Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Political Economy and Development

Autor:

Joseph Flavian Gomes

Director/es:

Klaus Desmet

DEPARTAMENTO/INSTITUTO

ECONOMÍA

Getafe, diciembre de 2013

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Political Economy and Development

Autor: *Joseph Flavian Gomes*

Director/es: **Klaus Desmet**

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Secretario:

Calificación:

Getafe, de de

Universidad
Carlos III de Madrid
www.uc3m.es

Abstract

My PhD thesis comprises of three chapters on Political Economy and Development. The first chapter contributes to a burgeoning literature that uses sub-national micro data to identify the causes of civil conflicts. In particular, I study the Maoist conflict in India by constructing a comprehensive district level database using conflict data from four different terrorism databases and combining it with socioeconomic and geography data from myriad sources. In addition to exploiting the within country regional heterogeneity, I use the micro structure of the data to construct group-level characteristics. Using data on 360 districts for 3 time periods, I find evidence on how land inequality and lower incomes are important for the conflict. Moreover, making use of the micro structure of the data I am able to ask whether exclusion of the low castes and tribes from the growth story of India is important. I find that the growth of incomes of Scheduled Tribes significantly decreases the intensity of the conflict. Finally, I show how historical property rights institutions from colonial times that go back centuries can affect present day conflict outcomes through their impact on economic outcomes, social relations and the political environment in the district.

In the second chapter, I compute new measures of religious diversity and intolerance and study their effects on civil conflict. Using a religion tree that describes the relationship between different religions, I compute measures of religious diversity at three different levels of aggregation. I find that religious diversity is a significant and robust correlate of civil conflict. While religious fractionalization significantly reduces conflict, religious polarization increases it. This is most robust at the second level of aggregation which implies that the cleavage between Hindus, Muslims, Jews, and Christians etc. is more relevant than that between either subgroups of religions like Protestants and Catholics, Shias and Sunnis, etc. or that between higher levels of aggregation like Abrahamic and Indian religions. I find religious intolerance to be a significant and robust predictor of conflict.

In the third and final chapter of the thesis, I show that ethnic distances can explain the huge disparities in child mortality rates across ethnic groups in Africa. Using high quality individual level micro data from the Demographic and Health Surveys for fourteen Sub-Saharan African countries combined with a novel high resolution dataset on the distribution of ethnic groups across space I show that children whose mothers have a higher linguistic distance from their neighbours have a higher probability of dying before reaching the age of five. On the other hand linguistic diversity measured by fractionalization or polarization reduces the probability of child death. One possible

explanation for my findings is that ethnic diversity reflects a higher stock of knowledge and information which leads to better health outcomes. However, such knowledge does not flow smoothly to groups which are linguistically distant and thus such groups lose out.

Resumen

Mi tesis doctoral consta de tres capítulos sobre Economía Política y Desarrollo Económico. El primer capítulo contribuye a la literatura que utiliza microdatos sub-nacionales para identificar las causas de los conflictos civiles. En particular, estudio el conflicto maoísta en la India mediante la construcción de una base de datos a nivel de distrito. Para construir este base utilizo datos de conflicto a partir de cuatro bases de datos distintas de terrorismo, y las combino con datos socioeconómicos y geográficos a partir de múltiples fuentes. Además de explotar la heterogeneidad regional dentro del país, uso la micro estructura de los datos para construir características a nivel de grupo. Utilizando datos de 360 distritos para 3 períodos de tiempo, encuentro evidencias de que la desigualdad en la distribución de tierra y el nivel de renta en el districto son factores importantes para explicar el conflicto. Por otra parte, haciendo uso de la microestructura de los datos estudio si la exclusión de grupos marginados (como castas bajas y tribus) del crecimiento experimentado por la India en los últimos años, es un factor importante para explicar el nivel de conflicto. Encuentro evidencia que el crecimiento de los ingresos de las tribus desfavorecidas disminuye significativamente la intensidad del conflicto. Finalmente, muestro cómo las instituciones de derechos de propiedad de la época colonial, que se remontan siglos atrás, pueden todavía afectar el nivel actual de conflicto a través de su impacto en los resultados económicos, las relaciones sociales y el entorno político en el distrito.

En el segundo capítulo, calculo nuevas medidas de diversidad religiosa e intolerancia y estudio sus efectos sobre el conflicto civil. Haciendo uso de un árbol de las religiones que describe la relación entre las diferentes religiones, puedo calcular medidas de la diversidad religiosa en tres niveles diferentes de agregación. Encuentro evidencia que la diversidad religiosa es un correlato significativo y robusto de los conflictos civiles. Mientras el fraccionamiento religioso reduce significativamente el nivel de conflicto, la polarización religiosa lo aumenta. Esto es más robusto en el segundo nivel de agregación, lo que implica que la división entre hinduistas, musulmanes, judíos, cristianos, etc. es más relevante que la que existe entre cualquiera de los subgrupos de religiones como protestantes y católicos, los chiitas y los sunitas, etc., o entre niveles más altos de agregación, como las religiones abrahámicas y las de la India. Finalmente, encuentro evidencia que la intolerancia religiosa es un predictor significativo y robusto de los conflictos.

