initial income level and $inv$ is the investment in physical capital. Mankiw *et al.* point out that it is unclear whether the investment in human capital or the level of schooling is most relevant, but the latter is considered to yield the most optimal results. Following their reasoning, this paper adopts a measure for the level of educational attainment in the population, referred to as $sch$.

This paper, however, deals with the influence of conflict on economic growth and especially the influence of conflict in neighbouring countries on economic growth. In order to perform such an analysis, one should have some idea about the influence conflict may exercise. In their different publications, Murdoch and Sandler have always focused on the negative effects of conflict. These effects appear obvious for host countries, but they hypothesise that neighbouring countries would suffer from similar obstacles. There are, on the other hand, several effects resulting from conflict that may positively affect countries close to a conflict country. Before setting up the rest of the model, I discuss some of the features that potentially influence economic growth in three different types of nations: host nations, directly contiguous nations (primary neighbours) and nations that are near to a conflict nation, but not directly contiguous (secondary neighbours).

The first channel through which conflict may exercise influence on economic growth is capital. This potential influence may take different forms and the different nation categories may therefore respond differently to conflict. The primary manner through which conflict and capital may influence growth is the actual destruction of capital stock as a result of conflict. This negative effect particularly concerns host nations, but even primary neighbours are possible victims of collateral damage in the form of capital stock destruction. Secondary neighbours, on the other hand, are unlikely to suffer from such collateral damage. The influence of investment on the state of the economy is unclear. The increased investment in e.g. weapons may in fact stimulate the economy in the very short run[40], but in the medium-to-long run the positive effects disappear. Whether or not a positive effect takes place also depends on whether investments are auxiliary or whether they crowd out investments in more productive assets. As the incentives to invest in these unproductive assets decrease in distance from the conflict, the investments are likely to be lower for neighbouring states and less crowding out may take place, particularly for secondary neighbours. When outside actors, such as NGOs are responsible for the increase in unproductive investments (e.g. refugee encampments), a fully positive effect may even be expected from these investments. The third and possibly most influential impact that conflict has on the capital growth rate is the signal given by conflict to potential investors. There are few industries that actually benefit from conflict[41] and the great majority of firms will in fact actively try to avoid investing in a country suffering from conflict. A good example is the withdrawal of all major oil companies (including Shell, ConocoPhillips and BP) from Somalia when the conflict intensified there during the late 1980s. For primary neighbours, this effect may be present as well, especially if investors fear the possibility of a conflict spill-over into neighbouring countries. However, it is interesting to consider the motivation for a particular investor to invest in a certain country. One of the motivations could be the desire to invest in a particular region and, if that is the case, this may actually benefit primary and secondary neighbours. The net effect of the capital channel on primary neighbours is ambiguous, but secondary neighbours may possibly benefit from conflict through the capital channel.

The second channel through which conflict may influence economic growth in a host country or its neighbours is through labour. The primary arguments are similar to those regarding the previous channel (capital), as the most influential problems are likely to

[40] Baker (2007) shows that, for the United States, the increased military expenditure stimulates the economy during the first six years after increasing spending to finance the Iraq war. After the initial boost, the negative effects resulting from increased inflation and interest rates outweigh the benefits.

[41] One of the first industries that comes to mind is the arms industry, but as Guidolin and La Ferrara (2007) show, the diamond industry is another industry that may actually benefit from the presence of (civil) conflict. However, while there may be some benefits for the diamond industry, these are not profits that accrue to the local economy, but instead to the international companies involved.

be the destruction of productive labour and the assignment of labour to less-productive activities (e.g. soldiering). In a similar vein as in the capital channel, primary neighbours may to some extent display the same effect as host nations, but there is little rationale to assume that secondary neighbours would also have an incentive to reassign workers to unproductive activities, such as border protection. In addition to these effects, there is also the influence of refugees. Primary neighbours are likely to suffer the bulk of the refugees, which negatively influences economic growth, particularly when these refugees are unskilled and poor, which is typically the case for refugees. Those refugees, on the other hand, who make their way through a primary neighbour and into a secondary neighbour are a different story altogether. These are more likely to have a high level of human or physical capital and are possibly not planning to return to their country of origin soon. For this reason, secondary neighbours could be hypothesised to actually benefit from any refugees that reach their countries.

The third channel through which conflict may turn out to be a problem is the potential spill-over effects of conflict itself (discussed in Sambanis, 2002). However, as mentioned earlier, while there is some evidence that primary neighbours are at risk of getting involved in neighbouring conflict[42], it is not rational to expect secondary neighbours to suffer from such a predicament.

Finally, the last channel through which conflict distortions may have a negative effect is associated with trade. In a host country, both domestic and international trade are likely to be disrupted, which causes harm to economic growth. Primary neighbours may also suffer from the diversion of trade flows with the host country, but the necessary substitution of trade partners may in fact benefit the secondary neighbours to some extent. Longo and Sekkat (2001) report that there is very little intra-African trade integration, but the little trade that does occur is predominantly among direct neighbours, so the secondary neighbours are less likely to suffer from their own diversion of trade flows, but could benefit from the diversion of trade of primary neighbours. Most of Africa's trade concerns trade with developed nations, however. When, due to conflict, these nations may have to switch trading partners, neighbouring countries, which are likely to have similar resources, are advantageous candidates. This would be a potentially beneficial phenomenon for both primary and secondary neighbours.

In summary, it is fairly obvious that host countries are likely to experience a negative growth shock as a result of conflict and that primary neighbours might also suffer from a

[42] For example, in chapter 2 of this dissertation, I show that conflict is likely to spill over to countries that are both contiguous and ethno-linguistically similar.

negative spillover, at least in the short run. Secondary neighbours, on the other hand, are less likely to suffer a similar predicament and may in fact benefit from the occurrence of conflict. For this reason, I augment the testable equation to include three conflict elements. First of all, the variable $conf$ has to be added to catch host-country effects. Additionally, there have to be two different variables to catch the effect of conflict in primary and secondary neighbour-states.

In order to catch spatial spill-over effects of conflict, I set up different kinds of contiguity matrices[43], some of which are familiar and some of which are new. Technically, a contiguity or weights matrix $W$ consists of a square matrix with all nations along the horizontal and vertical axes and with the matrix elements $e_{ij}$ consisting of some distance measure between $i$ and $j$, normalised over rows:

$$e_{ij} = \frac{\delta_{ij}}{\sum_i \delta_{ij}} \tag{3.7}$$

The most basic version of a $W$-matrix is the direct contiguity matrix, in which $\delta_{ij}$ is a binary value that takes the value 1 if countries $i$ and $j$ share a border and 0 if they do not. This contiguity matrix is referred to as $W_{dum}$. In the second commonly used contiguity matrix, which I call $W_{bor}$, $\delta_{ij}$ takes the value of the border length shared between countries $i$ and $j$. These two $W$-matrices are appropriate for picking up any primary-neighbour effect and I use them as such. In order to pick up effects from secondary neighbours, it is necessary to include some kind of distance measure. The first measure that is commonly used is the simple distance matrix, $W_{dist}$, in which $\delta_{ij}$ takes the value of the inverse distance between the centre points of countries $i$ and $j$. Theoretically, however, there should be a boundary on which spatial effects can still be expected to be present and for this reason I enforce a maximum distance on which any effect may still take place[44].

Finally, I want to make use of the minimal-distance data, which measures the distance of closest approach between two countries. A dataset reporting these data was originally set up by Gleditsch and Ward (2001) and Murdoch and Sandler (2002b, 2004) used these data to set up dummy variables that take a value 1 whenever the minimum distance falls within a particular boundary. In my opinion, this was a poor choice of representation of the minimal-distance data as considerable information is lost on the actual distance be-

---

43 Anselin (1988) thoroughly discusses all the relevant characteristics of performing spatial econometric analyses and setting up different types of contiguity matrices.

44 In my analysis, I use "average distance between all country-pairs" as a cut-off point, after which distances are *de facto* set to infinity.

tween countries. It is more advisable to distinguish a graded degree of proximity, instead of treating neighbours in such a binary fashion. The minimal-distance data, however, can not be used in the same way as the distance between centre points, as the lowest minimal distance is 0 km (direct contiguity). I therefore propose a new type of contiguity matrix, which I shall refer to as $W_{mdcut}$, where $cut$ stands for the minimal-distance cut-off used. The elements of this new contiguity matrix take the form of equation , where $\delta_{ij} = [(cut + 50) - (mindist_{ij} \mid mindist_{ij} < cut)]$ and $mindist_{ij}$ is the distance of closest approach between $i$ and $j$. The cut-off value $cut$ is increased in steps of 50 km, from 100 to 950 km, leading to a total of 18 different minimal-distance contiguity matrices. This measure of contiguity enables me to incorporate the minimal-distance data and to retain a continuous measure at the same time. Compared to $W_{dist}$, these new contiguity matrices unfortunately have one shortcoming. As $W_{dist}$ contains inverse distances, there is nonlinearity in the suspected spatial relationship. $W_{mdcut}$, on the other hand, is completely linear. In order to see to this problem, I suggest another type of contiguity matrix, which uses $\delta_{ij} = [(cut + 50) - (mindist_{ij} \mid mindist_{ij} < cut)]^2$ in order to construct its elements. This final contiguity matrix is referred to as $W_{mscut}$. Appendix E shows two examples to clarify how contiguity matrices are set up.

As can be perceived, two different types of contiguity matrices are created. The matrices for primary contiguity include $W_{dum}$ and $W_{bor}$, while the measures of secondary contiguity include $W_{dist}$, $W_{mdcut}$ and $W_{mscut}$, with $cut \in \{100; 950\}$.These, together with an actual conflict indicator, are added to equation  to arrive at:

$$\begin{aligned} gr \quad = \quad & \beta_0 + \beta_1 \ln(inv_{it}) + \beta_2 \ln(sch_{it}) + \beta_3 \ln(n_{it} + g_{it} + \delta_{it}) + \beta_4 \ln(y0_{it}) \\ & + \beta_5(conf_{it}) + \beta_6 W_{prim}(conf_{it}) + \beta_7 W_{sec}(conf_{it}) + \varepsilon_{it} \qquad (3.8) \end{aligned}$$

In this equation, the subscripts $i$ and $t$ stand for the country and time period[45]. Furthermore, $gr$ is the growth rate of a country, $inv$ is the level of investment, $sch$ is the level of schooling, $n$ is the growth rate of the working population, $g$ is the exogenous rate of technological progress and $\delta$ is the depreciation rate. $y0$ is the initial level of income per capita, $W_{prim}$ is a weights matrix of primary contiguity, $W_{sec}$ is a weights matrix for secondary contiguity, $conf$ is a measure of conflict, $\beta_i$ are coefficients and $\varepsilon$ is an error term. The regression includes period-fixed effects in order to account for the specific occurrences of region-wide shocks, such as the oil crises. In order to accommodate the possibility that different parts of the continent are affected in different ways, I also

[45] Equation  is the equation used in a panel data setting. For the long-run regressions, the variables are no longer time-dependent.

experiment with the use of a dummy for Northern Africa.

## Data

As previous work has confirmed, conflict influences nations mostly in the short term. The long-term influence is limited due to the phoenix effect and the conditional convergence theory. In this paper, I therefore mostly focus on the short-term consequences of conflict on economic growth. The period 1960-2000 is divided into 8 five-year periods and I perform a panel analysis on these periods and a sample of African nations. The long-run regressions are included in the results section, but are not the main focus of this paper[46].

The data actually used in this analysis are taken from a range of sources. Most individual variables are updated versions of the variables used by Murdoch and Sandler (2004), but in this paper, I limit my scope to Africa only, whereas Murdoch and Sandler considered a worldwide sample. The dependent variable, $gr$, is defined as $(\ln(y5) - \ln(y1))$, and thus represents the total growth rate over the entire period. The data source for this variable is the World Development Indicators (Worldbank, 2007). The level of investment, $inv$, is taken from the Penn World Table 6.2 (Heston, Summers and Aten, 2006) and refers to the average annual share of investment per period.

Mankiw *et al*. (1992) argue that one can either use the level of human capital attainment or the amount of human capital accumulation over a period and I choose to use the former in this analysis. In fact, I utilise the data collected by Barro and Lee (2000)[47] that measures the percentage of population over 25 which has attained at least some secondary schooling. For each individual period, the level of schooling is observed in year 0 of the appropriate period and used in the analysis.

Working population growth $n$ is also derived from the World Development Indicators[48], where I use the average of the logarithmic difference between start- and end-years of a period: $n = \frac{(\ln(pop5) - \ln(pop1))}{5}$. The variables $(g + \delta)$ are assumed to be 0.05 in total for all countries, as is assumed by Mankiw *et al*. (1992) as well.

[46] As is explained later, the long-run growth regressions only refer to the period 1970-2000, due to data availability problems.
[47] The Barro-Lee dataset is available for download from http://www.cid.harvard.edu/ciddata/ciddata.html.
[48] The World Development Indicators (2007) provide information on population sizes for each year, as well as the percentage of the population that is in the working age range (15-65). The multiplication of these two variables is used as the working population size.

Unfortunately, as is often the case in analyses of economic growth in Africa, data availability is limited. This is a significant problem, particularly if a bias can be expected in the set of countries for which data are available and it seems straightforward that this is indeed the case. Particularly, countries which suffer from conflict are more likely not to be reporting data. Therefore, I try and combine the available data with data from other sources, in order to prevent the loss of the valuable data that are available[49]. The rule of thumb, used in constructing this dataset is that, when for an observation only one piece of data is missing, there is a reliable alternative source available[50] and it is possible to make a reliable estimation with the information available, the information is imputed. This way, a considerably larger dataset is created and more reliable results can be achieved. As a robustness check, it is obviously to be considered whether these imputed values significantly influence the results.

$Conf$ is a measure of conflict. For this measure, I am using different specifications in order to assess whether there are differences in the results depending on the conflict type I employ. From the UCDP/PRIO Armed Conflicts Dataset 4[51] I construct three conflict databases using different types of conflict. The first one, $all$, contains all conflicts included in the dataset[52], the second, $violent$, contains only those conflicts in the highest conflict intensity category[53] and the third definition, $civil$, includes only intra-national conflict[54]. Finally, I also set up another conflict measure, used only to test the outcomes of the regressions using the previous definitions. This fourth conflict measure is a civil conflict indicator on basis of the data from the Correlates of War (COW) project[55].

All four conflict indicators are utilised in two different ways. I set up databases with dummies reflecting whether or not any conflict is recorded in a particular five-year pe-

---

[49] When, for a particular country, there is information available on conflict, growth, population size and schooling, it seems such a waste to simply discard all this valuable information, simply because investment data are not available from the exact same source as for all the other countries.

[50] For economic growth, the Penn World Table 6.2 (Heston, Summers and Aten, 2006) is used as an alternative source (21 cases) or extrapolation of the WDI data (1 case). For the investment data, the Penn World table 5.7 is used as an alternative source (10 cases). Finally, the education data are supplemented with the other data collected by Barro and Lee (2000): Education attainment for people over 15 (7 cases) and United Nations (2007) data on literacy rates (70 cases).

[51] The UCDP/PRIO Armed Conflicts Dataset 4 is available for download from http://new.prio.no/CSCW-Datasets.

[52] In order to be included in the UCDP/PRIO dataset, conflict is defined as “a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths” (Harbom and Högbladh, 2006).

[53] The low intensity category is defined as “between 25 and 999 battle-related deaths in a given year”, while a conflict falls in the high-intensity category if there are more than 1000 battle-related deaths in a given year (Harbom and Högbladh, 2006).

[54] The UCDP/PRIO dataset defines four types of conflict: Extrasystemic armed conflict, interstate armed conflict, internal armed conflict and internationalized internal armed conflict (Harbom and Högbladh, 2006). For my variable on civil conflict, I consider only the conflicts defined as “internal armed conflict”.

[55] The data from the COW project are available on http://www.correlatesofwar.org/.

riod, but, in line with Murdoch and Sandler, I also test for conflict duration. In order to do this, a variable is created that contains the number of months a country was in conflict during each five-year period.

The final data needed for my regression analysis concerns geographical data to construct the different $W$-matrices. For direct contiguity and border length, I use the information from the *CIA World Fact Book* (CIA, 2006)[56]. In addition to that, the CIA also provides information about the coordinates of the centre points of countries. Using those, I calculate the distances between different nations. For the cutoff in the bounded distance $W$-matrix, I use the average distance between all country-pairs. Finally, Murdoch and Sandler used the Gleditsch and Ward (2001) dataset as a source for the minimal-distance data, but there are a number of inconsistencies and problems in that dataset, so I have chosen to set up a new dataset. Using the same labour-intensive strategy as Gleditsch and Ward, I use the freely available computer programme *Google Earth*[57] to find the minimal distances between country pairs.

