Linguistic Cleavages and Economic Development*

Klaus Desmet	Ignacio Ortuño-Ortín	Romain Wacziarg
Southern Methodist University and CEPR	Universidad Carlos III de Madrid	UCLA and NBER

January 2015

Abstract

In this chapter we argue that different types of linguistic cleavages matter for different economic outcomes. Whereas deep cleavages, going back far in history, are important for determining a society's degree of redistribution, shallow cleavages are enough to hamper economic growth and development. This suggests that the degree of redistribution is determined by deeply rooted differences between linguistic groups. In contrast, economic growth may have more to do with the ease of coordination, collaboration and communication. There, more superficial linguistic differences suffice to have a negative impact on growth. The same finding holds when exploring the determinants of per capita income: even shallow linguistic differences are sufficient to hamper a country's level of development.

*Desmet: Southern Methodist University, Dallas, TX, kdesmet@smu.edu, Ortuño-Ortín: Universidad Carlos III, Getafe (Madrid), Spain, iortuno@eco.uc3m.es and Wacziarg: UCLA Anderson School of Management, Los Angeles, CA, wacziarg@ucla.edu.

This paper was prepared for the *Palgrave Handbook of Economics and Language*, Victor Ginsburgh and Shlomo Weber, eds. All errors are ours.

1 Introduction

What is the effect of linguistic diversity on economic and political outcomes? Much of the recent literature on this topic investigates how linguistic cleavages affect civil conflict, redistribution, economic growth, public goods and governance.[1] Most of the cross-country evidence suggests that linguistic diversity has negative effects on these political economy outcomes. These findings may help explain why the U.S. has a smaller welfare state than Europe, why some countries develop more slowly than others or why some African countries tend to have a higher incidence of civil conflict than others.

This chapter focuses on two important questions in this literature. The first question has to do with measurement, and in particular with defining the relevant linguistic groups used to measure linguistic fractionalization. For example, should we consider Flemish and Dutch to be two distinct groups? We will argue that the answer depends on the particular political economy outcome we are interested in: different linguistic cleavages matter for different outcomes. A second question has to do with the relationship between linguistic diversity and the level of development. In contrast to other political economy outcomes such as economic growth, less attention has been paid to the level of GDP and its relationship with linguistic fractionalization.

Diversity is usually measured by a fractionalization index that takes into account the number and the sizes of the different groups. One common criticism of this approach is that in many cases it is difficult to determine which dimension - language, ethnicity, religion, culture - defines the relevant groups (Laitin and Posner, 2001). Here we ask a related question, focusing exclusively on linguistic heterogeneity. Even when focusing only on language as the main dimension of heterogeneity, we are faced with the question of what constitutes the relevant linguistic classification. Almost everyone would consider Lombard and Piemontese to be variants of Italian, rather than two distinct languages. In contrast, most would consider Hindi and German to be distinct language groups, despite both belonging to the Indo-European family. But of course there are many in-between situations where doubts may arise: are Galician and Spanish or Icelandic and Norwegian sufficiently different to classify them as distinct groups?

In trying to determine the relevant groups to construct measures of linguistic diversity, in Desmet, Ortuño-Ortín and Wacziarg (2012) we argued that different cleavages may matter for different political economy outcomes. To make our point, we used a phylogenetic approach, based on information from language trees, to compute diversity measures at different levels of aggregation. At the highest level of aggregation, only the world's main language families, such as Indo-European and Nilo-Saharan, would define different groups, whereas at the lowest level of aggregation, even the different variants of Italian, such as Lombard and Venetian, would define different linguistic groups.

We used measures of linguistic diversity at different levels of aggregation to study the determinants of redistribution, conflict and growth. We found that for redistribution and conflict, diversity measures at high levels of aggregation matter most, whereas for economic growth, diversity measures at low levels of aggregation are more significant determinants. To interpret these results, we observed that linguistic trees give a historical dimension to the analysis. For instance it is estimated that the split between Indo-European languages and non-Indo-European languages happened about $8,700$ years ago. In contrast, the split between Icelandic and Norwegian occurred only after the 12^{th} century (Gray and Atkinson, 2003). Hence, these findings indicate that for redistribution, coarse divisions, going back far in time, matter most. Solidarity and empathy may not overcome deep cleavages, but can more easily bridge shallow divisions. In contrast, fine divisions are enough to hinder a country's economic growth, an outcome for which coordination and communication between economic agents matters for the economy to operate efficiently.

In this chapter, we build on our earlier work, extending our results to an analysis of how linguistic diversity affects the level of development. The recent literature in macro development has paid increasing attention to levels rather than growth, starting with Hall and Jones (1999) and Acemoglu, Johnson and Robinson (2001). Yet the effect of ethnolinguistic diversity on levels of development has not been the subject of a lot of research. If our interpretation is correct, we should expect shallow cleavages to also suffice to negatively impact a country's level of development. As noted by Parente and Prescott (1994), growth differences in income per capita across countries tend to be transitory, whereas level differences are not. Thus, the effect of linguistic diversity on growth could differ from its effect on income per capita levels. We find, in fact, that it does not. For per capita income levels, as for growth, heterogeneity measures based on finer linguistic distinctions matter more than those based on coarse ones. This finding constitutes a confirmation of our earlier interpretation, where coarse linguistic divisions created conflict and a lack of redistribution. In contrast, finer ones were sufficient to generate adverse effects on outcomes such as growth that require coordination and communication between heterogeneous groups.

The rest of the chapter is organized as follows. Section 2 explains the phylogenetic approach of using language trees to compute measures of diversity at different levels of aggregation. Section 3 illustrates the usefulness of this phylogenetic approach by briefly revisiting the main findings in Desmet, Ortuño-Ortín and Wacziarg (2012), comparing the impact of linguistic diversity on redistribution and growth. Section 4 analyzes the relationship between linguistic diversity and the level of development, and situates the new findings in the broader literature. Section 5 concludes by summarizing our economic interpretation of the empirical findings.

2 A Phylogenetic Approach to Linguistic Diversity

In this section we explain how to use language trees to compute measures of linguistic diversity, based on either coarse or fine divisions between languages. We then compute these different measures for 226 countries, and show that a country's measured linguistic diversity depends crucially on whether we take into account fine divisions between languages or not.

2.1 Linguistic Trees

Linguistic trees show the genealogical relationships between languages.[2] Linguistic differentiation occurs because populations become separated from each other. For example, the fall of the Roman Empire with the subsequent segmentation of populations and linguistic drift divided Latin into the different Romance languages that we know today. The degree of relatedness between languages in linguistic trees therefore gives a rough measure of the time that has elapsed since the two languages became separated. For example, Gray and Atkinson (2003) estimate that for the Indo-European language group, the split between the languages that would later give rise to present-day Hindi and German occurred about $6,900$ years ago, whereas the split between what would become Swedish and German goes back only $1,750$ years. Correspondingly, Hindi and German are separated by more branches in linguistic trees than Swedish and German.

Although this does not imply that linguistic trees act as precise clocks that measure the separation times of populations, as genetic distance does, deeper linguistic cleavages do correspond to greater linguistic differences between populations. In fact, Cavalli-Sforza et al. (1988) argues that there is a relationship between the world's main language groups and the world's most important genetic clusters.[3] This is consistent with several studies on Europe that have shown a significant correlation between genetic and linguistic diversity (Sokal, 1988). In a more recent, broader study, covering 50 populations across all continents, Belle and Barbujani (2007) reach a related conclusion. They find that language differences have a detectable effect on DNA diversity, above and beyond the effects of geographic distance. Like genes, language is passed on from generation to generation.

Since linguistic trees capture the degree of relatedness between languages, they can be used to compute different measures of diversity. Some of these measures can be based on coarse divisions, going back far in time, while others also include more shallow, recent divisions between languages.

Before calculating these different indices, recall that the standard ELF measure of fractionalization captures the probability that two individuals chosen at random belong to different groups. Formally, in a country with N groups, indexed by i, the ELF index is:

$$ELF = 1 - \sum_{i=1}^{N} [s_i]^2 \tag{1}$$

where s_i is the population share of group i.