En el tercer y último capítulo de la tesis, muestro que las distancias étnicas pueden explicar las enormes disparidades en las tasas de mortalidad infantil entre los grupos étnicos de África. Utilizando microdatos a nivel individual de alta calidad (de las Encuestas de Demografía y Salud (DHS)) para catorce países del África Subsahariana, combinados con una nueva base de datos de alta resolución sobre la distribución de los grupos étnicos sobre el espacio, encuentro evidencia que los niños cuyas madres tienen una mayor distancia linguística de sus vecinos tienen una mayor probabilidad de morir antes de cumplir los cinco años. Por otra parte, la diversidad linguística medida por fraccionamiento o polarización reduce la probabilidad de muerte del niño. Una posible explicación de mis hallazgos es que la diversidad étnica refleja un mayor stock de conocimientos e información que lleva a mejores resultados de salud. Sin embargo, tal conocimiento no fluye adecuadamente a grupos que son linguísticamente distantes, y por lo tanto éstos grupos terminan siendo perjudicados.

Acknowledgements

I am extremely grateful to Klaus Desmet for being such a wonderful supervisor. Not only has he dedicated a lot of time and effort towards my PhD, he has always treated me as a colleague rather than his PhD student. The many discussions we have had on myriad topics not necessarily related to my narrow research area have taught me a lot and have helped me think beyond my immediate research.

I am also very grateful to Ulrich Wagner and Ignacio Ortuño-Ortn. While on paper they were not my PhD advisors, in practice they have always responded to me as if they were. They have always been kind enough to discuss and read my research multiple times. Among the other professors at UC3M I would especially like to thank Irma Clots-Figueras for taking time out to talk to me, read my papers and giving me good feedback. I would also like to thank Jesus Carro for giving me a helping hand for different research related issues and doubts that I have had. I am also grateful to Ricardo Mora, Raquel Carrasco, Matilde Machado, Natalia Fabra, Juan Jose Dolado and Diego Puga for their feedback and help on my different projects.

I would like to thank Jean-Philippe Platteau and the Economics department (CRED) at the Universit de Namur for their hospitality. Likewise, I would like to thank Romain Wacziarg and UCLA Anderson for their hospitality. I had a great time while visiting both these institutions and have learnt a great deal while talking to different professors at these departments.

During the PhD I have been lucky to earn some friends for life. I would particularly like to thank Robert Kirkby. I have learnt a great deal from the many interesting discussions on Economics and many other issues with him at the university, at our favourite Chinese restaurant and many different places. I would also like to thank Francesco Risi for being such a good officemate and friend. The stories of his biking and other adventures have helped me not to forget about life outside the university. Anahi Francis and Cecilia Vives have been and will be invaluable friends to me. I would also like to thank my other university mates and friends. Had it not been for the company of these wonderful people this PhD journey would have been very rough.

I dont think I can thank my parents enough for everything they have done for me. The same holds true for all the support Lovleen has always extended towards me unhesitatingly. But either way, thanks a lot mom and dad. Thanks a lot Lovleen.

Contents

List of Figures

List of Tables

To mom and dad.

Chapter 1

The Political Economy of the Maoist Conflict in India: An Empirical Analysis

1.1 Introduction

In recent years the relation between economic performance and civil conflicts has generated a considerable amount of interest among economists. Within the span of a few years a lot has been written on the subject.[1] A good part of the literature has taken a cross-country approach using aggregate data to identify the causes of civil conflicts.[2] However, there is a small but burgeoning literature showing that going to the sub-national level is key. Conflicts are often localized, and have to do with the unequal spatial distribution of resources within countries. Thus, treating countries as being homogeneous is often problematic. In addition, the use of aggregate data in most studies limits the kind of questions one can address. For example, if ethno-linguistic diversity is shown to have an impact on the probability of conflict, aggregate data do not allow us to determine whether this is solely picking up the effect of cultural diversity or whether it is also proxying for economic heterogeneity across groups. To address this question, one needs microdata that give information on, say, income at the level of different ethnic groups.

This paper uses micro data at the sub-national level to analyze the Maoist (aka Naxalite) conflict in India. The conflict started as a localized land conflict in Naxalbari (hence the name Naxalite), a village in West Bengal in 1967. However, it has seen a terrifying

[1] See Blattman and Miguel [2010] for a recent survey of the existing literature.