# Results

## Short-run results

The first regression of which the results should be discussed is a very basic Mankiw *et al*. (1992) style regression, which includes only values for schooling, investment, labour growth and initial level of income. This result is shown in column 1 of table 6. Investment, schooling and the convergence term behave according to expectation, even though $y_0$ is not actually significant. Labour growth, on the other hand, unexpectedly shows a positive sign, but as the values are not significant, this should not pose much of a problem.

The next step, also shown in table 6, analyses the influence of host-country conflict. For each type of conflict, I run regressions with the number of conflict months and with a

[56] As often in spatial econometrics, islands present a problem. I deal with the issue on a case-by-case basis and propose that for the dummy-contiguity the following combinations are indicated as direct neighbours: Comoros-Madagascar, Comoros-Mozambique, Madagascar-Mozambique, Madagascar-Mauritius, Seychelles-Madagascar, Cape Verde-Senegal and Sao Tomé and Principe-Gabon. As for border lengths, I employ several formulas. The assumed border length influence of $i$ on $j$, when $i$ is either a coastal nation or an island and $j$ is an island is $\delta_{ij} = 100 \cdot \frac{coastline_i}{distance_{ij}}$; the border length influence of an island $i$ on coastal nation $j$ is $\delta_{ij} = coastline_i \cdot \frac{coastline_j}{\sum borderlength_j} \cdot \frac{distance_{ij} \cdot coastline_j}{\sum\limits_k (distance_{jk} \cdot coastline_k)}$, where $k$ stands for all the islands that are within reach of $j$.

[57] *Google Earth* can be downloaded from http://earth.google.com.

6.Baseline short-run regressions, using the violent conflict type

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | | *violent conflict* | | | |
| | *bench* | *confmonths* | *confdum* | *confmonths* | *confdum* |
| $constant$ | 0.171 | 0.169 | 0.148 | 0.220 | 0.203 |
| | (0.302) | (0.301) | (0.295) | (0.307) | (0.301) |
| $\ln(inv)$ | 0.057*** | 0.057*** | 0.053*** | 0.051*** | 0.045** |
| | (0.019) | (0.019) | (0.018) | (0.019) | (0.018) |
| $\ln(sch)$ | 0.031** | 0.031** | 0.030** | 0.039*** | 0.039*** |
| | (0.013) | (0.013) | (0.026) | (0.014) | (0.014) |
| $\ln(n+g+\delta)$ | 0.120 | 0.117 | 0.100 | 0.117 | 0.097 |
| | (0.096) | (0.095) | (0.091) | (0.094) | (0.090) |
| $\ln(y0)$ | -0.010 | -0.010 | -0.011 | -0.022 | -0.025 |
| | (0.017) | (0.017) | (0.017) | (0.020) | (0.020) |
| $conf$ | | | -0.066* | | -0.081** |
| | | | (0.038) | | (0.039) |
| $confmonths$ | | -0.033 | | -0.056 | |
| (x 100) | | (0.081) | | (0.083) | |
| $north$ | | | | 0.088*** | 0.102*** |
| | | | | (0.030) | (0.030) |
| $R^2$ | 0.147 | 0.147 | 0.161 | 0.162 | 0.181 |
| $N$ | 300 | 300 | 300 | 300 | 300 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

conflict dummy. In table 6, only the results for the $violent$ conflict type are shown, because these regressions give the most clear-cut results. Column 2 shows the result when the number of conflict months is included as an explanatory variable. This does not appear to influence the results at all and the conflict variable is insignificant. As can be seen in column 3 of table 6, this changes when a conflict dummy is included as an explanatory variable: this yields a significant result, implying that the occurrence of conflict within a nation's borders during a 5-year period decreases the 5-year growth rate by approximately 6.6 percentage points. The control variables still behave according to expectation, so this is not something to get worried about.

However, one could criticise the fact that Africa is considered as one region, even though there are a large differences in many respects between northern Africa and Africa south of the Sahara. To evaluate whether this is indeed a relevant consideration, a dummy for northern Africa, called $north$ is added and the result is shown in columns 4 and 5. These results add some significance to the control variables and increase the impact of the conflict dummy, but most of the results remain unchanged.

The next step is a virtual replication of the Murdoch and Sandler results, albeit with different data. In order to do this, I run 156 regressions for each type of conflict, varying $W_{conf}$, whether months-of-conflict or a conflict dummy is used and the inclusion of a northern Africa dummy. The results for the violent conflict type are displayed in table 7, in which columns 1, 3 and 5 are the optimal regressions without a northern Africa dummy and columns 2, 4 and 6 are the optimal regressions with a northern Africa dummy. To determine which of the contiguity matrices is to be used, $R^2$ is used to determine the most optimal fit. The reported results are not very strong. The control variables retain their desired outcomes, and host-country conflict is usually negative and significant. The spill-over effects of conflict on neighbouring countries' growth rates, on the other hand, is only (marginally) significant once. Overall, the regressions appear to indicate that host-country conflict has a negative impact of 4-5 percentage points, for the $all$ and $civil$ conflict indicators, while the $violent$ conflict indicator leads to an effect of 6.5-8.0 percentage points decrease in the five-year growth rate. Two other things should be noted from table 7. First, all the regressions are optimised when using the conflict dummy, instead of the number of months in conflict, which should imply that it is the presence and not the duration of conflict that does most of the damage. Second, all of the regressions arrive at $W_{dist}$ as their preferred contiguity matrix out of the 21 possible options.

While the results in table 7 are interesting, the difference between this and the results

7.Regressions using one type of conflict spill-over and varying over conflict type, spill-over type and the presence of a north-dummy

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *conflict type* | *all* | | *civil* | | *violent* | |
| $W_{conf} - type$ | **dist** | **dist** | **dist** | **dist** | **dist** | **dist** |
| $constant$ | 0.213 | 0.271 | 0.211 | 0.264 | 0.172 | 0.221 |
| | (0.309) | (0.314) | (0.313) | (0.318) | (0.300) | (0.304) |
| $\ln(inv)$ | 0.053*** | 0.045** | 0.053*** | 0.046** | 0.054*** | 0.047*** |
| | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) |
| $\ln(sch)$ | 0.029** | 0.038*** | 0.031** | 0.040*** | 0.032** | 0.041*** |
| | (0.013) | (0.014) | (0.013) | (0.014) | (0.014) | (0.014) |
| $\ln(n+g+\delta)$ | 0.110 | 0.109 | 0.112 | 0.112 | 0.101 | 0.098 |
| | (0.098) | (0.097) | (0.099) | (0.098) | (0.092) | (0.091) |
| $\ln(y0)$ | -0.011 | -0.025 | -0.012 | -0.025 | -0.012 | -0.025 |
| | (0.017) | (0.019) | (0.017) | (0.019) | (0.017) | (0.019) |
| $conf$ | -0.035 | -0.045** | -0.043* | -0.050** | -0.066* | -0.080** |
| | (0.022) | (0.021) | (0.025) | (0.025) | (0.039) | (0.040) |
| $confmonths$ (x 100) | | | | | | |
| $W_{conf} \cdot conf$ | -0.094 | -0.092 | -0.088 | -0.078 | -0.175* | -0.152 |
| | (0.082) | (0.079) | (0.077) | (0.075) | (0.102) | (0.098) |
| $W_{conf} \cdot confmonths$ | | | | | | |
| $north$ | | 0.096*** | | 0.092*** | | 0.098*** |
| | | (0.029) | | (0.029) | | (0.030) |
| $R^2$ | 0.158 | 0.176 | 0.162 | 0.178 | 0.168 | 0.186 |
| $N$ | 300 | 300 | 300 | 300 | 300 | 300 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

8.Comparison of the relevant results between those of Murdoch and Sandler (2002b) and the current paper

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *conflict type* | M&S(2002b) | | civil | | violent | |
| $W_{conf} - type$ | **md100** | **md100** | **md100** | **md100** | **md100** | **md100** |
| $conf$ | -0.084*** | | -0.046* | | -0.068* | |
| | (0.031) | | (0.024) | | (0.039) | |
| $confmonths$ | | -0.121* | | -0.071 | | -0.034 |
| (x 100) | | (0.071) | | (0.062) | | (0.082) |
| $W_{conf} \cdot conf$ | -0.109** | | 0.011 | | 0.035 | |
| | (0.047) | | (0.040) | | (0.047) | |
| $W_{conf} \cdot confmonths$ | | -0.151 | | 0.007 | | 0.014 |
| (x100) | | (0.108) | | (0.077) | | (0.108) |
| $R^2$ | 0.21 | 0.18 | 0.158 | 0.151 | 0.162 | 0.147 |
| $N$ | 235 | 235 | 300 | 300 | 300 | 300 |

*Note*: For each of these regressions, OLS regressions are used, with $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively. The regressions 3-6 use the adopt the same distance (100 km) as Murdoch and Sandler (2002b) reported by Murdoch and Sandler (2002b) are particularly noteworthy. The spill-over effects are not nearly as visible as they are in their analysis. In columns 1 and 2, table 8 shows the relevant parts of table IV of the Murdoch and Sandler (2002b) paper. The following columns use the $civil$ and $violent$ conflict indicators, in combination with the distance measure reported in the original study as yielding the optimal result (100 km), in order to see the difference. It is obvious that, while the host-country conflict indicator has the same sign, the spill-over effects do not. The current analysis does not yield any significant results, whereas Murdoch and Sandler at least found significant results for the conflict dummy. Neither limiting the sample to a one that is similar to that of Murdoch and Sandler, nor using the Correlates of War database makes any difference. Using cut-off distances different from their 100 km does not yield much of a result either. All in all, the difference between the results is striking.

As explained in the previous section, I now introduce an additional kind of conflict spill-over effect. Columns 1 and 2 of table 9 contain the results for the $general$ conflict type. Column one contains the result with a dummy conflict variable, the second column uses the number of conflict months and both include a $north$ dummy. The different columns report the optimal result of the 74 possible specifications (varying between dummy and borderlength primary spillovers and trying out 37 different kinds of secondary spill-over matrices), which is defined as the result with the highest $R^2$.

It is clearly visible that the levels of investment and schooling are significant determinants for economic growth. Population growth, on the other hand, is never significant and is actually of the wrong sign. The convergence term is not significant either, but it is at least of the correct sign. The host-country conflict indicator is only significant when the dummy conflict variable is applied, but the spill-over terms are highly significant in all specifications. Additionally, it should be noted that every specification is optimised at the same distance measures: dummies for the primary neighbours and the 250 km minimum distance measure. In order to calculate the impact conflict has on the growth rates in neighbouring states, one has to take into account the fact that each country has several neighbours and the conflict impact should thus be spread over all of these. For primary neighbours, this implies that the primary spill-over effect affects them only by $\frac{1}{4.08} = 0.245$ of the total influence, so that means that through this channel, a conflict leads to an average of $\frac{1}{4.08} \cdot (-0.366) = -0.090 = -9.0$ percentage points of growth reduction. However, the secondary spill-over effect, effects both secondary and primary neighbours, so that should be added to the first effect in order to arrive at reasonable estimates for the total effect[58]. The amount of influence a pri-

[58] It can be observed that the values for primary and secondary spillovers are in the same order, yet with

9.Short-run regressions, using two types of conflict measure and varying spill-over type and the types of conflict

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $type$ | all | | civil | | violent | |
| $W_{conf1} - type$ | **dum** | **dum** | **dum** | **dum** | **dum** | **bor** |
| $W_{conf2} - type$ | **md250** | **md250** | **md250** | **md250** | **md250** | **md250** |
| $constant$ | 0.234 | 0.248 | 0.261 | 0.238 | 0.215 | 0.262 |
| | (0.302) | (0.304) | (0.306) | (0.306) | (0.290) | (0.308) |
| $\ln(inv)$ | 0.047*** | 0.053*** | 0.047*** | 0.050** | 0.044** | 0.049** |
| | (0.018) | (0.020) | (0.018) | (0.020) | (0.018) | (0.019) |
| $\ln(sch)$ | 0.032** | 0.038** | 0.033** | 0.038** | 0.040*** | 0.040*** |
| | (0.014) | (0.015) | (0.014) | (0.015) | (0.014) | (0.014) |
| $\ln(n+g+\delta)$ | 0.116 | 0.114 | 0.118 | 0.112 | 0.092 | 0.120 |
| | (0.095) | (0.094) | (0.097) | (0.094) | (0.086) | (0.094) |
| $\ln(y0)$ | -0.022 | -0.028 | -0.024 | -0.026 | -0.031 | -0.028 |
| | (0.019) | (0.019) | (0.019) | (0.020) | (0.019) | (0.020) |
| $conf$ | -0.042** | | -0.053** | | -0.080** | |
| | (0.020) | | (0.023) | | (0.040) | |
| $confmonths$ | | -0.073 | | -0.091 | | -0.060 |
| (x 100) | | (0.059) | | (0.061) | | (0.084) |
| $W_{conf1} \cdot conf$ | -0.366*** | | -0.335** | | -0.607*** | |
| | (0.127) | | (0.133) | | (0.194) | |
| $W_{conf1} \cdot confmonths$ | | -0.823*** | | -0.537** | | -0.542*** |
| (x 100) | | (0.296) | | (0.264) | | (0.187) |
| $W_{conf2} \cdot conf$ | 0.421*** | | 0.377*** | | 0.655*** | |
| | (0.139) | | (0.139) | | (0.199) | |
| $W_{conf2} \cdot confmonths$ | | 0.855*** | | 0.561** | | 0.581*** |
| (x 100) | | (0.299) | | (0.269) | | (0.220) |
| $north$ | 0.094*** | 0.097*** | 0.094*** | 0.095*** | 0.112*** | 0.099*** |
| | (0.029) | (0.029) | (0.029) | (0.029) | (0.028) | (0.031) |
| $R^2$ | 0.200 | 0.191 | 0.195 | 0.177 | 0.210 | 0.180 |
| $N$ | 300 | 300 | 300 | 300 | 300 | 300 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

mary nation picks up from the secondary effect depends on the number of nations that fall within the given boundary and on the number of primary nations. In the case of the 250 km cutoff, there are, in addition to the $4.08$ primary neighbours, an average of $1.30$ secondary nations, who are an average minimum distance of $168$ kilometers away. The influence on a primary nation is equal to $\frac{300}{300 \cdot 4.08 + 1.30 \cdot (300-168)} \cdot 0.421 = 0.090 = 9.0$ percentage points. Combining the primary and secondary effects leads to the conclusion that primary nations actually have a benefit from conflict of $0.1$ percentage points[59]. Secondary neighbours, who only enjoy the secondary effect, have a positive effect of $\frac{(300-168)}{300 \cdot 4.08 + 1.30 \cdot (300-168)} \cdot 0.421 = 0.040 = 4.0$ percentage points. When looking at conflict duration and taking into account the average number of conflict months of $30.4$, the primary nations receive $-6.1 + 5.6 = -0.5$ percentage points and secondary neighbours benefit by $2.5$ percentage points.

The results for the $civil$ conflict type are similar to those for the $general$ conflict type, as is shown in columns 3 and 4 table 9. It becomes more interesting, however, when discussing the results for the $violent$ conflict-type in the final columns of table 9. The levels of investment and schooling are once again highly relevant. For interpretation purposes, the impact of an increase of one standard deviation in average investment leads to between $coef \cdot \left(\ln(\overline{inv} + \sigma_{inv}) - \ln\left(\overline{inv}\right)\right) = 0.044 \cdot (\ln(19.7) - \ln(10.52)) = 0.012 = 1.2$ and $0.049 \cdot (\ln(19.7) - \ln(10.52)) = 0.013 = 1.3$ percentage points of additional growth during a 5-year period. The education variable implies an increase of $coef \cdot \left(\ln(\overline{sch} + \sigma_{sch}) - \ln\left(\overline{sch}\right)\right) = 0.040 \cdot (\ln(52.7) - \ln(32.3)) = 0.008 = 0.8$ percentage points per 5-year period for a one standard deviation increase in average schooling. For the violent conflict type, the number of conflict months is again not significant in the host country. The dummy variable, on the hand, is significant and says that the existence of conflict leads to a growth reduction of $8.0$ percentage points. The spill-over effects are highly significant and imply the following for primary neighbours using the conflict dummy variable: $\frac{1}{4.08} \cdot (-0.607) + \frac{300}{300 \cdot 4.08 + 1.30 \cdot (300-168)} \cdot 0.655 = (-0.149) + 0.141 = -0.008 = -0.8$ percentage points. For secondary neighbours, $\frac{(300-168)}{300 \cdot 4.08 + 1.30 \cdot (300-168)} \cdot 0.655 = 0.062 = 6.2$ percentage points are gained. Using conflict months and taking into account the average $violent$ conflict length of $32.9$ months, primary neighbours suffer by $(-0.044) + 0.041 = -0.003 = -0.3$ percentage points and secondary neighbours by $1.8$ percentage points.

different signs. This should not be considered surprising, however, since the two are largely supposed to offset each other for primary neighbours, particularly with the optimal distances found in these results. With larger secondary distances, the point estimates of the absolute values of the coefficients diverge more.