In much of the literature the different groups i are taken as exogenously given. Instead, here we exploit the genealogical relationships between languages to define groups at different levels of coarseness. This is illustrated in Figure 1, showing the genealogical relationships between the main languages spoken in Pakistan. At the most disaggregated level, each of those seven languages (Panjabi, Pashto, Sindhi, Seraiki, Urdu, Balochi and Brahui) are taken to be a different group. Using the population shares that appear below the language names, this gives us an ELF index of 0.722. That is, the probability that two randomly chosen Pakistani individuals speak different languages is 72.2%. Because there are seven levels of aggregation in this language tree, we denote this measure of fractionalization as ELF(7).

[Insert Figure 1 here]

As we go up the language tree, some languages become part of the same group. For example, when going up two levels, Panjabi, Seraiki and Sindhi all belong to the same group. Together, they now account for a 0.714 share of the population. At that level of aggregation, the other four languages continue to constitute different groups. The corresponding ELF index, which we refer to as ELF(4), is now 0.460. That is, at aggregation level 4, the probability that two randomly chosen Pakistanis belong to a different group is only 46.0 percent. Of course, by construction, the ELF index decreases with the level of aggregation. At level 1, only two broad language families survive, Indo-European, accounting for 98.5 percent of the population, and Dravidian, accounting for 1.5 percent. Correspondingly, ELF(1) drops to 0.030, and by this account Pakistan no longer appears to be very linguistically diverse: when randomly choosing two Pakistanis, the probability that one speaks an Indo-European language and the other a Dravidian language is only 3 percent. As already mentioned, diversity at higher levels of aggregation capture deeper cleavages than diversity at lower levels of aggregation.

One issue when computing these different ELF indices is that in general not all languages are equidistant from the root. This can easily be seen in Figure 1. Although we have drawn all languages to be at the same distance from Proto-Human, in reality not all seven languages are removed by the same number of branches from the origin. While Urdu is seven branches away from the origin, Sindhi is six branches away, and Brahui is only three branches from the origin. To get around this issue, we move all languages down to the lowest level, thus making them equidistant from the origin. To be more precise, we are implicitly assuming that between Sindhi and the node called "Northwestern zone" there are two intermediate languages, one at level 5 and another at level 6, that capture the evolution of "Northwestern zone" into what today is Sindhi. The interested reader is referred to Desmet, Ortuño-Ortín and Wacziarg (2012) for a more detailed discussion of different ways of completing a tree to ensure that all languages are equidistant from the origin. These different methods do not yield vastly different empirical results or indices.

2.2 Fractionalization at Different Levels of Aggregation

Using data on the speakers of the $6,912$ world languages in the Ethnologue, together with information on linguistic trees, we can compute for each country different ELF measures at different levels of aggregation. The linguistic tree in the Ethnologue has a maximum of 15 levels.[4] By positioning all present-day spoken languages at the same distance from the origin, we can compute for each country 15 ELF measures, one for each level of disaggregation. More formally, for every level of disaggregation j, denote the partition of the country into $N(j)$ groups with population shares $s_{i(j)}$, where $i(j) = 1, 2, \ldots, N(j)$. We can then define a fractionalization index for any level of disaggregation j by

$$ELF(j) = 1 - \sum_{i(j)=1}^{N(j)} \left[s_{i(j)}\right]^2 . \tag{2}$$

[Insert Figure 2 here]

A country's relative level of diversity depends dramatically on the level of aggregation. To get a sense of how different things may look, Figure 2 shows maps of ELF(2) and ELF(15).[5] When computing ELF(2), French and German are allocated to different groups, but Spanish and French are not, whereas when computing ELF(15) all of the $6,912$ languages recorded in the Ethnologue are allocated to different groups, even if they are very similar. The differences are striking. Many countries in central and southern Africa have very high levels of diversity at Level 15, but relatively low levels of diversity at Level 2. Mozambique is a good example. According to Ethnologue the country has 43 languages, which explain why it ranks 10$^{\text{th}}$ out of 226 using ELF(15). However, 99.8 percent of Mozambicans speak a language of the Niger-Congo group, explaining why the country drops to the 200$^{\text{th}}$ position when using ELF(2). As a result, in Mozambique ELF(2) is 0.929 whereas ELF(2) is 0.004. Hence, depending on whether we consider deep cleavages or shallow cleavages, we would view Mozambique to be either a very diverse or a very homogeneous country.

In contrast, many countries in the Sahel region are highly diverse, independently of whether we look at ELF(2) or ELF(15). Chad, for example, ranks 6th when measuring diversity at Level 15, and is the most diverse country in our sample when measuring diversity at Level 2. In that country ELF(15) is 0.950 and ELF(2) is 0.805. This is the case because in Chad about a third of the population speaks an Afro-Asiatic language, about half a Nilo-Saharan language and the rest a language of the Niger-Congo family. Many of the Latin American countries, such as Bolivia, Ecuador or Peru, also have relatively similar levels of diversity, independently of whether we measure diversity at Level 2 or Level 15. Most of the diversity in those countries derives from the division between Spanish and non-Spanish speakers, where most of the non-Spanish speakers do not pertain to the Indo-European language family.

[Insert Table 1 here]

Table 1 provides further information about the different ELF measures. Panel A reports the summary statistics. As expected, the degree of diversity increases with the level of disaggregation. Panel B reports the correlations between the different measures. The correlation between ELF(1) and ELF(15) is only 0.526, indicating that these two measures are actually quite different. Of course, the correlations become much larger when we compare higher degrees of disaggregation. For example, the correlation between ELF(9) and ELF(15) is 0.943. This high correlation reflects the fact that the vast majority of languages are less than ten branches from the origin. As a result, in nearly three quarters of the countries ELF(9) and ELF(15) are identical. In only a handful of countries, mostly located in southern Africa, are the two measures substantially different. These countries include Gabon, South Africa, Zimbabwe, Uganda and Mozambique. For this reason it is usually sufficient to focus on a subset of the 15 measures of linguistic heterogeneity, as we sometimes do in the empirical work below.

3 Linguistic Diversity, Redistribution and Economic Growth

In this subsection we summarize the most important insights of Desmet, Ortuño-Ortín and Wacziarg (2012), where we let the data inform us which level is more relevant for the issue at hand. There are two reasons for this approach. First, it is not obvious which criterion one would use to choose the “right” level of aggregation, so that any attempt would likely be somewhat arbitrary. In fact, the arbitrariness of linguistic classifications characterizes common practice in the literature. This is the problem we are trying to address. Second, and more important, depending on the issue at hand, a different level of aggregation may be more or less relevant. By discovering which diversity measure has more predictive power, we can learn something economically meaningful. For example, if we were to find that fractionalization based on deep cleavages is what matters for redistribution, then we would conclude that solidarity and empathy have to do with deep fault lines in society that go back far in time and are deeply engrained. If, instead, we were to find that even shallow divisions reduce people's willingness to redistribute, then our interpretation would be quite different.

The main finding is that the relevant linguistic cleavages vary dramatically across different political economy outcomes. In the case of civil conflict and redistribution, deep divisions seem to be more important, whereas in the case of growth even shallow divisions are enough to hamper economic performance. These results are obtained by regressing the outcome of interest on linguistic fractionalization at successively greater levels of linguistic disaggregation, and a series of control variables often used for each dependent variable in the existing literature. The standardized beta on linguistic fractionalization is our summary measure of the magnitude of its effect on the outcome under scrutiny. Figure 3 compares the standardized betas on fractionalization at different levels of aggregation for redistribution (Panel A) and economic growth (Panel B).

[Insert Figure 3 here]

The figure in Panel A is based on an OLS regression of transfers and subsidies as a share of GDP on fractionalization, with a number of standard controls.[6] The regression is run 15 times, once for each level of aggregation, and Panel A then displays the effect of a one standard deviation increase in fractionalization on redistribution (expressed as percentage of a one standard deviation in redistribution). As can be seen, the effect of ELF(1) is -9.6%, and statistically significant at the 5% level. Once we pass the ELF(5) bar, fractionalization no longer has a statistically significant effect on redistribution. Hence, social solidarity travels well across shallow cleavages, but ceases to do so when divisions are deep.