[2] See for e.g. Collier and Hoeffler [2004], Fearon and Laitin [2003], Miguel et al. [2004], Ciccone [2010], Besley and Persson [2008].

increase in proportions only in the last decade. In the period 2004-2010 there have been more than 5000 lives lost (even by official estimates). Including the number of wounded and displaced would make the figure many times higher. In fact, it has been identified as "the single biggest security challenge to the Indian state" by Dr. Manmohan Singh, the Prime Minister of India. While on paper the aim of the movement is to establish a "people's democratic state under the leadership of the proletariat" (Harris [2010]), at the heart of the conflict is land (rights, acquisition and its unequal distribution) and "in practice land redistribution appears to be one of the main goals" (Iyer [2009]). In addition, there is an ethnic/caste element to the conflict, as some tribal groups are at the lower end of the income distribution, and feel they are being left behind the rising tide of the Indian economy in the last decades (Guha [2007]).[3]

Although the conflict has spread over several states across India, by no means is it affecting all regions in the same way. The goal of the paper is to exploit this spatial heterogeneity to understand the sources of the conflict. In particular we try to address some of the following questions: How important is land inequality? Are tribal groups resorting to violence because of being left behind? Does the spatial heterogeneity in colonial institutions help us explain the current distribution of violence?

With the above goal in mind we use district level conflict data for the period 1979-2009 along with socio-economic and geography data from multiple National Sample Surveys (NSS), Censuses etc. to build a district level dataset. To account for differences across tribal groups and castes, we use microdata to construct economic variables at the level of these groups for each district. This gives us a comprehensive dataset of 360 districts (for the 16 main states which constitutes $> 90\%$ of the population) over three time periods. We use Probit regressions to explain the probability of conflict and Negative Binomial regressions to explain the intensity of conflict at the district level. The main findings of the paper are listed below.

A first finding is that land inequality is one of the key determinants of the conflict controlling for all other factors that the literature has found to be important. Land inequality reflects not just the inequality in the distribution of land but also differences in the socio-economic lives of people in a predominantly agrarian society. Moreover, it also implies more scope for inadequate compensation under land acquisition.[4]

District income on the other hand, also comes out to be an important determinant of the conflict. In the Probit and Negative Binomial regressions low income comes out to be a significant predictor of conflict. Moreover, we obtain similar results in the

[3] Moreover, a lot of the land acquisition is taking place in the tribal areas.

[4] We use a time invariant measure of land inequality from a period prior to the conflict. However, there is still scope for omitted variable bias and thus caution should be exercised while giving a causal explanation to our results

2SLS, Instrumental variable regressions using its own lagged values as an instrument for income. We find that lower income significantly increases both the probability and the intensity of conflict.

The evidence on the being-left-behind hypothesis of the disadvantaged groups, i.e., the low castes and the tribes, is mixed. One often heard argument is that tribal groups and lower castes resort to violence because their groups are not equally benefiting from the high rates of growth. We find that the income levels of the Scheduled Castes or Scheduled Tribes have no significant impact on the conflict. However, when we look at the growth in incomes of the three groups separately, the picture changes. We find that a lower growth in incomes of the Scheduled Tribes significantly increases the intensity of conflict. Moreover, in some specifications, we also find that the presence of the Scheduled Tribes leads to more conflict.

A final finding is that historical institutions matter. Class antagonism driven by land institutions that have lingered for centuries has a significant impact on both conflict presence and intensity. Districts where land rights were traditionally enjoyed by landlords have higher conflict compared to districts where land rights were traditionally with the farmers themselves.

The existing literature has witnessed several different approaches to empirically identify the causes of civil conflicts. There are two clear directions in which this literature needs progress. The first direction is using sub-national micro data in order to overcome the shortcomings of the cross country analyses. The second crucial issue is to establish a causal relation between conflict and its determinants.

With regard to the use of cross country data in the analysis of civil conflict, Do and Iyer [2009] point out two caveats: (1) Data might not be comparable across countries. (2) Reasons for the conflict might vary from country to country.[5] Another serious shortcoming of such studies is that they ignore the within country heterogeneity by treating the country as a unit of observation.[6] Conflicts are often localized and depend on the unequal spatial distribution of resources within the country. For example, in the context of the Maoist conflict in India, in West Bengal, one of the severely affected states, the conflict is very pronounced in the Midnapore and Puruliya districts while it is completely absent in districts like Howrah, North and South 24 Parganas.[7] This is

[5] Further, even within a country focusing on one specific conflict might give us more interesting insights than looking at overall violence.

[6] Also, they do not allow us to control for all the factors that are constant within the country viz. the macroeconomic variables.