[59] The difference is due to rounding.

## Short-run robustness checks

In order to find out whether any of the results found previously were the result of a spurious correlation, a number of robustness checks are executed, as presented in table 10. In the first two columns, the conflict indicator is changed to the Correlates of War (COW) indicator, which was also employed by Murdoch and Sandler. As can be seen, most of the control variables are of the expected sign and often significant. The second column, however, shows that the conflict indicator has little explanatory power, both for domestic conflict and conflict spillovers from neighbours. The first column, using the conflict dummy, also does not achieve significance for home-country conflict. For primary neighbours, the total impact is implied to be $-0.175+0.136 = -0.039 = -3.9$ percentage points, while secondary neighbours benefit by a total of $6.0$ percentage points per five-year period.

In the third column all observations where one of the data points had to be estimated are dropped. As a result, the number of observations drops from 300 to 191 and the control variables generally lose their level of significance. Of the different conflict-related indicators, only the secondary neighbour spill-over effect is still significant, but all other indicators at least have the expected sign. As mentioned earlier, it is not surprising that the results lose significance when dropping these observations. After all, the selection of countries that have insufficient data is not random, and is more likely to be skewed towards countries currently suffering from conflict.

Another possibility for a robustness check is explored in the two final columns, in which certain observations are dropped. In the fourth column all observations are dropped where $|gr| > 1$, to leave out countries which were hit by extremely large or small shocks. After analysing the impact of different countries, the final column covers the results of the analysis when Botswana is dropped.

The aforementioned columns show that control variables remain significant and the variables of interest also continue to be significant and of the expected sign. Column 4 implies $-0.4$ percentage points for primary neighbours and $+3.1$ percentage points for secondary neighbours, while the sixth column actually predicts $+0.4$ percentage points for primary neighbours and $+2.1$ percentage points for secondary ones.

The final robustness check is an application of an indicator of Ethno-Linguistic Affinity (ELA), as described in De Groot (2009). The indicator uses detailed data on entire population groups to measure the dissimilarity between groups and finally between peoples

10.Robustness checks for the short-run regressions, displaying results with Correlates of War data (columns 1 and 2), without estimated data (3), without extreme observations (4) and without Botswana (5)

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $type$ | CoW | | | all | |
| $W_{conf1} - type$ | **dum** | **bor** | **dum** | **dum** | **bor** |
| $W_{conf2} - type$ | **ms300** | **md450** | **md250** | **md250** | **md250** |
| $constant$ | 0.229 | 0.242 | 0.225 | -0.053 | 0.259 |
| | (0.318) | (0.315) | (0.364) | (0.118) | (0.294) |
| $\ln(inv)$ | 0.053*** | 0.050*** | 0.039* | 0.034*** | 0.039** |
| | (0.019) | (0.019) | (0.022) | (0.013) | (0.018) |
| $\ln(sch)$ | 0.042*** | 0.040*** | 0.032 | 0.027** | 0.032** |
| | (0.014) | (0.014) | (0.020) | (0.013) | (0.014) |
| $\ln(n + g + \delta)$ | 0.106 | 0.114 | 0.130 | 0.015 | 0.107 |
| | (0.101) | (0.097) | (0.113) | (0.032) | (0.090) |
| $\ln(y0)$ | -0.030 | -0.024 | -0.013 | -0.009 | -0.027 |
| | (0.184) | (0.020) | (.025 | (0.015) | (0.020) |
| $conf$ | -0.030 | | -0.042 | -0.032** | -0.045** |
| | (0.042) | | (0.031) | (0.015) | (0.020) |
| $confmonths$ | | -0.113 | | | |
| (x 100) | | (0.072) | | | |
| $W_{conf1} \cdot conf$ | -0.714*** | | -0.227 | -0.302*** | -0.182** |
| | (0.260) | | (0.139) | (0.106) | (0.078) |
| $W_{conf1} \cdot confmonths$ | | 0.463 | | | |
| (x 100) | | (0.296) | | | |
| $W_{conf2} \cdot conf$ | 0.657** | | 0.299** | 0.327*** | 0.224** |
| | (0.269) | | (0.148) | (0.117) | (0.094) |
| $W_{conf2} \cdot confmonths$ | | -0.798* | | | |
| (x 100) | | (0.452) | | | |
| $north$ | 0.115*** | 0.115*** | 0.107*** | 0.087*** | 0.114*** |
| | (0.027) | (0.033) | (0.037) | (0.025) | (0.029) |
| $R^2$ | 0.190 | 0.186 | 0.220 | 0.209 | 0.183 |
| $N$ | 300 | 300 | 191 | 296 | 292 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

in countries. It increases both with increasing homogeneity of the populations and with increasing similarity between the populations. The measure proposed in the aforementioned paper is general, but for convenience purposes I use the same applied measure as in the previous paper. It uses five underlying identity characteristics: language, language group, self-identified ethnic group, ethnic cluster and religious subgroup. The application of ELA is a little complicated, because it requires a combination of geographic distance and ELA. In order to do this, the elements of the contiguity matrix are replaced with:

$$e_{ij} = \frac{ELA_{ij} \circ \delta_{ij}}{\sum_{i=1}^{N} (ELA_{ij} \circ \delta_{ij})}$$

For this robustness check, the rest of the method is followed exactly as described previously. The results can be found in table 11, which replicate the results from table 9. The first thing that shows up is the small increase in $R^2$ in all regressions, except for the *violent* conflict type with monthly data (column 6). Furthermore, the changes are small in all control variables. The variables of interest, regarding spillovers, on the other hand, do change somewhat. For the regressions using monthly conflict data, primary neighbours' growth effects become more negative, while the secondary' neighbours' growth effects become more positive. For example, in the *all* conflict case, primary neighbours' growth effect goes from $-0.5$ to $-0.8$ percentage pioints and secondary neighbours' growth effect shifts from $2.5$ to $2.7$ percentage points. For the binary conflict variable, the results are less clear-cut. One important thing to note is that the point estimate for the positive effect of *violent* conflict is reduced from $6.2$ to $3.3$ percentage points.

The next stage of the analysis executed here is to look at more local effects. After all, one could expect that actual conflict location within a nation will influence the size of its effect on neighbouring countries. However, for such an analysis it would make sense to perform the entire analysis on a local level and while conflict data has recently become available (see, for example, Buhaug and Lujala, 2005) at a local level in Africa, growth data is not. Future research should look into the possibilities of adjusting the current methodology to account for conflict location, but it is not within the scope of this paper to do so.

11.Robustness checks for the short-run regressions, replicating table 4 with an Ethno-Linguistic Affinity indicator, in combination with distance measures

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $type$ | all | | civil | | violent | |
| $W_{conf1} - type$ | **dum** | **dum** | **dum** | **dum** | **bor** | **dum** |
| $W_{conf2} - type$ | **md250** | **md250** | **md250** | **md250** | **md250** | **md250** |
| $constant$ | 0.285 | 0.268 | 0.289 | 0.247 | 0.258 | 0.248 |
| | (0.308) | (0.308) | (0.311) | (0.309) | (0.298) | (0.305) |
| $\ln(inv)$ | 0.049*** | 0.054*** | 0.048** | 0.051** | 0.042** | 0.053*** |
| | (0.018) | (0.020) | (0.019) | (0.020) | (0.018) | (0.020) |
| $\ln(sch)$ | 0.036** | 0.041*** | 0.035** | 0.040*** | 0.040*** | 0.043*** |
| | (0.014) | (0.015) | (0.014) | (0.015) | (0.014) | (0.015) |
| $\ln(n+g+\delta)$ | 0.124 | 0.118 | 0.124 | 0.114 | 0.110 | 0.116 |
| | (0.098) | (0.096) | (0.099) | (0.096) | (0.090) | (0.093) |
| $\ln(y0)$ | -0.028 | -0.031 | -0.027 | -0.027 | -0.029 | -0.030 |
| | (0.019) | (0.019) | (0.019) | (0.020) | (0.019) | (0.020) |
| $conf$ | -0.043** | | -0.053** | | -0.080** | |
| | (0.020) | | (0.023) | | (0.040) | |
| $confmonths$ | | -0.069 | | -0.089 | | -0.051 |
| (x 100) | | -0.059 | | (0.061) | | (0.082) |
| $W_{conf1} \cdot conf$ | -0.424*** | | -0.357** | | -0.284*** | |
| | (0.145) | | (0.140) | | (0.093) | |
| $W_{conf1} \cdot confmonths$ | | -0.924*** | | -0.560* | | -1.025** |
| (x 100) | | (0.318) | | (0.305) | | (0.494) |
| $W_{conf2} \cdot conf$ | 0.455*** | | 0.389*** | | 0.347*** | |
| | (0.151) | | (0.144) | | (0.113) | |
| $W_{conf2} \cdot confmonths$ | | 0.934*** | | 0.571* | | 1.025** |
| (x 100) | | (0.351) | | (0.310) | | (0.515) |
| $north$ | 0.099*** | 0.100*** | 0.097*** | 0.096*** | 0.108*** | 0.098*** |
| | (0.029) | (0.029) | (0.029) | (0.029) | (0.113) | (0.029) |
| $R^2$ | 0.202 | 0.192 | 0.196 | 0.177 | 0.212 | 0.174 |
| $N$ | 300 | 300 | 300 | 300 | 300 | 300 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

## Long-run results

While this paper mainly focuses on the short-term effects that conflict may have on economic growth, the long-term effects should also be examined. Previously, Murdoch and Sandler (2002a, 2002b and 2004) concluded that there was little long-run influence of conflict. As mentioned earlier, the phoenix effect (Organski and Kugler, 1980) has often been cited as an argument why there should not be much long-run influence.

In this section, however, I will re-examine these results. The analysis is performed in a very similar way as discussed in the short-run section, except that a cross-section instead of panel data OLS regression can be run. One important problem is the fact that data availability during the 1960s is particularly limited. For that reason, the long-run analysis will not cover 1960-2000, but instead cover 1970-2000, which increases the number of observations from 23 to 39. The procedure followed for the short-run results are the same as well. The following tables contain the unidimensional spill-overs in columns 1 (for a conflict dummy) and 2 (for the number of conflict months). Following that, columns 3 (dummy) and 4 (months) contain the results when there are 2 types of spill-over. Finally, the last 2 columns (5 and 6) include a dummy for northern Africa.

Table 12 contains the results for the $all$ conflict type, table 13 shows the $civil$ conflict type and finally, table 14 represents the $violent$ conflict type. In table 12 we can see that the results do not look promising. Even the control variables do not reach significance often. It is interesting to see that in none of the cases the host-conflict variable was significant. This could be interpreted as support for the theory that the phoenix effect indeed exists, because the regressions in the previous section have shown some evidence for damage to countries' economies in the short run. The spill-over effects take on the expected signs, but only achieve significance in the case of columns 3 and particularly 4. The difference between these regressions and the regressions in columns 5 and 6, however, is striking. It is clear that the $north$ dummy variable absorbs all the significance that was found in the terms, so it is safe to say that the significance found in regressions 3 and 4 was most likely spurious[60]. As for table 13, the results possibly look even worse. None of the variables of interest ever become significant, although they do mostly have the expected signs. Possibly, these results can be interpreted as support for the phoenix factor hypothesis as well, but further testing would be necessary to conclude that without doubt.

[60] Just to complete the analysis of column 4, the implication of the significant terms is that primary neighbours actually have a weighted influence of 0.0 percentage points, which is the sum of -42.9 for the first term and +42.9 for the second. Secondary neighbours, on the other hand receive a positive pay-off of 15.0 percentage points.

12.Long-run regressions for general conflict type, varying over conflict spill-over type and the presence of a north dummy

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $W_{conf1} - type$ | **md400** | **md950** | **dum** | **dum** | **dum** | **dum** |
| $W_{conf2} - type$ | | | **md250** | **ms150** | **ms250** | **ms150** |
| $constant$ | -3.008 | -2.458 | -3.890 | -2.021 | -1.966 | -1.377 |
| | (3.841) | (3.844) | (3.715) | (4.361) | (3.692) | (4.459) |
| $\ln(inv)$ | 0.409** | 0.403** | 0.337** | 0.431** | 0.266* | 0.332* |
| | (0.151) | (0.165) | (0.143) | (0.175) | (0.140) | (0.174) |
| $\ln(sch)$ | 0.193* | 0.081 | 0.163* | 0.055 | 0.237** | 0.122 |
| | (0.099) | (0.116) | (0.090) | (0.130) | (0.099) | (0.126) |
| $\ln(n+g+\delta)$ | -0.575 | -0.627 | -0.775 | -0.440 | -0.515 | -0.435 |
| | (1.132) | (1.336) | (1.315) | (1.507) | (1.326) | (1.550) |
| $\ln(y0)$ | -0.087 | -0.041 | 0.009 | -0.032 | -0.164 | -0.136 |
| | (0.123) | (0.119) | (0.132) | (0.121) | (0.141) | (0.123) |
| $conf$ | -0.043 | | -0.050 | | -0.182 | |
| | (0.258) | | (0.224) | | (0.229) | |
| $confmonths$ | | -0.032 | | -0.019 | | -0.024 |
| (x 100) | | (0.058) | | (0.066) | | (0.059) |
| $W_{conf1} \cdot conf$ | 0.960 | | -3.566 | | -4.205 | |
| | (0.695) | | (2.150) | | (3.053) | |
| $W_{conf1} \cdot confmonths$ | | 0.177 | | -1.937* | | -1.905 |
| (x 100) | | (0.233) | | (1.024) | | (1.172) |
| $W_{conf2} \cdot conf$ | | | 4.535* | | 4.877 | |
| | | | (2.243) | | (3.003) | |
| $W_{conf2} \cdot confmonths$ | | | | 2.006* | | 1.955 |
| (x 100) | | | | (1.040) | | (1.185) |
| $north$ | | | | | 0.734*** | 0.649*** |
| | | | | | (0.219) | (0.214) |
| $R^2$ | 0.243 | 0.204 | 0.302 | 0.284 | 0.386 | 0.361 |
| $N$ | 39 | 39 | 39 | 39 | 39 | 39 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

13.Long-run regressions for civil conflict type, varying over conflict spill-over type and the presence of a north dummy

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $W_{conf1} - type$ | **dum** | **md950** | **dum** | **dum** | **dum** | **dum** |
| $W_{conf2} - type$ | | | **md200** | **md400** | **ms250** | **md250** |
| $constant$ | -2.976 | -3.908 | -3.526 | -3.899 | -2.214 | -1.841 |
| | (4.046) | (4.127) | (3.803) | (4.147) | (3.697) | (4.58) |
| $\ln(inv)$ | 0.415** | 0.412** | 0.437*** | 0.458** | 0.274* | 0.342* |
| | (0.160) | (0.169) | (0.153) | (0.193) | (0.140) | (0.185) |
| $\ln(sch)$ | 0.094 | 0.070 | 0.125 | 0.060 | 0.222** | 0.099 |
| | (0.099) | (0.117) | (0.098) | (0.121) | (0.097) | (0.124) |
| $\ln(n+g+\delta)$ | -0.904 | -1.057 | -0.863 | -1.112 | -0.655 | -0.596 |
| | (1.381) | (1.443) | (1.344) | (1.453) | (1.318) | (1.598) |
| $\ln(y0)$ | -0.039 | 0.001 | -0.005 | -0.027 | -0.164 | -0.121 |
| | (0.132) | (0.119) | (0.131) | (0.121) | (0.142) | (0.126) |
| $conf$ | -0.169 | | -0.058 | | -0.193 | |
| | (0.285) | | (0.252) | | (0.241) | |
| $confmonths$ | | -0.017 | | -0.042 | | -0.052 |
| (x 100) | | (0.060) | | (0.062) | | (0.055) |
| $W_{conf1} \cdot conf$ | -0.102 | | -4.073 | | -3.991 | |
| | (0.572) | | (2.544) | | (2.985) | |
| $W_{conf1} \cdot confmonths$ | | 0.274 | | -0.668 | | -0.921 |
| (x 100) | | (0.280) | | (0.498) | | (0.815) |
| $W_{conf2} \cdot conf$ | | | 4.164 | | 4.568 | |
| | | | (2.525) | | (2.903) | |
| $W_{conf2} \cdot confmonths$ | | | | 0.898 | | 1.009 |
| (x 100) | | | | (0.571) | | (0.831) |
| $north$ | | | | | 0.903*** | 0.646*** |
| | | | | | (0.296) | (0.209) |
| $R^2$ | 0.195 | 0.207 | 0.247 | 0.263 | 0.371 | 0.330 |
| $N$ | 39 | 39 | 39 | 39 | 39 | 39 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