The results for growth are very different. The figure in Panel B is based on an OLS regression of growth in GDP per capita for the period 1970-2004 on fractionalization, with a number of standard controls.[7] Again, the regression is run 15 times, once for each level of aggregation. As shown in Panel B, the effect of fractionalization becomes more negative and statistically more significant at lower levels of aggregation. The standardized beta - the effect of a one standard deviation increase in fractionalization as a share of the standard deviation of growth - reaches a maximum of -24% at ELF(9), and after that more or less stabilizes. This suggests that shallow divisions are enough to hinder economic growth. This does not imply that deep cleavages are unimportant. However, if we focus exclusively on deep cleavages, we miss the shallow divisions, which also matter.

We argue that civil war and redistribution are more driven by differences in "preferences" (disagreements over policy or political control), whereas economic growth has more to do with the efficiency of "technology" (inability to coordinate and communicate). Our results indicate that when it comes to issues involving conflicts between groups, as in the case of war or redistribution, the deeper linguistic fault lines matter most. In contrast, when it comes to economic growth, the efficiency of an economy depends on the ease of trade, communication, coordination and collaboration. Shallow linguistic differences between groups are enough to have a negative impact on economic growth.[8]

4 Linguistic Diversity and Economic Development

In this section we explore which level of aggregation is more important for a country's *level* of development. This is of interest for several reasons. First, the relation between linguistic diversity and the level of economic development has been somewhat understudied. Much of the literature on linguistic diversity focuses on civil conflict, redistribution, economic growth, public goods and governance, with less attention being paid to the level of development. Notable exceptions are Fishman (1967), Pool (1972), and more recently, Nettle (2000) and Nettle et al. (2007).[9]

In this rather limited literature, there is a lack of consensus on the relation between linguistic diversity and GDP per capita. On the one hand, Pool (1972, p. 222) takes a negative view and goes as far as stating that "a country that is linguistically highly heterogeneous is always undeveloped or semideveloped, and a country that is developed always has considerable language uniformity". Pool's conclusions are based on the simple correlation between linguistic diversity and GDP per capita in a cross-section of countries, a notable weakness. However, other studies which do control for confounding variables, such as Nettle (2000), find a similar result.[10] On the other hand, Fishman (1991) takes a more positive (or neutral) view and claims that, when controlling for enough other explanatory variables, linguistic heterogeneity ceases to affect the level of economic development. Laitin and Ramachandran (2014) reach a similar conclusion: once they account for linguistic distance from the official language, diversity no longer influences GDP per capita. The lack of agreement in this literature is one of our motivations for revisiting the relation between linguistic diversity and the degree of development using our phylogenetic approach.

A second reason for our interest is that, as argued by Parente and Prescott (1994), long-run growth rates tend to converge across countries, but differences in the *level* of development are often quite persistent. Hence, to understand long-run relative differences across countries, it is more reasonable to look at levels, rather than growth rates. Of course, much of the empirical growth literature takes this into account by focusing on conditional convergence regressions. By controlling for initial GDP per capita, the other regressors can be interpreted as determinants of the steady-state differences in the levels of development. Here, instead, we look directly at the level of development. This has the additional advantage of getting around the issue of growth rates often being quite transitory, a problem pointed out by Easterly et al. (1993) and Hall and Jones (1999).

A third reason for investigating the effect of linguistic diversity on income levels is that if our earlier interpretation for the case of growth is correct, we would expect shallow divisions to hamper economic development as much as deep divisions. In that sense, we can interpret our analysis of economic development as constituting an additional test of our earlier interpretation of the effect of linguistic heterogeneity on growth.

To analyze the relation between fractionalization at different levels of aggregation and a country's level of development, we use the following standard econometric specification:

$$y = \delta D(j) + X\beta + \varepsilon \tag{3}$$

where y is income per capita in the year 2000, $D(j)$ is the ELF measure at aggregation level j, X is a matrix of controls, and ε is an error term. All data come from Desmet, Ortuño-Ortín and Wacziarg (2012), Ashraf and Galor (2013) and the references therein.

[Insert Table 2 here]

Table 2 starts by regressing a country's GDP per capita in 2000 on ELF at different levels of aggregation, with a basic set of geographic controls (latitude, percentage of arable land, mean distance to nearest waterway) and regional dummies. Comparing the first four columns, the effect of linguistic fractionalization is always negative. The statistical significance is maximized at ELF(9). The last four columns also control for legal origins and religious composition. This does not change the results: the effect of linguistic fractionalization is negative, and its predictive power is strongest at aggregation level 9. As in the case of economic growth, this suggests that relatively shallow divisions are enough to hurt economic development. Since there are 6 more levels of disaggregation - going from ELF(10) to ELF(15) - one could argue that ELF(9) represents an intermediate level of linguistic cleavages. Recall, however, that the correlation between ELF(9) and ELF(15) is 0.94, and that the difference between both indices is due to only a handful of mostly southern African countries.

[Insert Figure 4 here]

Figure 4 represents the standardized betas for all different levels of ELF corresponding to columns (1) to (4) in Table 2. As can be seen, the negative effect of fractionalization on economic development is maximized, both economically and statistically, at ELF(9). An increase by one standard deviation in ELF(9) lowers economic development by 16.7 percent when expressed as a share of the standard deviation in GDP per capita. As expected, the effect is largely unchanged for levels ELF(10) through ELF(15). To further illustrate the effect of ELF(9) on economic development, Figure 5 shows a scatterplot of column (7) from Table 2. It takes log of GDP per capita, partialled out from all the control variables in Column (7), and plots it against ELF(9), itself also partialled out from all the controls. The fitted line represents the negative partial relationship between ELF(9) and economic development.

[Insert Figure 5 here]

It is important to mention here that our results cannot strictly be interpreted as causal. As suggested by Greenberg (1956), among others, causality may run the other way, with economic development reducing the degree of linguistic diversity.[11] In fact, the two variables might have co-evolved in a complex ways. In order to provide a more convincing proof of causality, we would need data on linguistic diversity several generations ago. To the best of our knowledge, such data are not available for a large enough set of countries. Combined with the results on growth, however, where initial per capita income is controlled for on the right hand side, the level results are suggestive of an effect of linguistic diversity on growth.

Table 3 performs some further robustness checks. Hall and Jones (1999) argue that a country's level of development depends on its social infrastructure, which they define as policies favorable to productive activities and the accumulation of skills, rather than policies that promote rent-seeking, corruption and theft. In the first four columns of Table 3 we introduce the Hall and Jones (1999) measure of social infrastructure, which is a combination of government anti-diversion policies and the country's openness to free trade, as an additional control. Consistent with Hall and Jones (1999), social infrastructure has a positive effect on a country's level of development, but it does not change our basic insight. Although including social infrastructure somewhat weakens the statistical significance of linguistic fractionalization, ELF(9) continues to be significant at the 5 percent level.

[Insert Table 3 here]

Spolaore and Wacziarg (2009) find that the genetic distance to the technology leader constitutes a barrier to the diffusion of development. They argue that more closely related societies learn more from each other, so that the flow of ideas, knowledge and technology between two populations is facilitated if they share a more recent common ancestor. In the last four columns of Table 3, we therefore control for the genetic distance from the United States. As in Spolaore and Wacziarg (2009), we find that increasing the genetic distance to the U.S. lowers a country's income per capita. As for our variable of interest, the result is again unchanged: linguistic fractionalization continues to have a negative impact on a country's level of development, and its predictive power is maximized when ELF is measured based on linguistic groups at Level 9.

In recent work, Ashraf and Galor (2013) have found that development bears a hump-shaped relation with genetic diversity. In their theory diversity is good for innovation but bad for trust and coordination, so that there is an optimal level of diversity that maximizes development: on the one hand, higher diversity makes it harder to collaborate, which negatively affects efficiency and makes it harder for countries to operate at their production possibility frontier. On the other hand, higher diversity also implies more complementarities between people, making it more likely for countries to develop and adopt superior technologies, thus pushing out their production possibility frontier. Combining these two forces, they find that countries with intermediate levels of diversity perform best. Table 4 controls for genetic diversity and genetic diversity squared. It also allows for the timing of the Neolithic Revolution to affect today's level of development, a hypothesis advanced by Diamond (1997).[12] Our findings are consistent with those in Ashraf and Galor (2013). Turning to our variable of interest, the results are unchanged. ELF(9) continues to be statistically significant at the 5 percent level.

Taken together, these results suggest that fine divisions are enough to negatively impact a country's level of development. Even shallow cleavages can lead to inefficiencies. Markets become more segmented; trade and economic exchange encounter implicit barriers; and collaboration in productive activities becomes harder.