14.Long-run regressions for violent conflict type, varying over conflict spill-over type and the presence of a north dummy

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $W_{conf1} - type$ | **md950** | **md950** | **dum** | **dum** | **dum** | **dum** |
| $W_{conf2} - type$ | | | **md300** | **ms150** | **md250** | **ms150** |
| $constant$ | -4.515 | -3.632 | -6.723 | -3.437 | -4.831 | -2.357 |
| | (4.197) | (4.144) | (4.005) | (3.781) | (3.679) | (3.732) |
| $\ln(inv)$ | 0.377** | 0.426** | 0.434** | 0.402** | 0.304* | 0.314* |
| | (0.162) | (0.168) | (0.162) | (0.162) | (0.164) | (0.154) |
| $\ln(sch)$ | 0.086 | 0.071 | 0.106 | 0.091 | 0.184* | 0.179* |
| | (0.107) | (0.115) | (0.111) | (0.114) | (0.105) | (0.100) |
| $\ln(n + g + \delta)$ | -1.293 | -0.986 | -1.963 | -0.956 | -1.597 | -0.815 |
| | (1.477) | (1.432) | (1.394) | (1.296) | (1.311) | (1.288) |
| $\ln(y0)$ | 0.009 | -0.017 | 0.044 | -0.030 | -0.084 | -0.160 |
| | (0.112) | (0.117) | (0.111) | (0.112) | (0.112) | (0.108) |
| $conf$ | -0.223 | | -0.150 | | -0.214 | |
| | (0.175) | | (0.168) | | (0.178) | |
| $confmonths$ | | 0.007 | | -0.061 | | -0.027 |
| (x 100) | | (0.092) | | (0.094) | | (0.079) |
| $W_{conf1} \cdot conf$ | 0.584 | | -2.132** | | -2.523** | |
| | (0.428) | | (1.028) | | (1.179) | |
| $W_{conf1} \cdot confmonths$ | | 0.366 | | -3.749*** | | -3.990*** |
| (x 100) | | (0.362) | | (0.775) | | (0.843) |
| $W_{conf2} \cdot conf$ | | | 2.773** | | 2.763** | |
| | | | (1.209) | | (1.301) | |
| $W_{conf2} \cdot confmonths$ | | | | 3.883*** | | 4.000*** |
| (x 100) | | | | (0.814) | | (0.869) |
| $north$ | | | | | 0.753*** | 0.723*** |
| | | | | | (0.259) | (0.211) |
| $R^2$ | 0.244 | 0.212 | 0.308 | 0.346 | 0.390 | 0.440 |
| $N$ | 39 | 39 | 39 | 39 | 39 | 39 |

*Note*: For each of these regressions, OLS regressions are used, with period fixed effects and $gr$ as dependent variable. White's heteroskedasticity-adjusted standard errors are reported between brackets and * indicates significance at the 10% level, while ** and *** indicate significance at 5% and 1% levels respectively.

Finally, table 14 actually reveals some interesting results. As before, host-country $violent$ conflict again turns out to have a negative, but insignificant influence. Once again, this can be taken as evidence that conflict-affected countries actually benefit from the post-war growth increase known as the phoenix effect. However, it is interesting to see that the neighbouring countries are significantly affected by the conflict, nonetheless. It has been theorised that neighbouring countries do not enjoy a phoenix effect like the actual conflict countries because the type of damage is different (or the neighbours actually benefit). Additionally, neighbouring countries may not receive aid the way the actual conflict countries do, as they are not seen as conflict victims by the international community. To see the influence of an additional conflict, it is necessary to look at the sum of the two different influencing terms.

In the case of column 5 (using a conflict-dummy variable), the model uses dummy spill-overs for the direct neighbours and applies the secondary effect for countries that have a distance of closest approach of maximally 250 kilometers. On average, a country has $\frac{216}{53} = 4.08$ primary neighbours and $\frac{69}{53} = 1.30$ secondary neighbours, who are on average $\frac{11577}{69} = 168$ kilometers away. The primary effect of $-2.523$ thus gives a negative effect per neighbour of $\frac{1}{4.08} \cdot (-2.523) = -0.619$. However, the primary neighbours also enjoy a share of the secondary effect: $\frac{300}{4.08 \cdot 300 + 1.30 \cdot (300-168)} \cdot 2.763 = 0.594$, adding these two factors together gives a total benefit for primary neighbours of $-0.619 + 0.594 = -0.025 = -2.5$ percentage points. Secondary neighbours are only affected by the second term and these thus receive a positive influence of $\frac{(300-168)}{4.08 \cdot 300 + 1.30 \cdot (300-168)} = 0.262 = 26.2$ percentage points for the average non-contiguous neighbour within a 250 kilometer minimum distance over a thirty-year period.

The calculations regarding the size of the influence in column 6 are complicated by the fact that the optimal results are achieved with the squared minimum distance factor. With the 150 kilometer minimum distance, the average nation still only has 4.08 direct neighbours, but it also has 0.42 secondary neighbours. $E\left[(200-\delta)^2\right] = 13984\, km^2$ for secondary neighbours and $40000\, km^2$ for primary neighbours. Of course, it is also necessary to take into account the length of an average $violent$ conflict, which is 82.5 months during the whole thirty-year period. The influence on primary neighbours consists of two elements again, of which the first element takes the value $82.5 \cdot \frac{1}{100} \cdot \frac{1}{4.08} \cdot (-3.990) = -0.808$, while the second element takes the value $82.5 \cdot \frac{40000}{\sum_i E[(200-\delta_i)^2]} = 0.782$, leading to a cumulative effect of $-0.808 + 0.782 = -0.026 = -2.6$ percentage points. The secondary neighbours benefit from the conflict by $82.5 \cdot \frac{E\left[(200-(\delta_i|\delta_i>0))^2\right]}{\sum_i E[(200-\delta_i)^2]} = 0.273 = 27.3$

percentage points over the entire thirty-year period. The data clearly show that there is a similar effect in the long run as there is in the short run, in which violent conflict (slightly) damages directly contiguous countries, whereas countries at greater distances actually benefit. For the other types of conflict, the results are much less clear-cut.

## Conclusion

This paper revisits the analyses of conflict spillovers executed by Murdoch and Sandler in their different papers, but reaches a different conclusion altogether. Previous results provided evidence that, particularly in Africa, countries in the general neighbourhood of countries suffering from conflict were influenced negatively as well. In this paper, on the other hand, I propose the hypothesis that proximate conflict is not necessarily all bad. In fact, it is suggested that there is a trade-off effect that benefits countries close to conflict, but punishes countries that are directly contiguous to it. This paper does not research the sources of the different spill-over effects, but this would be an interesting topic for future research.

The question remains why the results in this paper differ so much from the previous analyses. In part, this is due to the fact that a lot of new data have been used and the sample of countries is different[61]. This implies that the results found by Murdoch and Sandler would not hold with the updated data sources. In addition to this, the inclusion of a northern Africa dummy may have reduced the results of Murdoch and Sandler as well.

The most important difference between the previous analyses and the current one, however, is found in the contiguity matrix. Accounting for the possibility of a non-linear effect has a strong impact on the results, which can be observed in the difference between tables 17 and 9. The other modification I propose in the usage of the contiguity matrix is the way the minimal-distance database is exploited. Using the minimal-distance data in the way suggested by Murdoch and Sandler reduces the amount of information contained in it considerably. The contiguity matrix this paper proposes is simple and continuous, which is important in order to retain the available information.

The practical use of the current analysis is mainly found in the refutation of many of the policy recommendations made by Murdoch and Sandler in their previous papers.

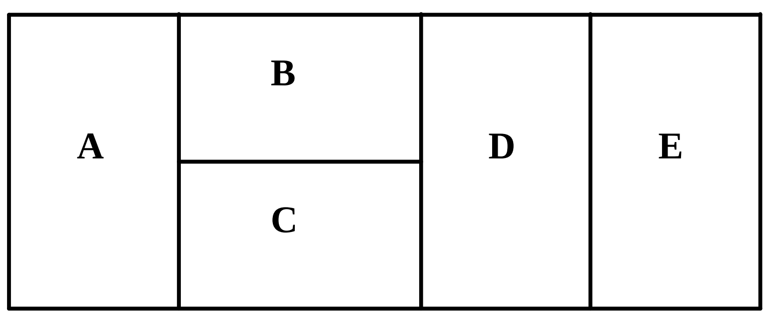
[61] I have a total of 300 observations, from 47 countries, while the paper by Murdoch and Sandler (2002b) that separates between different geographical regions has 235 observations from approximately 35-38 countries. The 2004 paper does not distinguish between different regions, but includes 217 observations for Africa, from 37 different countries.

Their recommendations that aid providers should consider entire conflict regions and their call for an extended geographical scope of these aid providers is invalid when the results of the current research are taken into account. I do agree, however, that directly contiguous neighbours should be considered as victims of violent conflict, even when there is no actual conflict within their own borders. In fact, the long-run results seem to indicate that the primary neighbours are more likely to suffer permanent damage than the host countries themselves. Another observation that can be derived from the results presented in this paper is that only major conflict matters. Smaller conflicts seem to have little effect, which should not really come as a surprise.

One important thing should be noted, however. The analysis could be improved upon a lot if better data were available, particularly concerning education. Education really is the bottleneck of the current analysis, as there is simply very little good information available and further research into that subject is adamant if this analysis were to be generalised. Finally, it is also uncertain whether the conclusions from this paper can be generalised to other continents as well. Murdoch and Sandler did a larger-scale comparison between the different continents and concluded that effects were most pronounced in Africa, which may imply that little result would be found if I performed a similar analysis for regions such as Latin America or Asia.

# E Shape of Contiguity Matrix W

In this appendix, an example will be given of the practical application of the contiguity matrix. As discussed in equation , the elements of a contiguity matrix are calculated as follows: $e_{ij} = \frac{\delta_{ij}}{\sum_i \delta_{ij}}$, where $\delta_{ij}$ is an operator for defining the distance between countries $i$ and $j$. Of the several types of indicators used in the current study, I show examples of how to set up the relevant contiguity matrices for the border-length contiguity and for a minimal-distance contiguity matrix. For this, a simple sample geography is used, with the following shape:




## Border-length contiguity

When border-length contiguity is used, the transition matrix, which is the first step in creating a contiguity matrix and contains the elements $\delta_{ij}$, looks as follows:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{bmatrix}, \text{ so row-summing, } \sum = \begin{bmatrix} 2 \\ 4 \\ 4 \\ 4 \\ 2 \end{bmatrix}, \text{ which makes it possible to}$$

arrive at the actual border-length contiguity matrix, by dividing the transition matrix of

the sum of the rows: $W_{bor} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$.

The contiguity matrix now shows the size of the influence of the different neighbours on the countries.

## Minimum-distance contiguity

In the case of minimum-distance contiguity, $\delta_{ij} = [(cut + \Lambda) - (mindist_{ij} \mid mindist_{ij} < cut)]$, where $\Lambda$ is a pre-determined size, depending on what is practical. In the actual analysis, $\Lambda = 50$, but in this example, $\Lambda = 0.5$. The distance matrix between all the nations looks as follows

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} - & 0 & 0 & 2 & 3 \\ 0 & - & 0 & 0 & 1 \\ 0 & 0 & - & 0 & 1 \\ 2 & 0 & 0 & - & 0 \\ 3 & 1 & 1 & 0 & - \end{bmatrix}$$

Taking a cut-off value of $1.5$ and $\Lambda = 0.5$, the transition matrix will look as follows:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} 0 & 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 2 & 1 \\ 2 & 2 & 0 & 2 & 1 \\ 0 & 2 & 2 & 0 & 2 \\ 0 & 1 & 1 & 2 & 0 \end{bmatrix}, \sum = \begin{bmatrix} 4 \\ 7 \\ 7 \\ 6 \\ 4 \end{bmatrix}, \text{ so } W_{md1.5} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{2}{7} & 0 & \frac{2}{7} & \frac{2}{7} & \frac{1}{7} \\ \frac{2}{7} & \frac{2}{7} & 0 & \frac{2}{7} & \frac{1}{7} \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \end{bmatrix}$$

In this case, the weights are obviously different from the weights found when using the border-length contiguity. Finally, as an example, taking a cut-off value of $2.5$ and $\Lambda = 0.5$, the following happens:

$$\begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} 0 & 3 & 3 & 1 & 0 \\ 3 & 0 & 3 & 3 & 2 \\ 3 & 3 & 0 & 3 & 2 \\ 1 & 3 & 3 & 0 & 3 \\ 0 & 2 & 2 & 3 & 0 \end{bmatrix}, \sum = \begin{bmatrix} 7 \\ 11 \\ 11 \\ 10 \\ 7 \end{bmatrix}, \text{so } W_{md2.5} = \begin{bmatrix} 0 & \frac{3}{7} & \frac{3}{7} & \frac{1}{7} & 0 \\ \frac{3}{11} & 0 & \frac{3}{11} & \frac{3}{11} & \frac{2}{11} \\ \frac{3}{11} & \frac{3}{11} & 0 & \frac{3}{11} & \frac{2}{11} \\ \frac{1}{10} & \frac{3}{10} & \frac{3}{10} & 0 & \frac{3}{10} \\ 0 & \frac{2}{7} & \frac{2}{7} & \frac{3}{7} & 0 \end{bmatrix}$$

# 4 The Influence of Conflict on the Demand for Education in the Basque Region

**Abstract**

It has previously been shown that civil conflict influences many economic factors, including education, which play an important role in development and economic growth. Previous authors working on the influence of conflict on education have, however, always focused strongly on the supply-side effects, whereas this paper examines the influence of conflict on the demand for education. It is theoretically shown that, under relatively general conditions, individuals living in a conflict area have an incentive to increase their level of education and that this effect depends on the individual's skill level. This hypothesis is then tested using the conflict in the Basque Region as a case study, which is an example of a conflict in which one would not expect strong supply-side effects. Using the other Spanish regions, an artificial region is created in which the population has a similar educational distribution as in the Basque Region. When comparing the true and artificial regions, it can clearly be seen that for individuals with a medium level of education, there is a strong incentive to increase their education level, which is in concordance with the theoretical model[62].

**Keywords**: Conflict; Education; Matching; Spain

**JEL code**: I21, D74, C15

## Introduction

The adverse effects of civil conflict on many factors that influence social well-being can be considered to be a well-established fact. For example, conflict researchers have previously examined the impact of conflict on health (Ghobarah *et al.*, 2004), human capital (Hoeffler and Reynal-Querol, 2003), physical capital (Abadie and Gardeazabal, 2003), economic growth (Murdoch and Sandler, 2002), and education (Lai and Thyne, 2007 and Ichino and Winter-Ember, 2004). Furthermore, the causes of civil war (Djankov and

[62] This chapter is co-authored by Idil Göksel from Bocconi University.

Reynal-Querol, 2007), particularly the impact of ethnicity (Fearon and Laitin, 2003) on conflict and its spillover effects (Murdoch and Sandler, 2002 and chapter 3 of this dissertation) are also topics discussed extensively in this literature.

On the other hand, it is well known that education plays an important role in economic growth (Barro, 2001 and Wong and Yip, 1999), equality (Gradstein, 2003), and the role of females in society (Schultz, 1995b). Moreover, returns to education (Garcia Prieto *et al.*, 2005 and Psacharopoulos and Patrinos, 2004) is regarded as an important topic as well as the basic consideration when making human capital investment decisions. Furthermore, education has an additional downstream effect: higher educated parents tend to have higher incentives to invest more on their children's human capital (Göksel, 2007).

Extending from the aforementioned literature, this paper aims to analyse the impact of conflict on the demand for education. Observing this impact will help to clarify the implications of conflict on this particular element of social well-being. It is important to stress that, unlike previous papers (Lai and Thyne, 2007, particularly), we aim to analyse the demand-side of education, instead of the combination of demand and supply-side effects. We have chosen the Basque conflict as a case study because this conflict has not had strong supply-side effects on education, which gives us the opportunity to look at the isolated influence of conflict in the demand for education.

In the following section of the paper, we provide a brief literature review. We introduce our theoretical model in the third section. In the fourth section, the empirical analysis and data are presented. The fifth section includes the results and the final section concludes.

## Literature Review

To our knowledge, the only other paper that investigates the impact of conflict on education is by Lai and Thyne (2007), in which they use cross-sectional and time series methods to analyse the issue. The authors consider two mechanisms: the first channel entails the fact that civil wars are likely to destroy a state's system of education through the loss of infrastructure and personnel. The second channel is the reallocation of resources from education to military expenses. Lai and Thyne use UNESCO education data for all states from 1980 until 1997 and examine the percentage change in educational expenditures for all education levels. They find evidence for their first claim, as both expenditures and enrolment decline during periods of civil war, but they do not find any proof for the reallocation of education funds towards military spending during conflicts. Finally, the supposed decrease in education expenses is only valid for higher level conflicts.

In their paper, Hoeffler and Reynal-Querol (2003) investigate the costs of conflict. They separate the costs into two subgroups consisting of economic and human costs. They also consider the long-term effects of civil war, taking into account the mortality rates among children, HIV in the military and the psychological damage of conflict. Surprisingly enough, they do not consider the impact of civil wars on education while performing these analyses.

According to Arrazola and De Hevia (2006), who use the Spanish Civil War as an Instrumental Variable to research the rates of return to education for men and women, there are three main reasons why educational attainment decreases during war periods. This includes difficulties in the physical access to schools, the decline in financial means for school attendance and need for children to leave school to earn money for their family. In this paper, we are interested in the impact of the Basque conflict, which is not an actively armed conflict and thus does not cause physical bans or damage to the schooling system. Our claim is that it still may influence the incentives to acquire education, due to changes in the returns to schooling. Basically, our main interest is the demand side of the schooling system instead of the supply side.

## Brain Gain versus Brain Drain

Our hypothesis is related to the ongoing debate of brain gain versus brain drain in migration literature. It has been claimed that migration decreases economic growth and the average education level of the source country due to the departure of the more highly educated and more intelligent people (Bhagwati and Hamada, 1974 and Haque and Kim, 1995). Building on that, Wong and Yip (1999) claim that brain drain reduces the economic growth rate and has a detrimental effect on non-emigrants in the source country through income-distributional effects and the reduced human capital accumulation. Wong and Yip show that if the initial rate of human capital accumulation is relatively low, a representative non-emigrant's sum of discounted income and life time utility could deteriorate. But recently researchers have started to show that migration may in fact have positive effects for the source country (e.g. Mountford, 1997 and Stark, 2004). Borjas (1994) in his paper investigates the economics of immigration in detail and evaluates both positive and negative consequences of migration. Using cross section evidence, Boucher *et al.* (2007) claim that the access of households to high-skill internal migration networks increases the likelihood that children will attend school beyond compulsory level. Furthermore, they provide evidence that if the returns to education are higher in the destination country, the incentive to invest in human capital will increase in the source country with a positive probability of migration. Likewise, Stark *et al.* (1997) claim that optimising workers in the source country will invest more in education if they have an opportunity to migrate and hence have higher expected returns to investment in human capital. Moreover, they find that costs of acquiring human capital are lower for more able workers than for less able ones. Both of these papers stress that only a proportion of the educated residents finally migrates and that therefore, in the end, the average level of education for the remaining population also increases. On the other hand, Beine *et al.* (2001) consider both beneficial and detrimental effects of migration. Their supposed "brain effect" refers to the improved incentives for investment in education as a result of improved migration opportunities, similar to the previously mentioned authors. The second impact of migration, referred to as the "drain effect" is the departure of some, if not all, educated agents. Beine *et al.* go on to claim that the sign of the total impact of the migration depends on which effect dominates.

As mentioned before, the aim of this paper is to analyse the impact of conflict on the demand for education. Civil conflict can have two different effects on education. It may either increase the incentives for education in order to be able to migrate out of the conflict region or reduce the incentives on human capital accumulation due to the problems caused by conflict and demoralisation.

## Methodological Literature

The methodology used in this paper is inspired by Abadie and Gardeazabal (2003), who construct a synthetic control region, which resembles the Basque region, in order to be able to compare the economic evolution of that artificial region to the Basque region during the conflict era. The artificial region is constructed as a weighted combination of other Spanish regions chosen to resemble the characteristics of the Basque country before the conflict. The authors find a 10% average GDP gap and provide evidence that changes in the per capita GDP gap are associated with the intensity of the conflict.

In the same style, Guidolin and La Ferrara (2007) employ a similar methodology to conduct an event study regarding the sudden end of the conflict in Angola due to the death of the rebel movement leader in 2002. They aim to find how the value of diamond mining firms responds to conflict episodes and to estimate the relationship between abnormal return and political tension. In order to achieve this, two portfolios are constructed: a portfolio with firms that have significiant Angolan interests, and a control portfolio that consists of companies that do not have such interests. The control group is formed by assigning weights to companies in order to minimise the Euclidean distance between two vectors containing the mean of abnormal returns, the variance of abnormal returns and the OLS beta of a world market portfolio model that regresses daily control returns on the world market for the period of four years before the end of the conflict. Their results show that the end of the conflict decreases the abnormal returns of the Angolan portfolio rather than increase them. The authors interpret this as a signal that incumbent firms benefit from the existence of conflict due to the barriers to entry and reduction in a government's bargaining power caused by the instability created by the civil war. Their paper is an important contribution to the literature not only because of the methodology employed, but also because it proves that conflict may positively impact the interests of some agents.

## Basque Conflict

Before continuing this paper, it is important to clarify the outline of the Basque conflict[63]. The main actor of the Basque conflict is ETA (Basque for "Basque Homeland and Freedom"), whose main aim is to promote establishment of an independent Basque country. Although it was founded in 1959, ETA did not claim its first victim until 1968 (Abadie and Gardeazabal, 2003). Since then, ETA has killed 823 people and committed dozens of kidnappings. Abadie and Gardeazabal show a table listing the number of killings and kidnappings by ETA between 1968 and 2000, which is reproduced here in table 15. From table 15, it can be seen that the number of killings and kidnappings were low before 1973, but started to increase during the mid 1970s, peaking during the years of 1978-1980 (235 victims). After 1980, the number of killings decreased gradually. During the 1980s, on average, ETA killed 39 people per year and this number is reduced to 19 per year during the 1990s. In September 1998, ETA declared a cease-fire, which lasted for 14 months and in 2000 ETA killed 23 people (Abadie and Gardeazabal, 2003). ETA's main financial sources are kidnapping, extortion and some robberies. For these activities, they have been targeting mainly the Basque entrepreneurs. Another

[63] This subsection including the description of the Basque conflict leans heavily on the description provided by Abadie and Gardeazabal (2003).

15.Number of deaths from ETA attacks

| year | deaths | year | deaths | year | deaths |
|---|---|---|---|---|---|
| *1968* | 2 | *1979* | 76 | *1990* | 25 |
| *1969* | 1 | *1980* | 92 | *1991* | 46 |
| *1970* | 0 | *1981* | 30 | *1992* | 26 |
| *1971* | 0 | *1982* | 37 | *1993* | 14 |
| *1972* | 1 | *1983* | 32 | *1994* | 13 |
| *1973* | 6 | *1984* | 32 | *1995* | 15 |
| *1974* | 19 | *1985* | 37 | *1996* | 5 |
| *1975* | 16 | *1986* | 41 | *1997* | 13 |
| *1976* | 17 | *1987* | 52 | *1998* | 6 |
| *1977* | 11 | *1988* | 19 | *1999* | 0 |
| *1978* | 67 | *1989* | 19 | *2000* | 23 |

*Note: table is a replication from Abadie and Gardeazabal (2003).*

interesting thing is the fact that although ETA conducts its activities in almost all Spanish regions, most of its activities are concentrated in the Basque Region. According to the calculations of Abadie and Gardeazabal, almost seventy percent of deaths cause by ETA in Spain during 1968-1997 took place in the Basque Region.

## Theoretical Model

For the current paper, we develop a naive theoretical model, which is merely used to outline our hypothesised changes in education as a result of a civil conflict. As mentioned earlier, the model is not generally applicable to simply any conflict, because, different from Lai and Thyne (2007), it focuses on the supply side of education, instead of the demand side. It is assumed that in the conflict region, as well as in the non-conflict region, education is sufficiently available and the level of education is determined primarily by the demands of the individual, instead of supply constraints. This assumption may not necessarily hold when considering longer time periods, as freely accessible (tertiary) education is something that has only developed during recent decades, but the assumption can be adjusted to read that there is no difference in the supply of education between the conflict area and the non-conflict area.

The outline of the model is as follows: An economy is populated by individuals who all have a particular level of ability. Using their level of ability, they decide to obtain a certain level of education and subsequently they decide whether to work in the home region, or whether to migrate and work in the migrant region. Individuals are fully rational and therefore make all their decisions through rational optimisation.

## Model Outline

As mentioned above, the population consists of a continuum of individuals. These individuals differ in only two respects. First, they have an ability level $A_i \in [0;1]$ and second, they are either susceptible to migration or not. In total, a proportion $\gamma$ is potentially interested in moving, while the rest $(1-\gamma)$ is not[64]. The first group is the group that we are interested in and whom our model concerns, whereas the latter is simply staying in the home region and not involved in any migration decision. The Ability level $A_i$ is the main determinant in the cost of acquiring education, where it obviously has a negative sign. The relationships between the Cost of Education ($CE_i$) and both $A_i$ and the chosen Education level $E_i \in [0;1]$ are convex, which yields the following type of cost function:

$$CE_i\left(E_i, A_i\right) = E_i^2 \cdot \left(2 - A_i\right)^2 \tag{4.1}$$

There are also other potential cost functions, but as long as it satisfies the initial conditions, the actual form does not significantly influence the results. During the first stage of the model, individuals use this cost function to determine their optimal level of education, taking into account the expected payoffs from education at a later stage. That second stage is when the individuals take up jobs and start working either in the domestic market or the migrants' market. The wages in these markets are the same in principle and depend on the level of education of an individual and a factor $\omega > 0$:

$$w_d\left(E_i\right) = w_m(E_i) = \left(1 + E_i\right) \cdot \omega \tag{4.2}$$

However, living as a migrant has one additional benefit. Individuals who live as migrants receive a bonus of $\lambda \in [-\infty;\infty]$, which represents the benefit of living out of your own region. This benefit can come from different sources, such as the expansion of job opportunities or an increased appreciation for highly developed skills. Another possibility, however, is that there is a peace bonus to living in another region, if indeed there

[64] This assumption is merely in place to avoid an outcome in which the entire population decides to migrate, which is unrealistic and undesirable.

is a civil conflict taking place in the home region. This is one of the channels through which we expect to see results. The second relevant channel is the costs incurred during migrating. These Costs of Migration ($CM_i$) consist of the fact that one is away from the home region, which may lead to problems of adjusting to a different culture and/or language. Particularly the language aspect is expected to be important and therefore, the Costs of Migration are decreasing in Education. One way this function could look is as follows:

$$CM_i(E_i) = (1 - E_i) \cdot \mu \tag{4.3}$$

where $\mu \geq 0$ is related to the difficulty of the adjustment process between the two different cultures.

The individuals have all this information available and decide whether to migrate or not. For this decision, they weigh the different levels of utility they would derive from migration and non-migration:

$$\begin{aligned} U_i &= (1 + E_d) \cdot \omega - E_d^2 \cdot (2 - A_i)^2 \text{ if } migr = 0 \\ U_i &= (1 + E_m) \cdot \omega - E_m^2 \cdot (2 - A_i)^2 + \lambda - (1 - E_m) \cdot \mu \text{ if } migr = 1 \end{aligned} \tag{4.4}$$

## Solution

It is easy to see that this model is simply a game with three stages:

1. Decide on education level $E_i$, given $A_i$.
2. Decide whether to migrate or not.
3. Work and earn.

We therefore solve the game through backward induction. At the third stage, no real decisions are taken and this stage is therefore ignored. At the second stage, however, there is an important decision to be made. The outcome of this decision depends on the Utility levels that can be derived from migration and from staying in the domestic region. To be more precise, defining $\widehat{E_d}$ and $\widehat{E_m}$ as the optimal levels of education under the assumption that the individuals are either working domestically or as migrants, an individual decides to migrate when the following is true

$$\left[\left(1+\widehat{E_m}\right)\cdot\omega-\widehat{E_m}^2\cdot(2-A_i)^2+\lambda-(1-\widehat{E_m})\cdot\mu\right]-\left[\left(1+\widehat{E_d}\right)\cdot\omega-\widehat{E_d}^2\cdot(2-A_i)^2\right]>0 \tag{4.5}$$

$$\lambda-\widehat{E_d}\cdot\omega+\widehat{E_m}\cdot(\mu+\omega)-\mu>\left(\widehat{E_m}^2-\widehat{E_d}^2\right)\cdot(2-A_i)^2 \tag{4.6}$$

$$\sqrt{\frac{\lambda-\widehat{E_d}\cdot\omega+\widehat{E_m}\cdot(\mu+\omega)-\mu}{\left(\widehat{E_m}^2-\widehat{E_d}^2\right)}}>(2-A_i) \tag{4.7}$$

$$\overline{A_i}=A_i>2-\sqrt{\frac{\lambda-\widehat{E_d}\cdot\omega+\widehat{E_m}\cdot(\mu+\omega)-\mu}{\left(\widehat{E_m}^2-\widehat{E_d}^2\right)}} \tag{4.8}$$

From equation , it can be seen that the decision to migrate depends on the equilibrium levels of Education, the individual Ability level and several parameters. An individual prefers to migrate when her threshold ability level ($\overline{A_i}$) is larger than the right-hand side of the equation. Taking this into account, we move back to the first stage of the game in which individuals decide on their education levels. We calculate separately what the equilibrium level of education is when migrating and when staying in the home region:

$$\max_{E_i} U_i=(1+E_d)\cdot\omega-E_d^2\cdot(2-A_i)^2 \text{ if } migr=0 \tag{4.9}$$

$$\max_{E_i} U_i=(1+E_m)\cdot\omega-E_m^2\cdot(2-A_i)^2+\lambda-(1-E_m)\cdot\mu \text{ if } migr=1 \tag{4.10}$$

Simply taking the First Order Conditions for each of these two expressions yields the following:

$$\begin{aligned}\frac{\partial U_{i,d}}{\partial E_d} &= \omega-2\cdot E_i\cdot(2-A_i)^2=0 \\ E_d &= \frac{\omega}{2\cdot(2-A_i)^2} \text{ if } migr=0\end{aligned} \tag{4.11}$$

$$\begin{aligned}\frac{\partial U_{i,m}}{\partial E_m} &= \omega-2\cdot E_i\cdot(2-A_i)^2+\mu=0 \\ E_m &= \frac{\omega+\mu}{2\cdot(2-A_i)^2} \text{ if } migr=1\end{aligned} \tag{4.12}$$

These equilibrium levels of education depend on the parameters of the model, and the individual level of Ability, $A_i$. It is important to note here that for a given level of Ability, an individual who decides to migrate will acquire a higher level of education than an individual who decides to stay at home, as long as $\mu > 0$. At this stage, we can insert expressions and into equation , to find the equilibrium after which individuals will find it more attractive to migrate and find work in the migration region:

$$\overline{A_i} > A_i = 2 - \sqrt{\frac{\lambda - \widehat{E_d} \cdot \omega + \widehat{E_m} \cdot (\mu + \omega) - \mu}{\left(\widehat{E_m}^2 - \widehat{E_d}^2\right)}} = 2 - \sqrt{\frac{\lambda - \frac{\omega^2}{2\cdot(2-A_i)^2} + \frac{(\omega+\mu)^2}{2\cdot(2-A_i)^2} - \mu}{\left(\left(\frac{\omega+\mu}{2\cdot(2-A_i)^2}\right)^2 - \left(\frac{\omega}{2\cdot(2-A_i)^2}\right)^2\right)}}$$

$$\lambda - \frac{\omega^2}{2\cdot(2-A_i)^2} + \frac{(\omega+\mu)^2}{2\cdot(2-A_i)^2} - \mu > (2-A_i)^2 \cdot \left(\left(\frac{\omega+\mu}{2\cdot(2-A_i)^2}\right)^2 - \left(\frac{\omega}{2\cdot(2-A_i)^2}\right)^2\right)$$

$$\lambda - \mu + \frac{1}{2} \cdot \frac{(\omega+\mu)^2 - \omega^2}{(2-A_i)^2} > \frac{1}{4} \cdot \frac{(\omega+\mu)^2 - (\omega)^2}{(2-A_i)^2}$$

$$\overline{A_i} > A_i = \left(2 - \frac{1}{2} \cdot \sqrt{\frac{(\omega)^2 - (\omega+\mu)^2}{\lambda - \mu}}\right) \tag{4.13}$$

So, apart from a number of coefficients, the decision whether or not to migrate depends solely on the level of Ability.