5 Conclusion

The depth of linguistic cleavages matters for political economy outcomes. Deep cleavages are associated with deleterious outcomes related to disagreements over the control of resources and common policies. For instance, measures of linguistic diversity based on deep cleavages, going back thousands of years, have a negative effect on civil conflict and redistribution. In contrast, more recent linguistic cleavages are sufficient to introduce barriers between populations, reducing their ability to communicate, interact and coordinate. These more superficial linguistic differences hinder growth and economic development by segmenting markets and limiting the scope for fruitful economic transactions.

Our explanation for these contrasting findings is based on drawing a distinction between the effects of linguistic cleavages on preferences (a demand-side explanation) versus their effect on technology (a supply-side explanation). Deep cleavages, because they originate earlier in history, are associated with starker differences in preferences, norms, values, attitudes and culture. In our recent work using data from the World Values Survey (Desmet, Ortuño-Ortín and Wacziarg, 2014) we show indeed that the degree of overlap between cultural values and ethnolinguistic identity is highly predictive of civil conflict. That is, countries where ethnicity helps predict cultural values and preferences are more likely to experience civil wars. This is entirely consistent with what we argue here, namely that deep cleavages - those most likely to be associated with deep cultural and preference differences between linguistic groups - are those most likely to generate conflict and low solidarity between groups.

In contrast, more superficial linguistic differences, sufficient to limit intelligibility and communication between distinct groups, introduce transactions costs and barriers, i.e. technological hindrances. These differences may be insufficient to generate deep disagreements in terms of preferences and culture, but are sufficient to create limits to coordination, cooperation and transactions, segmenting markets and reducing the scope of economic interactions. Our finding, detailed in this chapter, that linguistic diversity measured at fine levels of disaggregation has a negative effect on growth and development, is entirely consistent with this interpretation.

These findings shed some light on the mechanisms through which linguistic heterogeneity affects political economy outcomes, but much remains to be done. The precise mechanisms linking linguistic heterogeneity should be the subject of further research using a wide array of methodologies - not only cross-country comparative approaches but also more microeconomic and experimental approaches. Scholarly inquiry into these important questions is only in its infancy.

Notes

[1]Salient references include: 1) on civil conflict, Fearon and Laitin (2003), Montalvo and Reynal-Querol (2005) and Esteban et al. (2012). 2) On redistribution, Alesina et al. (2001), Alesina and Glaeser (2004), Desmet et al.

(2009) and Dahlberg et al. (2012). 3) On economic growth, Easterly and Levine (1997) and Alesina et al., (2003). 4) On public goods and governance, La Porta et al. (1999), Alesina et al. (2003), Habyarimana et al. (2007). For more general surveys of this vast and expanding literature, see Alesina and La Ferrara (2005) and Stichnoth and Van der Straeten (2013).

[2]See Ginsburgh and Weber (2015) in this book for a further discussion of how language trees are constructed.

[3]For a further discussion and an empirical analysis of the relationship between genetic and linguistic distances between countries, see the chapter by Spolaore and Wacziarg (2015), in this volume.

[4]See Barrett at al. (2001) for an alternative language classification with only 7 levels.

[5]The complete data set is available at `http://faculty.smu.edu/kdesmet/`.

[6]This regression corresponds to Table 4 in Desmet, Ortuño-Ortín and Wacziarg (2012) and is based on 103 countries. The exact list of controls, in addition to ELF at different levels of aggregation, is log GDP per capita, log population, a small island dummy, latitude, legal origin dummies and regional dummies.

[7]This regression corresponds to Table 6 in Desmet, Ortuño-Ortín and Wacziarg (2012) and is based on a single cross-section of 100 countries. The exact list of controls is log initial GDP per capita, investment share of GDP, average years of schooling, growth of population, log population, interaction between openness and log population, openness, legal origin dummies and regional dummies.

[8]One could wonder why the effect of diversity on growth is maximized at ELF(9), rather than at ELF(15). However, as already mentioned before, in nearly all countries ELF(9) and ELF(15) are identical, which also explains why in Panel B of Figure 3 the difference between ELF(9) and ELF(15) is minimal.

[9]For a discussion of some of this literature, see also the chapter by Sonntag (2015) in this book.

[10]One drawback is that these papers measure linguistic diversity as the share of the population who are speakers of the most widespread language, although Nettle (2000) also considers the number of languages per million of people and Nettle et al. (2007) considers an ELF index of diversity.

[11]See also De Grauwe (2006), Alesina and Reich (2014), and Amano et al. (2014).

[12]Note that genetic diversity and the timing of the Neolithic Revolution are "ancestry adjusted", meaning that it is based not on a country's geography, but on a country's ancestral population (Putterman and Weil, 2010). For example, the timing of the Neolithic Revolution for Australia is coded as closer to that of England due to the presence of a large population of English descent in Australia.

Bios

Klaus Desmet is the Altshuler Centennial Interdisciplinary Professor of Cities, Regions and Globalization at Southern Methodist University and a Research Fellow at the Centre for Economic Policy Research. He holds an MSc in Business and Engineering from the Université catholique de Louvain and a PhD in Economics from Stanford University. He previously was professor at Universidad Carlos III de Madrid. His research focuses on regional economics, international trade, economic growth, and diversity. In 2010 he was awarded the August Lösch Prize, together with Esteban Rossi-Hansberg, for his contributions to regional science. He has published in journals such as the *American Economic Review*, the *Journal of Economic Theory* and the *Journal of Development Economics.*

Ignacio Ortuño-Ortin is Professor of Economics at Universidad Carlos III de Madrid. He holds a PhD in Economics from University of California, Davis. His research interests focus on political

economy, public economics, intergenerational social mobility and cultural diversity. He has published in journals such as the *Journal of Public Economics*, *Journal of the European Economic Association*, *PlosOne* and the *Journal of Development Economics*.

Romain Wacziarg is Professor of Economics at the UCLA Anderson School of Management, and a research associate at the National Bureau of Economic Research. He holds a PhD from Harvard University, and was previously a professor at the Stanford Graduate School of Business. His research concerns the determinants of economic development, with a focus on institutional, cultural, and demographic factors. His papers have been published in such outlets as the *American Economic Review*, the *Quarterly Journal of Economics*, the *Journal of Economic Growth*, the *Journal of Development Economics* and the *Journal of International Economics*.

Index Terms

This chapter:
civil conflict, economic growth, ELF index, Ethnologue, fractionalization, GDP per capita, language families, language tree, level of development, levels of aggregation, linguistic cleavages, linguisitc diversity, phylogenetic, redistribution.

More general:
causality, communication, coordination, cultural values, Dravidian, genetic distance to the U.S., genetic diversity, Indo-European, intelligibility, long-run growth, Neolithic revolution, Nilo-Saharan, segmented markets, social infrastructure, solidarity, standardized beta, steady-state level of development, technology, transaction costs.

References

[1] D. Acemoglu, S. Johnson and J. Robinson (2001) 'The Colonial Origins of Comparative Development' *American Economic Review*, 91, 1369–1401.

[2] A. Alesina, A. Devleeschauwer, W. Easterly, S. Kurlat and and R. Wacziarg (2003) 'Fractionalization' *Journal of Economic Growth*, 8, 155–194.

[3] A. Alesina and E. Glaeser (2004) *Fighting Poverty in the U.S. and in Europe: A World of Difference* (New York: Oxford University Press).

[4] A. Alesina, E. Glaeser and B. Sacerdote (2001) 'Why Doesn't the U.S. Have a European-style Welfare System?' *Brookings Papers on Economic Activity*, 2, 187–254.

[5] A. Alesina and E. La Ferrara (2005) 'Ethnic Diversity and Economic Performance' *Journal of Economic Literature*, 43, 762–800.

[6] A. Alesina and B. Reich (2014) 'Nation Building' *NBER Working Paper* #18839.

[7] T. Amano, B. Sandel, H. Eager, E. Bulteau, J. C. Svenning, B. Dalsgaard, C. Rahbek, R. G. Davies and W.J. Sutherland (2014) 'Global Distribution and Drivers of Language Extinction Risk' *Proceedings of the Royal Society*, B 2014 281, 20141574.

[8] Q. Ashraf and O. Galor (2013) 'The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development' *American Economic Review*, 103, 1–46.