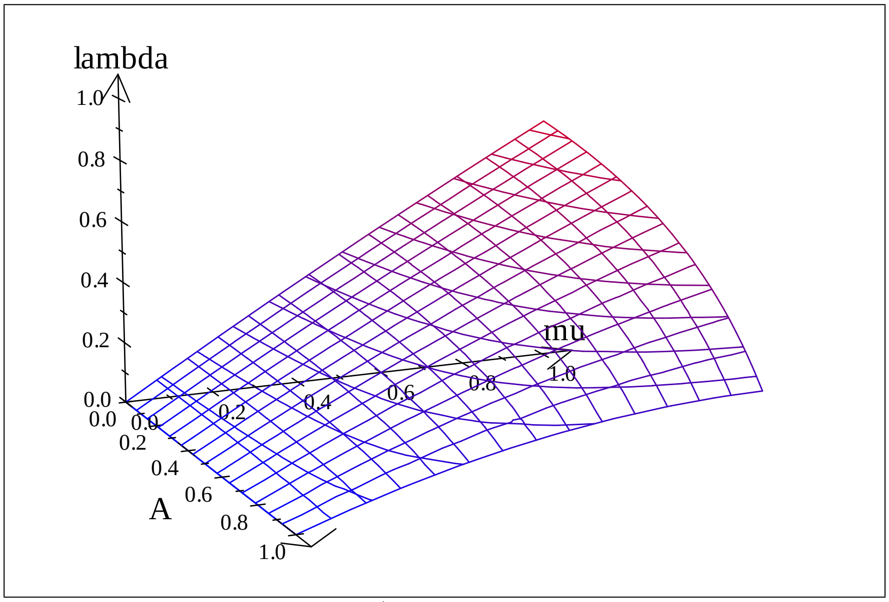
## Interpretation

Unfortunately, the interpretation of equation is not very straightforward at first sight. There is, however, a way of graphically showing the implications from this equation. If we assume unity wage, so $\omega = 1$, equation becomes:

$$\overline{A_i} > A_i = \left(2 - \frac{1}{2} \cdot \sqrt{\frac{1 - (1+\mu)^2}{\lambda - \mu}}\right) \tag{4.14}$$

There are now only two coefficients left ($\lambda$ and $\mu$), in addition to the level of Ability $A_i$. The way this can be represented is by looking at the combinations $\lambda$ and $\mu$ that will give a particular threshold level of Ability above which it is more attractive to migrate. In graph

, the threshold for choosing to migrate is shown as function the relevant characteristics. As can be seen, given $\lambda$, an increased level of $\mu$ will require a higher Ability level in order to migrate. Conversely, keeping $\mu$ constant, a higher $\lambda$ indicates that the cut-off level of Ability is lower.




*The combinations of $\lambda$, $\mu$ and $A_i$ where individuals decide to migrate or not.*

It is important to stop for a moment and consider the interpretation of graph . Looking at an example, we know that with $\lambda = 0.4$ and $\mu = 0.4$, even individuals who have $A_i = 0$ are interested in migrating and working in the foreign region. On the other hand, when $\lambda = 0.2$ and $\mu = 0.5$, not even the top of the Ability distribution is going to be willing to migrate and work abroad. Of course, it would be expected that the real values of $\lambda$ and $\mu$ are somewhere in the middle of these extreme examples.

The implications of graph  should be obvious in the context of the current paper. An outbreak of a relatively small-scale civil conflict in the domestic region is going to increase $\lambda$, which leads to an increase in migration, as long as the ex ante equilibrium is somewhere in the medium region. However, an autonomous shift of $\lambda$ will not merely change the size of migration, but also the composition. After all, when e.g. $\lambda \approx 0.36$ and $\mu = 0.5$, any individual who is potentially interested in migrating[65] needs $A_i \geq 0.5$ in order for him to have an incentive to migrate. This implies that the level of education of migrants is distributed between $0.5$ and $1$. Now, if due to the occurrence of a civil

---

[65] Remember that we said that only a portion $\gamma$ of the population is potentially interested. This is to make sure that there is no situation in which the entire population chooses to migrate.

conflict, $\lambda$ increases and becomes $\lambda_{new} \approx 0.40$, the threshold level of Ability will go down to $0.3$. This leads to an increase in migration, but also to a reduction of the Ability level, which will now be distributed between $0.3$ and $1$.

The implication for education can be seen as follows: Of those individuals born in the domestic region, the increase in potential migrants will in fact increase the over-all level of education. The average level of education for migrants, on the other hand, is going to fall, due to the increased presence of low-Ability individuals in the total pool of migrants.

### The Basque case

The case of the Basque civil conflict fits into the current model very well, particularly because Spain and the Basque region are politically unified. This reduces the Cost of Migration and increases the percentage of people who would potentially be interested in migrating. In this particular case, it is only the cultural and linguistic differences that make the move more difficult. Another major difference with the standard migration literature (e.g. Borjas, 1994), is the fact that there are no legal differences between the education systems of the Basque and Spanish regions, so the problem that highly educated migrants end up doing lower class jobs does not take place here.

During the conflict, there have been no significant effects on the supply of education and any changes in education achievement are due to demand changes. In the next section, we analyse empirically what the presence of the Basque conflict did to the education achievements of Basque-born individuals. For this, we could look at the average educational achievement, but according to the theoretical model described here, the effects should be stronger in one particular part of the educational distribution. In fact, assuming that the combinations of $\lambda$ and $\mu$ do not take on extreme values that would lead to no or full migration, it can be expected that the shift as a result of the conflict takes place in the middle part of the educational distribution. This leads to an increased overall level of education and to a lower level of education for the population of domestic migrants from the Basque Region.

## Empirical Analysis

In the previous section, we described a hypothesis on how the presence of civil conflict may influence education outcomes, even when, contrary to Lai and Thyne (2007), a conflict does not influence the supply of education. To empirically test whether this hypothesis is true, we employ a method based on Abadie and Gardeazabal (2003) and in this section we explain how. In a nutshell, the idea is that we use a dataset on realised education for people born in different Spanish provinces and in different years, which is then used in an ordered probit analysis to take out the most obvious explanatory factors for education accomplishments. This first stage yields a dataset on "residual education" for individuals from all different Spanish regions. At the second stage, the dataset is split into one pre-conflict dataset and one dataset that takes place during the conflict. Using the pre-conflict dataset, we then apply a matching method to create an artificial region that exhibits the same characteristics as the Basque region, pre-conflict. After doing this, we compare the distributions of education in the true and artificial Basque Regions during the conflict, from which it is possible to conclude what the effect of the conflict on education is.

In the following subsections, the details of the analysis are explained more precisely.

## First Stage Regression

Different from the Abadie and Gardeazabal (2003) analysis, we choose to filter out the alternative explanatory factors of education level before creating the artificial region during the second stage. The dataset we employ are the 1991 and 2001 Census results, which report, among many other things, the realised education level of all interviewed individuals. Of course, in order to make sure that individuals will have completed their entire education, only those individuals who have reached the age of 25 at the time of the Census are included.

The education levels are given in levels between 1 and 10, representing everything from $illiterate$ until $PhD - level$. As these are categorical non-continuous observations, the use of Ordinary Least Squares is ruled out. Instead, we perform an ordered probit analysis, with education level as the dependent variable. The independent variables are the most ex ante obvious variables to explain the level of education. In an ideal situation, this would include variables like the wealth status of the individual's family at the time of birth or the parental level of education, but unfortunately these data are impossible to obtain. The only totally exogenous explanatory variables that are available at the individual level are year of birth and gender. So while it is not possible to add further individual-specific control variables, there are several variables at the provincial level that add a significant amount of explanatory power. First of all, there is Provincial

GDP level. As this may be endogenous to the level of education of the population, it is important to have an ex ante level of GDP. Unfortunately, due to the specifics of the Spanish situation, there is very little available in terms of old provincial data and we therefore use the oldest trustworthy source of data that can be found: GDP in 1967. Most importantly, this year is before the actual start of the conflict and can therefore be considered as a relevant explanatory variable.

The other provincial-level variables are related to the supply of education. After all, obtaining a higher level of education is largely dependent on the availability of educational institutions. As a proxy for the availability of lower level of education, we use population density. After all, primary and secondary education are nearly universally available as long as there is a minimal mass of people. The only possible impediment to accessing these levels of education is the potential travel time to and from schools, which, again, is directly related to the population density in a province. In order to make sure there are no endogeneity issues, we use the population density at the provincial level at the nearest decade (1930, 1940, etc) before birth. For tertiary education, we have used a more sophisticated method and set up a database on the Spanish Higher Education. In subsection 4.3, there is further explanation concerning the data collection, but it suffices to say that we use a dummy variable that says whether or not an institution of higher education is available in the province, 18 years after birth. Finally, in order to pick up any other province-specific effects, we add province-dummies for each of the 50 provinces in our dataset[66].

Empirically, the Ordered Probit regression we estimate during the first stage looks as follows:

$$\begin{aligned}\Pr(edu_{i,y,p} &= lvl) = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{50} \end{bmatrix} \begin{bmatrix} D_1 & \cdots & D_{50} \end{bmatrix} + \beta_{51} \cdot sex_i + \beta_{52} \cdot birth_i + \\ &\beta_{53} \cdot \ln\left(gdp_{1967,p}\right) + \beta_{54} \cdot dens_{p,|y|} + \beta_{55} \cdot uni_{p,y+18} + \varepsilon_{i,y,p}\end{aligned} \tag{4.15}$$

where $edu_{i,y,p}$ is the level of education of individual $i$, born in year $y$ in province $p$. $lvl$ represents the possible level outcomes for education[67]. $D_1 \cdots D_{50}$ are the provincial dummies, $sex_i$ is a dummy variable taking value 1 when an individual is male, $birth_i$

[66] As explained at a later stage, the provinces of Ceuta and Melilla are dropped due to their specific status.
[67] This is an ordinal variable, which is why we use the Ordered Probit methodology. The situation would be different if the education level were, for example, measured in total years of schooling. The total number of levels is ten.

is the year of birth of the individual, $gdp_{1967,p}$ is the 1967 GDP level in the province of birth, $dens_{p,|y|}$ is the population density in the nearest full decade before birth and finally, $uni_{y+18,p}$ is a dummy variable that takes value 1 when a university is present in the birth province when the individual reaches age 18.

After running the regression in equation , we have probabilities for each educational outcome for every single individual. Using a simple measure, we then calculate the so-called residual education, $res_i$ which is taken to the second stage of the analysis:

$$res_i = edu_i - \frac{\sum_{\eta=1}^{10} \Pr(\eta_i)}{10} \tag{4.16}$$

where $\eta_i$ are the different levels of education that are obtainable.

## Second Stage

The inspiration for the second stage comes from the paper by Abadie and Gardeazabel (2003), who analyse the influence of the Basque conflict on economic growth. However, there are a number of significant differences between their methodology and ours. Abadie and Gardeazabel construct an artificial region, with all the non-Basque regions as potential elements. While we do the same in principle, an important distinction is the fact that we attempt to replicate the entire distribution of educational achievement, and not simply the average. For that reason, we split up all our data in ten different deciles, and perform the matching analysis upon each separate one. The next important difference is that Abadie and Gardeazabel replicate the explanatory factors of economic growth (such as the physical and human capital stocks) and then look at the resulting levels of GDP. As we have taken out the (few) obvious explanatory variables in the first stage of the analysis, it is feasible for us to simply create a matching of the outcome (level of education).

First, with $D_J$ as the total number of deciles from all $J$ potentially contributing regions to the artificial region ($D_J = 10 \cdot J$), we define $\mathbf{W} = (w_1, \ldots, w_J)$ as a $(D_J \times 1)$ vector of weights for each decile of each contributing region $j$. $W = \{((w_1, \ldots, w_J))\}$ is the set of possible different combinations of $w_j$, under the conditions that $w_1 + \ldots + w_J = 1$ and $w_j \geq 0 \, \forall \, j = 1, \ldots, J$. $\mathbf{Z}_1$ is a $T \times 1$ vector containing the educational outcomes for the decile under analysis, where $T$ is the number of pre-conflict time periods used. $\mathbf{Z}_0$ is a $T \times D_J$ matrix which contains the same outcomes for all $J$ regions during all $T$ time periods. We then use the following method to find the outcome of $\mathbf{W}$ that minimizes the difference between the real Basque region and the artificial one:

$$\mathbf{W}^* = \underset{w \in W}{\arg\min} \left(\mathbf{Z}_1 - \mathbf{Z}_0\mathbf{W}\right)' \left(\mathbf{Z}_1 - \mathbf{Z}_0\mathbf{W}\right) \tag{4.17}$$

This method is repeated ten times for each decile in order to get estimations for each one. As a basis, we utilise the data from people born between 1930 and 1955 as the source of data for the pre-conflict births. As the conflict breaks out around 1970, this means that all individuals who reach age 15 after the initiation of the conflict are assumed to be influenced by its presence. As this choice may seem rather arbitrary, we show in the robustness checks that the results do not change significantly for other reasonable pre- and post-war assumptions. The post-war generation, as discussed in the next section, finishes with the generation born in 1976, after which there is no data available on completed educations.

## Data

In this section, we give a description of our data sources and a description of what is done to make them suitable for this analysis. The dependent variable is the accomplished education level, which runs, as said before, from 1 until 10. Like all the individual data, these data come from the National Censuses in 1991 and 2001 (Instituto Nacional de Estadistica, 1991, 2001). Unfortunately, the two Censuses use different definitions of the Education system, but using information from the Ministry of Education (retrieved from http://www.mec.es), we can transform the fifteen categories in the 1991 Census to conform to the ten categories in the 2001 Census[68]. The 1991 Census used a 10% sample of the population, yielding a total of 3,888,692 observations and the 2001 Census interviewed 2,039,274 individuals, giving us a total of 5,927,966 observations. However, we drop those individuals who are born abroad or in either Ceuta or Melilla, the two small city-regions on the North-African coast and retain 5,714,097 individuals. The next step

[68] More particularly, the categories “Formacion Professional 2° grado, Maestria Industria” and “Otras titulaciones media” were combined to become “FP Grado Superio”. “Arquitecto e Ingeniero Técnico y Diplomado (aprovado compieto 3er curso) de Escuelas Téchnicas Superiores”, “Diplomado de Escuelas Universitarias y Diplomado (aprobado completo 3er curso) de Facultades y Colegios Universitarios”, “Arquitecto o Ingeniero Superior” and “Titulaciones de Estudios Superiores no universitarios” form “ Diplomatura” and finally, “Titulaciones de Estudios de Posgrado o Especializacion para Licenciados” is added to the category of “Doctorado”.

is dropping the individuals who may not yet have completed their entire education. Assuming that the maximum level of education is in principle reached at the age of 25, we drop all observations aged less than 25 at the time of their Census interview, thus retaining 3,842,997 observations. As can be seen in the results section, we experiment with the starting date for the analysis and drop more observations that way, but in principle we can use nearly 4 million observations.

From the Census, we also retrieve the gender, the birth year and the birth province upon which the provincial dummies are based. Next up is the provincial density, which proxies the availability of primary and secondary education. These data are also retrieved from the Instituto Nacional de Estadistica's Census results (2008). Goerlich Gisbert and Mas Ivars (2001) are the source for the 1967 GDP data.

The final variable in the first-stage ordered probit regression is the dummy for whether or not there is an institution of Higher Learning in the Province of birth. To create this variable, we obtained a list of current institutions of Higher Learning from the Ministry of Education. We then performed a web-based research on all the individual institutions and set up a database on their respective histories. This database includes the founding dates of each of the institutions, and whether they are follow-ups of other institutions. If they are, we then looked up the founding dates of the previous institutions. For most institutions, this method worked very well and for those few for which it did not, we contacted the institutions directly to obtain the required data. Overall, this yielded a database with an observation for each province-year and whether or not an institution of higher learning was present.

## Results

As explained in the previous section, we start with a first-stage regression that aims to take out the most obvious effects that influence the level of education. The results of this regression, following equation , are shown in table 16. The different columns all use different sample periods, which is going to be useful at the next stage when it has to be determined which time period is most appropriate. It can already be seen that the differences in the coefficients between different periods is relatively small, so the sample selection is not expected to have a strong influence.

A disadvantage of using ordered probit analysis is the difficulty of interpreting the coefficients. Therefore, for ease of interpretation, an Ordinary Least Squares (OLS) regression

16.Results from the first-stage Ordered Probit regression

| | all_data | >1909 | >1929 | >1934 | OLS, >1909 |
|---|---|---|---|---|---|
| $sex$ | 0.163*** | 0.160*** | 0.127*** | 0.108*** | 0.208*** |
| | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 |
| $birthyear$ | 0.040*** | 0.041*** | 0.047*** | 0.047*** | 0.054*** |
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\ln(gdp_{1967})$ | 1.500*** | 1.504*** | 1.536*** | 1.503*** | 2.022*** |
| | 0.011 | 0.011 | 0.011 | 0.012 | 0.018 |
| $\ln(popdens)$ | -0.284*** | -0.317*** | -0.484*** | -0.483*** | -0.101*** |
| | 0.004 | 0.005 | 0.006 | 0.006 | 0.008 |
| $unipres$ | 0.051*** | 0.041*** | 0.020*** | 0.034*** | 0.128*** |
| | 0.002 | 0.002 | 0.002 | 0.003 | 0.004 |
| $prov\_dummies$ | $yes$ | $yes$ | $yes$ | $yes$ | $yes$ |
| $N$ | 3834854 | 3736709 | 2958818 | 2640918 | 3736709 |
| $LR - \chi^2$ | 1620372.7 | 1519346.1 | 922046.3 | 684278.3 | - |
| *df* | 53 | 53 | 53 | 53 | $\overline{R^2} = 0.26$ |

*Note: The results of the first stage regression show the values of some of the most obvious explanatory variables that are able to explain long-run education achievements. The table shows coefficients and standard errors and *** implies a significance level of more than 99%.*

is added in table 16, of which the coefficients do not necessarily convey much meaning, but the signs of the coefficients do. Unsurprisingly, sex has a positive effect (with men given the value 1 for this dummy variable) and the year of birth does as well. As expected, richer regions also have higher levels of education. The effect of population density, on the other hand, has an effect opposite to what might be expected. After all, it was hypothesised that an increase in density should be associated with easier access to schools and therefore an increase in education. It turns out that this effect is actually going in the opposite direction and an increase in population density is in fact associated with a lower level of education.

Following equation , we calculate the residuals of the level of education and end up with our variable referred to as “residual education”, with which we continue to go on to the second stage of our analysis. At this second stage, we use equation  to determine the values for the artificial region. The different components that make up the artificial Basque Region are shown in table 17. Our artificial region is composed of very different elements than Abadie and Gardeazabal’s, whose artificial region consists of only Catalunya and Madrid. For our results, on the other hand, all regions, except the Balearic

Islands, contribute to the behaviour of the synthetic Basque region. For conciseness, we have summarised the data, and we show for the artificial low, medium and high levels of education what the contributing regions are. As for the contributing deciles, the table shows whether the contribut**ing** decile is equal to the contribut**ed** decile (*same)*, whether it is from one of the lower deciles (*lower*) or whether it is from a higher decile (*higher*). The final column shows the contribution of all contributing elements to all deciles. For example, when creating the lowest three deciles, 54.2% of the contributions come from same-level deciles in other regions and particularly Madrid and Navarra play a large role. At the same time, 33.2% of the contributions to the lowest deciles come from higher-ranked deciles (Cantabria and Navarra, in particular), while 12.6% come from lower-ranked deciles[69]. Of course, this is only a summary measure and the more detailed information could be interesting too. The contributors can be quite surprising in fact. For example, for the recreation of the fifth decile, the contributors are as follows: Navarra decile 1 (13.7%), Cantabria dec.3 (1.9%), Galicia dec.3 (18.1%), Navarra dec.3 (10.2%), Canary Islands dec.4 (8.2%), Catalunya dec.5 (26.3%), Cantabria dec.6 (4.3%), Aragon dec.6 (9.6%), La Rioja dec.6 (5.9%) and Madrid dec.10 (1.9%). Overall, the greatest contributors are Navarra with 28.9%, Madrid with 14.0% and Asturias with 9.4%

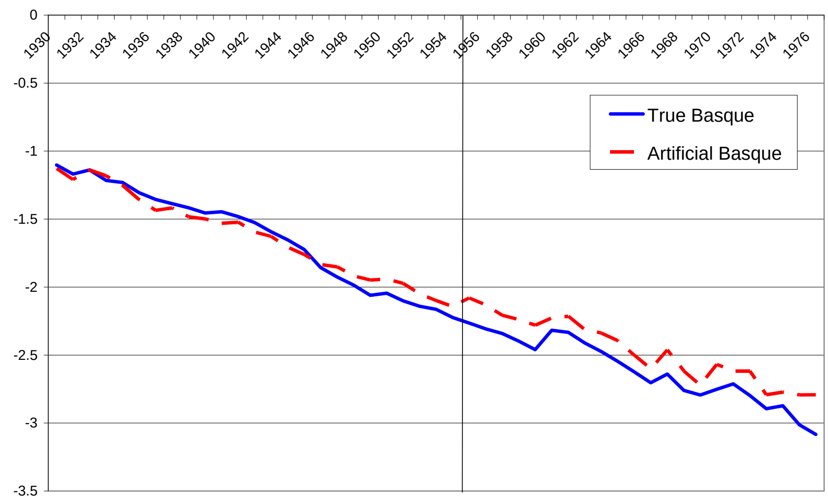
Of course, the question remains whether these results reflect reality well. This can be shown graphically, by showing the graphs that compare the true Basque region with the artificial one. Figures 1, 2 and 3 show the residual education values for the low, middle and high deciles of the true and artificial Basque Regions. For the lowest deciles, it can clearly be seen from the graph that there is a disparity between the true Basque region and the artificial region, but it also shows that this divergence starts before the conflict-generation is born. In fact, for those born after 1955, the disparity is not any larger than for those born before 1955. For the highest deciles (figure 3), there is also no division at all between the pre-conflict generation and the post-conflict generation. However, when looking at the middle levels of education in figure 2, a clear divergence takes place. As we hypothesised in section 3, the level of education for this group of individuals actually increases compared to the artificial Basque Region. This is compatible with what we have stated earlier, if indeed a larger portion of these individuals decides to seek opportunities outside the Basque Region. An quantification is possible, with the total divergence between the true and artificial Basque Regions being approximately 0.5 points

69 Of course, for the first decile, there are no lower-ranked deciles, so a specific value for the first deciles has to be 0. The contributions of Vastile & Leon, Madrid and Navarra take place in the estimation of the second and third deciles.

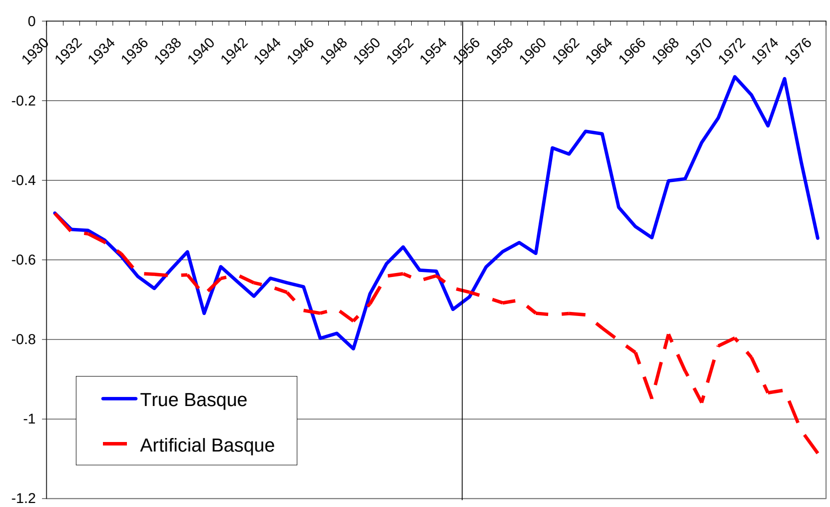
17.Weights distribution within artificial Basque Region

| | | Deciles | | | |
|---|---|---|---|---|---|
| | | *lowest 3* | *middle 4* | *highest 3* | *All* |
| | *Valencia* | 0 | 0.092 | 0 | 0.037 |
| | *Andalusia* | 0 | 0 | 0.071 | 0.021 |
| | *Asturias* | 0 | 0.099 | 0.058 | 0.057 |
| L | *Vastile&Leon* | 0.049 | 0 | 0 | 0.015 |
| O | *Cantabria* | 0 | 0.005 | 0.014 | 0.006 |
| W | *Galicia* | 0 | 0.126 | 0 | 0.051 |
| E | *La Rioja* | 0 | 0.007 | 0 | 0.003 |
| R | *Madrid* | 0.054 | 0.033 | 0 | 0.030 |
| | *Murcia* | 0 | 0.064 | 0.022 | 0.032 |
| | *Canary Islands* | 0 | 0.021 | 0 | 0.008 |
| | *Navarra* | 0.022 | 0.093 | 0 | 0.044 |
| | **TOTAL** | **0.126** | **0.539** | **0.0165** | **0.303** |
| | *Catalunya* | 0.018 | 0.139 | 0.008 | 0.063 |
| S | *Cantabria* | 0.009 | 0 | 0.075 | 0.025 |
| A | *La Rioja* | 0 | 0 | 0.007 | 0.002 |
| M | *Madrid* | 0.237 | 0 | 0.125 | 0.108 |
| E | *Navarra* | 0.278 | 0 | 0.199 | 0.143 |
| | **TOTAL** | **0.542** | **0.139** | **0.415** | **0.343** |
| | *Catalunya* | 0 | 0 | 0.013 | 0.004 |
| | *Castille-La Mancha* | 0 | 0.029 | 0 | 0.012 |
| | *Asturias* | 0 | 0.009 | 0.112 | 0.037 |
| H | *Vastile&Leon* | 0 | 0 | 0.135 | 0.040 |
| I | *Extremadura* | 0 | 0 | 0.033 | 0.010 |
| G | *Cantabria* | 0.055 | 0.014 | 0.067 | 0.042 |
| H | *Galicia* | 0 | 0.016 | 0 | 0.007 |
| E | *Aragon* | 0 | 0.049 | 0.060 | 0.038 |
| R | *La Rioja* | 0 | 0.130 | 0 | 0.052 |
| | *Madrid* | 0 | 0.005 | 0 | 0.002 |
| | *Murcia* | 0 | 0.021 | 0 | 0.009 |
| | *Navarra* | 0.277 | 0.048 | 0 | 0.102 |
| | **TOTAL** | **0.332** | **0.322** | **0.421** | **0.355** |

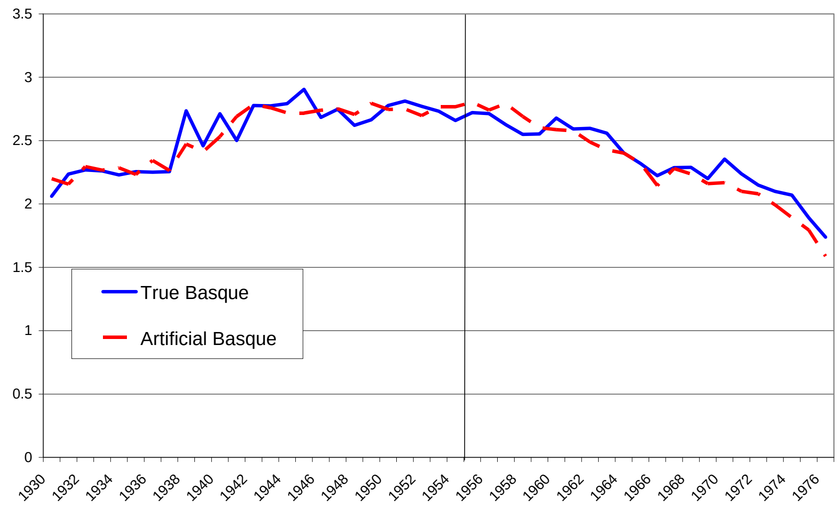
*Note: The table contains the weights of "same", "lower-ranked" and "higher ranked" deciles for all relevant regions in the construction of the different artificial decile groups. The table is explained further in the accompanying text.*




1.*This figure contains the values of residual education for the true Basque and the arti-cial Basque regions for the three lowest deciles.*




2.*This figure contains the values of residual education for the true Basque and the arti-cial Basque regions for the four middle deciles.*




3.*This figure contains the values of residual education for the true Basque and the artificial Basque regions for the three highest deciles.*

of residual education. In order to put this value in perspective, we have calculated the average difference between any two contiguous deciles, which turns out to be approximately 0.6. Therefore, the relative increase in education is nearly equal to an upward movement of one decile for the entire middle population, which is a surprisingly large result.

## Additional evidence

Our results in the previous section indicate that in the Basque Region, the middle deciles increased their levels of education disproportionally compared to individuals who did not reside in the conflict region. In section 3, we hypothesised that this is due to the increased incentives to migrate. Of course there may be other channels through which the changes in educational achievements change, but since we have no alternative theories available, it makes sense to test the migration theory. We do this by looking at the actual migration behaviour of individuals. Here we use a relatively straightforward technique and simply consider the diff-in-diff for within-country migration from the Basque Region and from all non-Basque regions. As is shown in the top part of table 18, both before and after the conflict, Basque-born individuals are less likely to migrate to other Spanish regions than individuals born elsewhere. This may be due to the cultural difference between the Basque Region and the other Regions, but it may also be due to the superior economic position held by the Basque Region. The same table, however, also shows that there is quite a large decrease in the probability of migration for non-Basque individuals after the cutoff point. The Basque-born individuals, on the other hand, actually become slightly more likely to migrate. This implies that, according to our difference-in-difference estimation, the Basque-born indeed increase their probability of within-country migration.

Another important implication from the theoretical model is a (relative) decrease in the level of education of migrants. After all, the education cutoff beyond which individuals decide to migrate moves to a lower part of the distribution and as a result, the average migrant's level of education should go down. Considering the difference-in-difference, the lowest section of table 18 shows that the level of education of migrants from the Basque Region indeed decreases, compared to migrants from other regions. However, in contrast, the middle section of the table shows that the overall education level of Basque-born individuals increases faster than that of non-Basque born, despite the fact that the Basque-born have a higher level of education to start with. Again, this is evidence that supports our theory that the occurrence of conflict increases the incentive to migrate and that the potentiality of migration requires a higher level of education.

Three things should, however, be considered in this case as well. The first is the geographical size of the regions. After all, their sizes differ immensely and it can be expected that fewer people from large regions migrate than people from geographically smaller ones. For two reasons, this cannot be driving the effect we find. First of all, the Basque Region is in fact among the smallest of the regions and one therefore should in principle expect a higher level of migration, which is not the case. Second, as the sizes of the regions do not change over time, this cannot be expected to influence the results from our difference-in-differences analysis.

The second thing that needs to be considered is foreign migration. As we use data from the Spanish Census, this merely includes data regarding those individuals who are still living in Spain. The percentage of Spaniards migrating internationally has been relatively small (compared to within-country migration) and would not be able to explain our results. However, even if there were a relatively large amount of international migration, this would have only two effects: First, it would increase the amount of migration, particularly from the Basque Region, as it is close to the international border with France. Second, with the increasing integration between Spain and the rest of Europe, including their joining of the European Union, one would particularly expect the later

18.Naive diff-in-diff estimation of effects of civil conflict on migration and education levels

| | *Spain* | *Basque* | Difference | |
|---|---|---|---|---|
| **%Migrants** | | | | **N=3,995,162** |
| *Pre-conflict* | *30.5* | *15.1* | -15.4 | |
| *Post-conflict* | *25.2* | *15.4* | -9.8 | |
| Difference | -5.3 | 0.3 | **5.6*** ** | (0.227) |
| **Educ_all** | | | | **N=3,986,468** |
| *Pre-conflict* | *3.09* | *3.85* | 0.77 | |
| *Post-conflict* | *4.79* | *5.69* | 0.90 | |
| Difference | 1.70 | 1.83 | **0.13*** ** | (0.0096) |
| **Educ_migr** | | | | **N=1,093,197** |
| *Pre-conflict* | *3.33* | *4.42* | 1.09 | |
| *Post-conflict* | *4.82* | *5.76* | 0.94 | |
| Difference | 1.49 | 1.34 | **-0.15*** ** | (0.026) |

*Note: This table shows a naive difference-in-difference comparison between the Basque region and the rest of Spain, in which it can be seen that migration out of the Basque Region has increased, that the average education level has increased and that the average migrant's education level has decreased. Between brackets, the standard deviations of the diff-in-diff estimations are included.*

period to have increased total migration, because of the international component. This would imply that the result we find now is in fact an underestimation of migration effect rather than an overestimation.