[9] D. B. Barrett, G. T. Kurian and T.M. Johnson (2001) *World Christian Encyclopedia; A Comparative Survey of Churches and Religions in the Modern World*, 2nd ed. (Oxford: Oxford University Press).

[10] E. M. S. Belle and G. Barbujani (2007) 'Worldwide Analysis of Multiple Microsatellites: Language Diversity Has a Detectable Influence on DNA Diversity' *American Journal of Physical Anthropology*, 133, 1137–1146.

[11] L. L. Cavalli-Sforza, A. Piazza, P. Menozzi and J. Mountain (1988) 'Reconstruction of Human Evolution: Bringing Together Genetic, Archaeological and Linguistic Data' *Proceedings of the National Academy of Sciences of the United States of America*, 85, 6002–6006.

[12] M. Dahlberg, K. Edmark and H. Lundqvist (2012) 'Ethnic Diversity and Preferences for Redistribution' *Journal of Political Economy*, 120, 41–76.

[13] P. De Grauwe (2006) 'Language Diversity and Economic Development' unpublished manuscript, Katholieke Universiteit Leuven.

[14] K. Desmet, I. Ortuño-Ortín and R. Wacziarg (2012) 'The Political Economy of Linguistic Cleavages' *Journal of Development Economics*, 97, 322–338.

[15] K. Desmet, I. Ortuño-Ortín and R. Wacziarg (2014) 'Culture, Identity and Diversity' Working Paper, UCLA.

[16] K. Desmet, I. Ortuño-Ortín and S. Weber (2009) 'Linguistic Diversity and Redistribution' *Journal of the European Economic Association*, 7, 1291–1318.

[17] J. Diamond (1997) *Guns, Germs and Steel: The Fates of Human Societies* (New York: W.W. Norton).

[18] W. Easterly, M. Kremer, L. Pritchett and L. Summers (1993) 'Good Policy or Good Luck? Country Growth Performance and Temporary Shocks' *Journal of Monetary Economics*, 32, 459–483.

[19] W. Easterly and R. Levine (1997) 'Africa's Growth Tragedy: Policies and Ethnic Divisions' *Quarterly Journal of Economics*, 112, 1203–1250.

[20] J. Esteban, L. Mayoral and D. Ray (2012) 'Ethnicity and Conflict: An Empirical Study' *American Economic Review*, 102, 1310–1342.

[21] Ethnologue (2005) *Ethnologue: Languages of the World*, 15^{th} Edition (SIL International).

[22] J. Fearon and D. Laitin (2003) 'Ethnicity, Insurgency, and Civil War' *American Political Science Review*, 97, 75–90.

[23] J. A. Fishman (1968) 'Some Contrasts between Linguistically Homogeneous and Linguistically Heterogeneous Polities' in J. Fishman, C. Ferguson and J. Das Gupta *Language Problems of Developing Nations* (New York: Wiley).

[24] J. A. Fishman (1991) 'An Inter-polity Perspective on the Relationships between Linguistic Heterogeneity, Civil Strife and per capita Gross National Product' *International Journal of Applied Linguistics*, 1, 5–18.

[25] V. Ginsburgh and S. Weber (2015) 'Linguistic Distances and their Use in Economics' in V. Ginsburgh and S. Weber *Palgrave Handbook of Economics and Language*, forthcoming.

[26] R. D. Gray and Q. D. Atkinson (2003) 'Language-tree Divergence Times Support the Anatolian Theory of Indo-European Origin' *Nature*, 426, 27 November, 435–439.

[27] J. H. Greenberg (1956) 'The Measurement of Linguistic Diversity' *Language*, 32, 109–15.

[28] J. Habyarimana, M. Humphreys, D. Posner and J. Weinstein (2007) 'Why Does Ethnic Diversity Undermine Public Goods Provision?' *American Political Science Review*, 101, 709–725.

[29] R. E. Hall and C. I. Jones (1999) 'Why do Some Countries Produce So Much More Output Per Worker than Others?' *Quarterly Journal of Economics*, 114, 83–116.

[30] R. La Porta, F. Lopez-de-Silanes, A. Shleifer and R. Vishny (1999) 'The Quality of Government' *Journal of Law, Economics, and Organization*, 15, 222–279.

[31] D. Laitin and D. Posner (2001) 'The Implications of Constructivism for Constructing Ethnic Fractionalization Indices' Apsa-cp (*Newsletter of the Organized Section in Comparative Politics of the American Political Science Association*), 12, 13–17.

[32] D. Laitin and R. Ramachandran (2014) 'Language Policy and Human Development' unpublished manuscript.

[33] J. G. Montalvo and M. Reynal-Querol (2005) 'Ethnic Polarization, Potential Conflict and Civil War' *American Economic Review*, 95, 796–816.

[34] D. Nettle (2000) 'Linguistic Fragmentation and the Wealth of Nations: The Fishman-Pool Hypothesis Reexamined' *Economic Development and Cultural Change*, 48, 335–348.

[35] D. Nettle, J. B. Grace, M. Choisy, H. V. Cornell, J. F. Guégan and M. E. Hochberg (2007) 'Cultural Diversity, Economic Development and Societal Instability' *PlosOne*, DOI: `10.1371/journal.pone.0000929`.

[36] S. L. Parente and E. C. Prescott (1994) 'Barriers to Technology Adoption and Development' *Journal of Political Economy*, 102, 298–321.

[37] J. Pool (1972) 'National Development and Language Diversity' in J. A. Fishman *Advances in the Sociology of Language*, Volume II (The Hague: Mouton).

[38] L. Putterman and D. N. Weil (2010) 'Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality' *Quarterly Journal of Economics*, 125, 1624-1682.

[39] R. R. Sokal (1988) 'Genetic, Geographic and Linguistic Distances in Europe' *Proceedings of the National Academy of Sciences of the United States of America*, 85, 1722–1726.

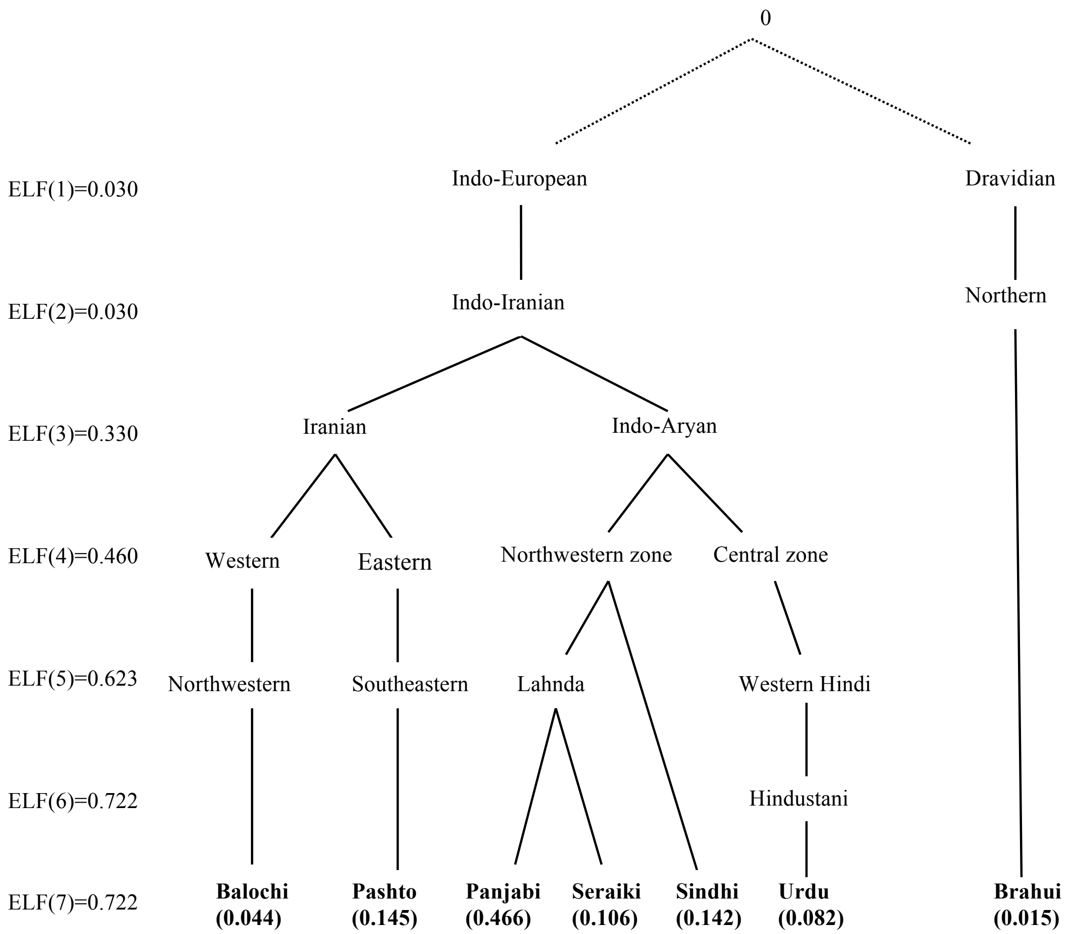
[40] S. K. Sonntag (2015) 'Languages, Regional Conflicts and Economic Development in South Asia' in V. Ginsburgh and S. Weber *Palgrave Handbook of Economics and Language*, forthcoming.