Finally, the third thing that further strengthens our results is government policy. For political reasons, successive Spanish governments have been trying to give incentives to individuals to move towards the Basque region. This would imply that overall, within-country migration has been stimulated more in the non-Basque regions than in the Basque Region itself, which again biases our results in the direction of an underestimation, rather than an overestimation.

## Conclusion

In this paper, we analyse the potential effect that conflict has on the incentives to acquire education. According to our naive theoretical model, the demand-side effects of civil conflict particularly influence the middle level of education, although this depends on the parameters of the model. We test our theoretical model in an analysis of the Basque conflict region in Spain. After all, this is exactly the kind of conflict in which our theoretical model is applicable, because it is a conflict that is relatively low-key and influences only the demand for education and not the supply thereof. After all, at no time during the conflict, was there any significant interruption of the supply of education, which makes it an interesting case to analyse. When using a relatively advanced matching technique to set up an artificial region that has an educational distribution similar to that of the Basque Region, it turns out that the effects are indeed particularly visible among the median levels of education. These individuals, it turns out, acquire a significantly higher level of education after the initiation of the conflict than beforehand, which leads us to believe that there exists in fact a causal relationship. The probability of migration of this group also increases over time, which is in concordance with our results.

The results found in this paper contrast previous findings in which conflict has been shown to have a negative impact on human capital accumulation. We argue that this negative relationship is indeed true, but fully due to supply-side effects, whereas demand-side effects actually increase human capital accumulation. As a result, the supply-side effect found in previous papers is probably an underestimation of the true effect.

While it is a little difficult to see usable policy recommendations on basis of our results, we do believe there is scope for some. One of the major differences between the conflict in the Basque Region and many other civil conflicts is its geopolitically limited scope. The conflict is limited to the Basque Region only, and the rest of the same political entity (i.e. Spain) is unaffected. Apart from potential cultural-linguistic barriers, there are no barriers to migration between the Basque Region and the rest of Spain, which makes sure that, even when individuals expect only a modest utility improvement from migrating, they still increase their demand for education. In many other conflicts, momentarily ignoring the supply-side effects, civil conflict in one country may in fact increase the difficulty of migration to other places, when a conflict engulfs an entire nation. And this is where we can suggest one potential policy recommendation, because it is clear that when individuals have the option of migration in the future, like in the rest of the brain gain literature, this improves their incentives for education. It is therefore not wise to lock refugees/migrants into positions where there is little scope for improvement, because in the long run, this strongly reduces their incentives for education, which exacerbates the long-run negative impact of conflict.

# References


[1] Abadie, Alberto and Javier Gardeazabal (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country", *American Economic Review*, Vol. 93 (1), pp. 113-132

[2] Abreu, Maria, Henri L.F. de Groot and Raymond J.G.M. Florax (2004), "Space and Growth", *Tinbergen Institute Discussion Paper*, TI 2004-129/3

[3] Ades, Alberto and Hak B. Chua (1997), "Thy Neighbour's Curse: Regional Instability and Economic Growth", *Journal of Economic Growth*, vol. 2, pp. 279-304

[4] Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003), "Fractionalization", *Journal of Economic Growth*, Vol. 8, pp. 155-194

[5] Alesina, Alberto, William Easterly and Janina Matuszeski (2006), *Artificial States*, NBER Working Paper 12328

[6] Annett, Anthony (2001), "Social Fractionalization, Political Instability, and the Size of Government", *IMF Staff Papers*, Vol. 48 (3), pp. 561-592

[7] Anselin, Luc (1988), *Spatial Econometrics*, Kluwer Academic Publishers, Dordrecht, the Netherlands

[8] Arrazola, María and José De Hevia (2006), "Gender Differentials in Returns to Education in Spain", *Education Economics*, Vol. 14(4), pp. 469-486

[9] Barro, Robert J. (2001), *Education and Economic Growth*, Working Paper

[10] Barro, Robert J. and Sala-i-Martin, Xavier (1992), "Convergence", *Journal of Political Economy*, vol. 100(2), pp. 223-251

[11] Barro, Robert J. and Jong-Wha Lee (2000), "International Data on Educational Attainment: Updates and Implications", CID Working Paper, No. 42

[12] Baker, Dean (2007), "The Economic Impact of the Iraq War and Higher Military Spending", CEPR Research Report

[13] Barrett, David B., George Thomas Kurian and Todd M. Johnson (eds.) (2001), *World Christian Encyclopedia: A Comparative Survey of Churches and Religions in the Modern World*, Oxford University Press, 2nd edition

[14] Beck, Nathaniel, Kristian Skrede Gleditsch and Kyle Beardsley (2006), "Space Is More Than Geography: Using Spatial Econometrics in the Study of Political Economy", *International Studies Quarterly*, Vol. 50, pp. 27-44

[15] Beine, M., F. Docquier and H. Rapoport (2001), "Brain Drain and Economic Growth: Theory and Evidence", *Journal of Development Economics*, Vol. 64, pp. 275-289

[16] Bhagwati, J.N. and K. Hamada (1974), "The Brain Drain, International Integration of Markets for Professionals and Unemployment", *Journal of Development Economics*, Vol. 1(1), pp.19-42

[17] Borjas, George J. (1994), "The Economics of Immigration", *Journal of Economic Literature*, Vol. 32 (4), pp. 1667-1717

[18] Boucher, S., O. Stark and J.E. Taylor (2005), *A Gain with a Drain? Evidence from Rural Mexico on the New Economics of the Brain Drain*, Invited paper 14th World Congress of the International Economic Association, Marrakech, Morocco, August 29 – September

2, 2005

[19] Bossert, Walter, Conchita D'Ambrosio and Eliana La Ferrara (2008), *A Generalized Index of Fractionalization*, Bocconi University Working Paper

[20] Buhaug, Halvard and Kristian Skrede Gleditsch (2008), "Contagion or Confusion? Why Conflicts Cluster in Space", *International Studies Quarterly*, Vol. 52(2), pp. 215-233

[21] Buhaug, Halvard and Päivi Lujala (2005), "Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War", *Political Geography,* vol. 24 (4), pp. 399-418

[22] Cavalli-Sforza, L. Luca, Paolo Menozzi and Alberto Piazza (1994), *The History and Geography of Human Genes*, Princeton University Press

[23] Center for Global Policy (2008), "Polity IV Project", <http://www.cidcm.umd.edu/polity>, retrieved January 2008

[24] CIA (2007), "The World Factbook", <https://www.cia.gov/cia/publications/factbook/>, retrieved December 2006 and November 2007

[25] Collier, Paul (1999), "On the Economic Consequences of Civil War", *Oxford Economic Papers*, vol. 51, pp. 168-183

[26] Collier, Paul and Anke Hoeffler (2004), "Greed and Grievance in Civil War", *Oxford Economic Papers*, Vol. 56 (4), pp. 563-595

[27] Djankov, Simeon and Marta Reynal-Querol (2007), *The Causes of Civil War*, World Bank Policy Research Working Paper, No 4254

[28] Easterly, William and Ross Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions", *Quarterly Journal of Economics*, Vol. 112 (4), pp. 1203-1250

[29] Esteban, Joan and Debraj Ray (1994), "On the Measurement of Polarization", *Econometrica*, Vol 62 (4), pp. 819-851

[30] Fearon, James D. (2003), "Ethnic and Cultural Diversity by Country", *Journal of Economics Growth*, Vol. 8, pp. 196-222

[31] Fearon, James D. (2005), "Primary Commodity Exports and Civil War", *Journal of Conflict Resolution*, Vol. 49 (4), pp. 483-507

[32] Fearon, James D. and David D. Laitin, (2003), "Ethnicity, Insurgency and Civil War", *American Political Science Review*, Vol. 97(1), pp. 75-90

[33] García Prieto, Carmen, Angel Martín Román and Carlos Pérez Domínguez (2004), "Actual and Potential Returns to Schooling", *Education Economics*, Vol. 13(4), pp. 387-407

[34] Gerrard, A.J.W. (2000), *What is a Mountain? Background Paper to definition of mountains and mountain regions*, World Bank mimeo

[35] Ghobarah, Hazem Adam, Paul Huth and Bruce Russett (2004), "The Post-war Health Effects of Civil Conflict", *Social Science & Medicine*, Vol. 59, pp. 869-884

[36] Gleditsch, Kristian Skrede (2007), "Transnational Dimensions of Civil War", *Journal of Peace Research*, Vol. 44 (3), pp. 293-309

[37] Gleditsch, Kristian S. and Michael D. Ward (2001), "Measuring Space: A Minimum-Distance Database and Applications to International Studies", *Journal of Peace Research*, vol. 38(6), pp. 739-758

[38] Gleditsch, Nils Petter, Peter Wallensteen, Mickael Eriksson, Margareta Sollenberg and

Håvard Strand (2002), “Armed Conflict 1946-2001: A New Dataset”, *Journal of Peace Research*, vol. 39(5), pp. 615-637

[39] Goerlich Gisbert, Francisco J. and Matilde Mas Ivars (2001), *La Evolución económica de las proviñcias Espanolas (1955-1998) Volumen II, Desigualdad y convergenci*a, Fundación BBVA, Bilbao

[40] Göksel, Idil (2007), *Determinants of School Attainment in Turkey and the Impact of Extention of Compulsory Education*, Working Paper

[41] Gordon, Raymond G., Jr. (ed.) (2005), *Ethnologue: Languages of the World, Fifteenth edition*, Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/

[42] Gradstein, Mark (2003), *The Political Economy of Public Spending on Education, Inequality, and Growth*, World Bank Research Working Paper, No 3162

[43] Greenberg, Joseph H. (1956), “The Measurement of Linguistic Diversity”, *Language*, Vol. 32 (1), pp. 109-115

[44] Guidolin, Massimo and Eliana La Ferrara (2007), “Diamonds are Forever, Wars are Not: Is Conflict Bad for Private Firms?”, *American Economic Review*, Vol. 97 (5), pp. 1978-1993

[45] Guiso, Luigi, Paolo Sapienza and Luigi Zingales (2007), *Cultural Biases in Economic Exchange?*, EUI Working Papers, ECO 2007/42

[46] Haque, N.U. and S.J. Kim (1995), “Human Capital Flight: Impact of Migration on Income and Growth”, *IMF Staff Papers*, Vol. 42 (3), pp. 577-607

[47] Harbom, Lotta and Stina Högbladh (2006), *UCDP/PRIO Armed Conflict Dataset Codebook*, Uppsala Conflict Data Program (UCDP) and International Peace Research Institute, Oslo (PRIO)

[48] Heston, Alan, Robert Summers and Bettina Aten (2006), *Penn World Table Version 6.2*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania

[49] Hoeffler, Anke and Marta Reynal-Querol (2003), *Measuring the Costs of Conflict*, Working Paper

[50] Ichino, Andrea and Rudolf Winter-Ebmer (2004), “The Long-Run Educational Cost of World War II”, *Journal of Labour Economics*, Vol. 22(1), pp. 57-86

[51] Instituto Nacional de Estadistica (1991), *Censos de Población y Viviendas 1991*, <http://www.ine.es>, retrieved March 2008

[52] Instituto Nacional de Estadistica (2001), *Censos de Población y Viviendas 2001*, <http://www.ine.es>, retrieved March 2008

[53] Instituto Nacional de Estadistica (2001), *Historical Population Figures*, <http://www.ine.es>, retrieved March 2008

[54] Johnstone, Patrick (2007), “Affinity Blocks and People Clusters: An Approach Toward Strategic Insight and Mission Partnership”, *Mission Frontiers*, March-April, pp. 8-15

[55] Joshua Project (2007), “Joshua Project: Bringing Definition to the Unfinished Task”, <www.joshuaproject.net>, retrieved November 2007

[56] Koubi, Vally (2005), “War and Economic Performance”, *Journal of Peace Research*, vol. 42(1), pp. 67-82

[57] Lai, Brian and Clayton Thyne (2007), "The Effect of Civil War on Education 1980-97", *Journal of Peace Research*, Vol. 44(3), pp. 277-292.

[58] Laitin, David D. (2000), "What Is a Language Community?", *American Journal of Political Science*, Vol. 44 (1), pp. 142-155

[59] Longo, Roberto and Khalid Sekkat (2001), "Obstacles to Expanding Intra-African Trade", OECD Development Centre Working Paper No. 169

[60] Mankiw, N. Gregory, David Romer and David N. Weil (1992), "A Contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics*, vol. 107, pp. 31-77

[61] Mauro, Paolo (1995), "Corruption and Growth", *The Quarterly Journal of Economics*, Vol. 110 (3), pp. 681-712

[62] Michalopoulos, Stelios (2007), *Ethnolinguistic Diversity: Origins and Implications*, working paper, Brown University

[63] Miguel, Edward, Shanker Satyanath and Ernest Sergenti (2004), "Economic Shocks and Civil Conflict: An Instrumental Variables Approach", *Journal of Political Economy*, Vol. 112 (4), pp. 725-753

[64] Montalvo, José G. and Marta Reynal-Querol (2005), "Ethnic Polarization, Potential Conflict, and Civil Wars", *American Economic Review*, Vol. 95 (3), pp. 796-816

[65] Mountford, A. (1997), "Can a Brain Drain be Good for growth in the Source Economy?", *Journal of Development Economics*, Vol. 53, pp. 287-303

[66] Murdoch, James C. and Todd Sandler (2002a), "Economic Growth, Civil Wars and Spatial Spillovers", *Journal of Conflict Resolution*, vol. 46(1), pp. 91-110

[67] Murdoch, James C. and Todd Sandler (2002b), "Civil Wars and Economic Growth: A Regional Comparison", *Defence and Peace Economics*, vol. 13(6), pp. 451-464

[68] Murdoch, James C. and Todd Sandler (2004), "Civil Wars and Economic Growth: Spatial Dispersion", *American Journal of Political Science*, vol. 48(1), pp. 138-151

[69] Organski, A.F.K. and Jacek Kugler (1980), *The War Ledger*, University of Chicago Press

[70] Posner, Daniel N. (2004), "Measuring Ethnic Fracionalization in Africa", *American Journal of Political Science*, Vol. 48 (4), pp. 849-863

[71] Psacharopoulos, George and Harry Anthony Patrinos (2004), "Returns to Investment in Education: A Further Update", *Education Economics*, Vol. 12(2), pp. 111-134

[72] Roeder, Philip G. (2001), "Ethnolinguistic Fractionalization (ELF) Indices, 1961 and 1985." February 16, <http//:weber.ucsd.edu/~proeder/elf.htm>, retrieved January 3, 2008

[73] Rosetta Project (2007), "The Rosetta Project Digital Language Archive", <http://www.rosettaproject.org>, retrieved December 2007

[74] Sambanis, Nicholas (2002), "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War", *Defence and Peace Economics*, vol. 13(3), pp. 215-243

[75] Schultz, T.W. (1995), "The Economics of Women's Schooling" in *The Politics of Women's Schooling: Perspectives from Asia, Africa and Latin America*, by J.K. Conway and S.C. Bourque (eds.), University of Michigan Press, Ann Arbor, MI

[76] Solow, Robert M. (1956), "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics*, vol. 25, pp. 65-94

[77] Spolaore, Enrico and Romain Wacziarg (2006), *The Diffusion of Development*, NBER Working Paper Series 12153

[78] Stark, Oded (2004), “The Economics of the Brain Drain Turned on its Head”, Proceedings of the 5th World Bank Conference on Development Economics-Europe, pp. 335-345

[79] Stark, O., C. Helmenstein and A. Prskawetz (1997), “A Brain Gain with a Brain Drain”, *Economic Letters*, Vol. 55, pp. 227-234

[80] United Nations (2007) *UNStats*, retrieved March 2007, from http://unstats.un.org

[81] Ward, Michael D. and Kristian Skrede Gleditsch (2002), “Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace”, *Political Analysis*, Vol. 10 (3), pp. 244-260

[82] Wong, K. and C.K. Yip, (1999), “Education, Economic Growth, and Brain Drain”, *Journal of Economic Dynamics and Control*, Vol. 23, pp. 699-726

[83] Worldbank (2007), *World Development Indicators*, retrieved February 2007, from http://go.worldbank.org/B53SONGPA0