[41] E. Spolaore and R. Wacziarg (2009) 'The Diffusion of Development' *Quarterly Journal of Economics*, 124, 469–529.

[42] E. Spolaore and R. Wacziarg (2015) 'Ancestry, Language and Culture' *Working paper, UCLA and Tufts University*, January.

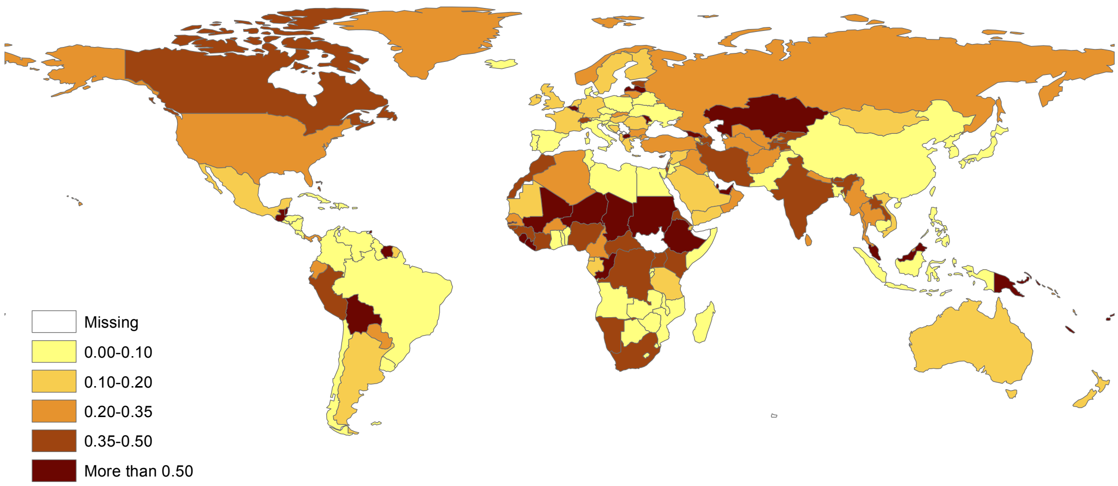
[43] H. Stichnoth and K. Van der Straeten (2013) 'Ethnic Diversity, Public Spending, and Individual Support for the Welfare State: A Review of the Empirical Literature' *Journal of Economic Surveys*, 27, 364–389.

Figure 1: Phylogenetic Tree of Main Languages Spoken in Pakistan

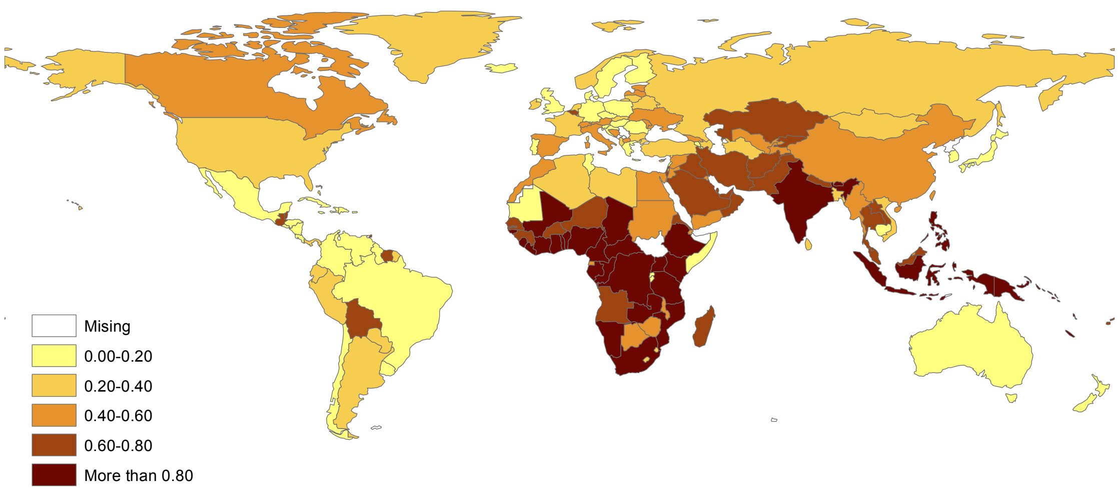


Source: Desmet, Ortuño-Ortín and Wacziarg (2012)

Figure 2: Linguistic Fractionalization at Different Levels of Aggregation: ELF(2) and ELF(15)

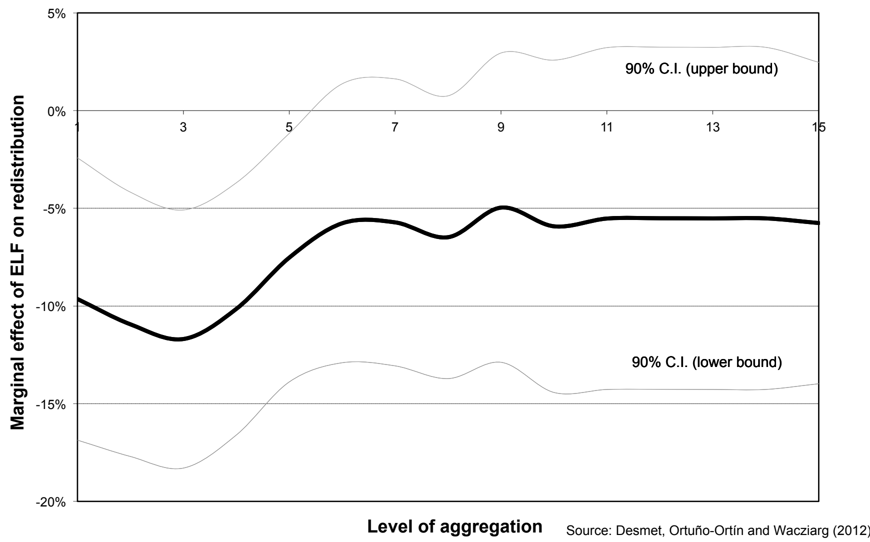


A. ELF(2)

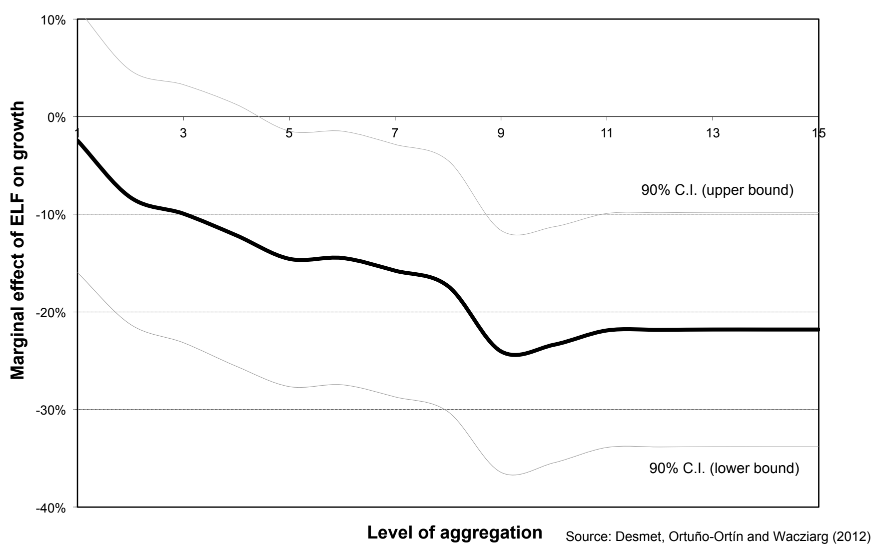


B. ELF(15)

Figure 3: Effect of a One Standard Deviation Increase in ELF



A. On Redistribution (as % of standard deviation of redistribution)



B. On Growth (as % of standard deviation of growth)

Figure 4: Effect of a One Standard Deviation Increase in ELF on GDP per Capita (expressed as % of s.d. in GDP per capita)

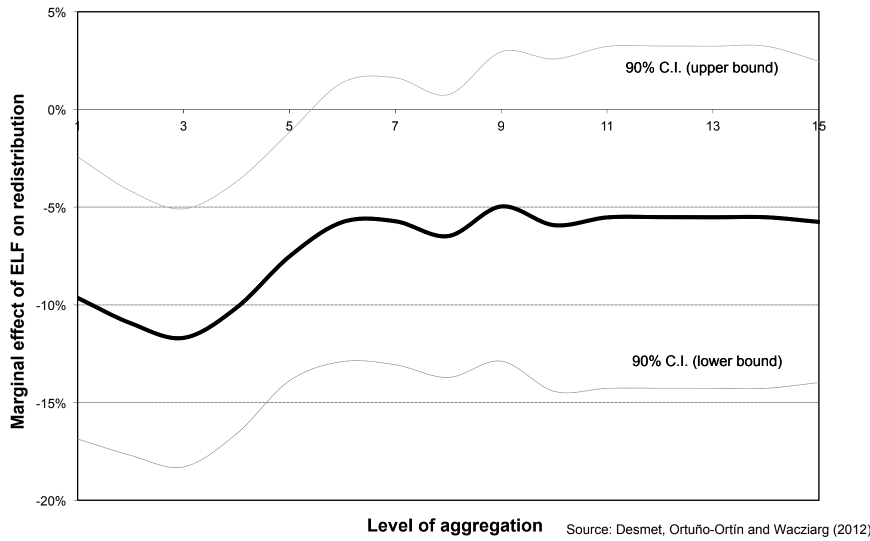


Figure 5: Conditional Log GDP per Capita vs ELF(9)

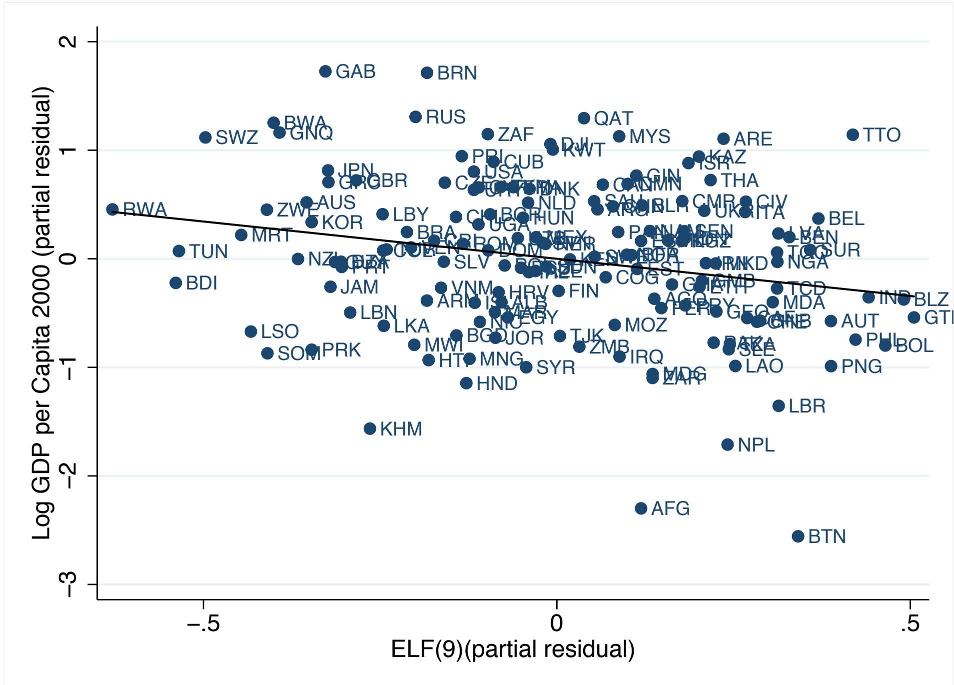


Table 1 – Summary Statistics ELF

Panel A. Means and Standard Deviations

Variable	Mean	Std. Dev.	Min	Max
ELF(1)	0.156	0.18	0	0.647
ELF(3)	0.241	0.221	0	0.818
ELF(6)	0.328	0.272	0	0.941
ELF(9)	0.377	0.292	0	0.987
ELF(15)	0.412	0.308	0	0.99

(226 observations)

Panel B. Correlations

	ELF(1)	ELF(3)	ELF(6)	ELF(9)	ELF(15)
ELF(1)	1				
ELF(3)	0.77	1			
ELF(6)	0.579	0.826	1		
ELF(9)	0.56	0.748	0.9	1	
ELF(15)	0.526	0.672	0.798	0.943	1

(226 observations)

Source: Desmet, Ortuño-Ortín and Wacziarg (2012)

Table 2 – Log Income per Capita in 2000 and ELF at Different Levels of Aggregation

	(1) ELF1	(2) ELF6	(3) ELF9	(4) ELF15	(5) ELF1	(6) ELF6	(7) ELF9	(8) ELF15
ELF	**-0.44**	**-0.833*** **	**-0.931*** **	**-0.659** **	**-0.234**	**-0.433* **	**-0.686*** **	**-0.490* **
(different levels of aggregation)	**[-1.05]**	**[-3.25]**	**[-3.58]**	**[-2.37]**	**[-0.59]**	**[-1.71]**	**[-2.76]**	**[-1.88]**
Log absolute latitude	0.161	0.145	0.116	0.129	0.192*	0.190**	0.167*	0.173*
	[1.53]	[1.46]	[1.16]	[1.25]	[1.93]	[1.98]	[1.76]	[1.79]
Percentage of arable land	-0.020***	-0.021***	-0.021***	-0.020***	-0.018***	-0.018***	-0.018***	-0.018***
	[-3.52]	[-3.86]	[-3.92]	[-3.74]	[-3.33]	[-3.36]	[-3.48]	[-3.47]
Mean distance	-0.687***	-0.700***	-0.676***	-0.698***	-0.479***	-0.486***	-0.450***	-0.467***
to nearest waterway	[-4.06]	[-4.39]	[-4.26]	[-4.29]	[-2.94]	[-3.07]	[-2.88]	[-2.95]
Latin America and Carribean	-0.520**	-0.702***	-0.759***	-0.697***	-0.984***	-1.037***	-1.130***	-1.116***
	[-2.21]	[-2.98]	[-3.20]	[-2.84]	[-3.99]	[-4.22]	[-4.60]	[-4.42]
Sub-Saharan Africa	-1.618***	-1.611***	-1.530***	-1.477***	-1.694***	-1.698***	-1.668***	-1.623***
	[-6.92]	[-7.28]	[-6.98]	[-6.50]	[-7.63]	[-7.91]	[-7.94]	[-7.58]
East and Southeast Asia	-0.702**	-0.715**	-0.699**	-0.708**	-0.580**	-0.578**	-0.563**	-0.580**
	[-2.47]	[-2.60]	[-2.56]	[-2.53]	[-2.09]	[-2.11]	[-2.09]	[-2.12]
French legal origin					-0.275	-0.153	-0.011	-0.083
					[-0.48]	[-0.27]	[-0.02]	[-0.14]
German legal origin					0.562	0.653	0.722	0.682
					[0.87]	[1.01]	[1.14]	[1.06]
Socialist legal origin					-0.443	-0.381	-0.304	-0.333
					[-0.77]	[-0.67]	[-0.54]	[-0.59]
UK legal origin					-0.017	0.077	0.204	0.156
					[-0.03]	[0.14]	[0.39]	[0.29]
Share of Muslims					0	0	0	0
					[-0.10]	[0.03]	[-0.07]	[-0.18]
Share of Roman Catholics					0.010***	0.009***	0.009***	0.010***
					[3.29]	[3.01]	[2.97]	[3.12]
Share of Protestants					0.010*	0.010*	0.010**	0.010*
					[1.91]	[1.90]	[2.00]	[1.98]
Constant	9.157***	9.499***	9.651***	9.492***	8.825***	8.888***	8.982***	8.936***
	[21.03]	[22.64]	[22.52]	[21.09]	[12.46]	[12.70]	[13.03]	[12.76]
Observations	152	152	152	152	150	150	150	150
R-squared	0.5078	0.5381	0.5447	0.5227	0.6295	0.6364	0.6484	0.6381

t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 3 – Log Income per Capita in 2000 and ELF at Different Levels of Aggregation: Robustness

	(1) ELF1	(2) ELF6	(3) ELF9	(4) ELF15	(5) ELF1	(6) ELF6	(7) ELF9	(8) ELF15
ELF	**0.046**	**-0.231**	**-0.414****	**-0.18**	**-0.124**	**-0.548****	**-0.724*****	**-0.451***
(different levels of aggregation)	**[0.14]**	**[-1.17]**	**[-2.10]**	**[-0.87]**	**[-0.32]**	**[-2.15]**	**[-2.94]**	**[-1.75]**
Log absolute latitude	0.159**	0.147**	0.133*	0.145**	0.133	0.104	0.089	0.11
	[2.15]	[2.08]	[1.90]	[2.02]	[1.30]	[1.04]	[0.90]	[1.10]
Percentage of arable land	-0.014***	-0.014***	-0.015***	-0.014***	-0.018***	-0.018***	-0.018***	-0.018***
	[-3.09]	[-3.18]	[-3.31]	[-3.22]	[-3.32]	[-3.47]	[-3.57]	[-3.50]
Mean distance	-0.422**	-0.410**	-0.391**	-0.411**	-0.428**	-0.405**	-0.376**	-0.410**
to nearest waterway	[-2.27]	[-2.27]	[-2.20]	[-2.27]	[-2.59]	[-2.54]	[-2.38]	[-2.55]
Latin America and Carribean	-0.410*	-0.442**	-0.528**	-0.473**	-0.895***	-0.928***	-1.027***	-1.015***
	[-1.85]	[-1.99]	[-2.36]	[-2.03]	[-3.55]	[-3.74]	[-4.14]	[-3.94]
Sub-Saharan Africa	-1.189***	-1.210***	-1.202***	-1.182***	-1.238***	-1.153***	-1.153***	-1.190***
	[-5.99]	[-6.25]	[-6.31]	[-6.08]	[-3.89]	[-3.74]	[-3.80]	[-3.85]
East and Southeast Asia	-0.462*	-0.428*	-0.391	-0.437*	-0.288	-0.211	-0.22	-0.292
	[-1.92]	[-1.78]	[-1.65]	[-1.81]	[-0.91]	[-0.68]	[-0.72]	[-0.94]
French legal origin	0.173	0.254	0.327	0.246	-0.112	0.142	0.243	0.084
	[0.39]	[0.58]	[0.75]	[0.55]	[-0.19]	[0.24]	[0.41]	[0.14]
German legal origin	0.378	0.421	0.443	0.41	0.732	0.923	0.958	0.853
	[0.79]	[0.88]	[0.94]	[0.86]	[1.11]	[1.41]	[1.49]	[1.30]
Socialist legal origin	0.382	0.425	0.428	0.41	-0.202	0.001	0.036	-0.08
	[0.75]	[0.83]	[0.85]	[0.80]	[-0.34]	[0.00]	[0.06]	[-0.13]
UK legal origin	0.269	0.335	0.393	0.337	0.163	0.4	0.486	0.347
	[0.67]	[0.84]	[1.00]	[0.84]	[0.29]	[0.71]	[0.88]	[0.61]
Share of Muslims	-0.004	-0.004	-0.004	-0.004	0	0	0	-0.001
	[-1.60]	[-1.42]	[-1.54]	[-1.62]	[-0.14]	[0.05]	[-0.10]	[-0.20]
Share of Roman Catholics	0.002	0.001	0.001	0.002	0.010***	0.009***	0.009***	0.010***
	[0.66]	[0.54]	[0.54]	[0.62]	[3.36]	[3.03]	[3.03]	[3.19]
Share of Protestants	0.003	0.003	0.004	0.003	0.013**	0.014**	0.014**	0.013**
	[0.70]	[0.78]	[0.85]	[0.78]	[2.24]	[2.47]	[2.53]	[2.34]
Social infrastructure	2.042***	2.026***	1.971***	2.002***				
	[5.57]	[5.57]	[5.48]	[5.45]				
Genetic distance to the U.S.					-0.050*	-0.063**	-0.059**	-0.049*
					[-1.86]	[-2.34]	[-2.26]	[-1.87]
Constant	7.719***	7.787***	7.900***	7.817***	9.052***	9.203***	9.258***	9.164***
	[12.44]	[12.73]	[13.04]	[12.62]	[12.77]	[13.17]	[13.45]	[13.05]
Observations	112	112	112	112	148	148	148	148
R-squared	0.838	0.8403	0.8451	0.8393	0.6348	0.6469	0.6571	0.6428

t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 4 – Log Income per Capita in 2000, Predicted Genetic Diversity and ELF at Different Levels of Aggregation

	(1) ELF1	(2) ELF6	(3) ELF9	(4) ELF15
ELF	0.313	-0.392	-0.590**	-0.306
(different levels aggregation)	[0.78]	[-1.41]	[-2.17]	[-1.25]
Log absolute latitude	0.183	0.168	0.159	0.16
	[1.60]	[1.55]	[1.51]	[1.42]
Percentage of arable land	-0.021***	-0.022***	-0.022***	-0.022***
	[-3.88]	[-4.30]	[-4.50]	[-4.32]
Mean distance	-0.423*	-0.410*	-0.398*	-0.404*
to nearest waterway	[-1.76]	[-1.84]	[-1.83]	[-1.79]
Latin America and Carribean	-0.967***	-1.048***	-1.136***	-1.077***
	[-3.90]	[-3.92]	[-3.95]	[-3.87]
Sub-Saharan Africa	-1.427***	-1.229***	-1.150***	-1.268***
	[-4.51]	[-3.92]	[-3.74]	[-4.09]
East and Southeast Asia	-0.522	-0.498	-0.434	-0.492
	[-1.31]	[-1.35]	[-1.18]	[-1.28]
French legal origin	-0.319	-0.139	-0.058	-0.168
	[-0.66]	[-0.29]	[-0.12]	[-0.35]
German legal origin	0.271	0.37	0.374	0.327
	[0.51]	[0.75]	[0.82]	[0.65]
Socialist legal origin	-0.593	-0.487	-0.484	-0.508
	[-1.18]	[-1.00]	[-1.04]	[-1.04]
UK legal origin	-0.161	0.016	0.086	0.002
	[-0.36]	[0.04]	[0.20]	[0.00]
Share of Muslims	-0.009***	-0.008***	-0.009***	-0.009***
	[-3.35]	[-3.13]	[-3.30]	[-3.43]
Share of Roman Catholics	0.005*	0.004	0.004	0.005
	[1.72]	[1.53]	[1.58]	[1.60]
Share of Protestants	0.005	0.007	0.007	0.006
	[0.76]	[1.13]	[1.24]	[1.05]
Predicted diversity	292.464***	259.711***	247.288***	257.583***
(ancestry adjusted)	[3.57]	[3.35]	[3.11]	[3.28]
Predicted diversity squared	-205.384***	-183.971***	-175.261***	-181.806***
(ancestry adjusted)	[-3.55]	[-3.35]	[-3.12]	[-3.27]
Neolithic Revolution timing	0.317	0.543**	0.578**	0.454**
(ancestry adjusted)	[1.26]	[2.17]	[2.55]	[2.00]
Constant	-97.279***	-86.708***	-82.528***	-85.474***
	[-3.40]	[-3.20]	[-2.97]	[-3.10]
Observations	144	144	144	144
R-squared	0.669	0.673	0.682	0.673

t-statistics in brackets

*** p<0.01, ** p<0.05, * p<0.1

Bootstrapped standard errors, accounting for the use of generated regressors, as in Ashraf and Galor (2013)