[7] Again, if one looks at Maharashtra (which is a state located in the west of the country), almost all the Maoist incidents have been concentrated solely in the Gadchiroli district (formerly part of the Chandrapur district). Again, while the conflict affects virtually all the districts of Bihar, it is almost completely absent in the states of Rajasthan, Punjab and Haryana.

the kind of heterogeneity that one cannot take into account using even states as units of analysis. Moreover, by making use of micro data we can take into account the differences in incomes of disadvantaged groups vis-a-vis others and also the heterogeneity in the distribution of these groups.

The other critical issue in this literature is establishing a causal relation between conflict and its determinants. This is due to two main problems, as highlighted by Do and Iyer [2009] (1) There might exist unmeasured factors that affect both conflict intensity and pre conflict characteristics. (2) Districts that are experiencing more violence might also be districts that have experienced high past conflict. Recent studies have tried to address this issue using an instrumental variable approach when clearly exogenous instruments have been available. Miguel et al. [2004] and Ciccone [2010] have used weather shocks. Dube and Vargas [2008] have used exogenous shocks to agricultural and resource prices. When clear instruments have not been available authors have tried to use data on covariates from the pre conflict period in order to prevent endogeneity arising out of reverse causality.[8] Also, all potential covariates are controlled for in order to reduce endogeneity arising out of omitted variables. Following in the same vein, in this paper we use data from the pre conflict period; control for all possible variables (that the literature suggests are important) that might play a role subject to the data availability and always control for the presence of past conflict. In addition, we also run IV regressions with previous period income as an instrument for present income.

As far as the literature on the Naxalite conflict itself is concerned, there are very few rigorous empirical studies. Barooah [2008] relying on a simple cross section OLS analysis finds that the probability of conflict in the district is increasing in the poverty rate and is decreasing in the literacy rate. Hoelscher et al. [2011] also using a cross section and relying on probit and negative binomial techniques, find forest cover, prevalence of conflict in the neighbouring district and presence of Scheduled Castes and Tribes to be important. Gawande et al. [2012] using a district level panel find that negative natural resource shocks increase the intensity of conflict. Vanden Eynde [2011] on the other hand also using a district level panel finds that negative labour income shocks increases violence against civilians to prevent them from being recruited as police informers. While all these papers are important for understanding the nature and causes of the Maoist conflict in India, they ignore some important factors like land inequality, historical land institutions and the exclusion of the tribals in India, which are crucial in understanding the conflict.[9]

---

[8] e.g. Do and Iyer [2009]; Mitra and Ray [2010]

[9] There exist several descriptive case studies on the issue. See Harris [2010] for a summary.

Thus, to summarize, this paper contributes to several different strands of the literature. The first strand is the research using sub-national-micro data exploiting the spatial heterogeneity within a country and the micro characteristics of the data to pin down the causes of civil conflict.[10] This is a clear progress over existing cross country literature. Moreover, we show how horizontal inequality in growth rates matters rather than growth itself. We see that while overall growth/or the lack of it does not affect the conflict, the low growth in incomes of the Scheduled Tribes significantly increases the intensity of conflict, controlling for income growth of other ethnic groups. Finally this paper also contributes to the broad class of literature that traces divergences in current economic outcomes to differences in historical institutions in a country.[11] We show that, in addition to economic underdevelopment, land relations and historical institutions within a country could lead to conflict.

In the next section we discuss the Maoist conflict in India, in Section 3 we list the main hypotheses of the study. Section 4 gives the empirical analysis and results and Section 5 concludes.

1.2 The Maoist/Naxalite Conflict

The start of the Maoist conflict is marked by a peasant uprising in the year 1967 in Naxalbari, a small village in the Darjeeling district of West Bengal. "A tribal youth having obtained a judicial order went to plough his land on 2 March 1967. The local landlords attacked him with the help of their goons. Tribal [peasants] of that area retaliated and started forcefully recapturing their land" [Kujur, 2008]. The rebellion left nine tribal people and one police personnel dead and the Naxalite movement in India was born. The rebellion itself was contained by government forces within 72 days with the use of force, but had already gathered huge visibility from Communist revolutionaries from across the country.

After West Bengal the movement spread to the state of Andhra Pradesh where the formation of the People's War Group (PWG) in 1980 marks the revival of the movement post the Naxalbari uprising. It has since then spread across various states in India including Bihar, Jharkhand, Madhya Pradesh, Orissa, Chhattisgarh, Maharashtra and Karnataka across many districts and has existed in varying degrees across the country.[12] However, it was the 2004 merger of the PWG with the Maoist Communist Center (MCC) that lead to the formation of the Communist Party of India-Maoist (CPI-Maoist) that

[10] Some of the recent papers in this literature are Do and Iyer [2009], Dube and Vargas [2008], Jha [2008], Iyer [2009].

[11] A very novel concept in the conflict literature. Jha [2008] is the other paper in this context.

[12] See map in Figure B.1 in appendix B.

marks the modern revival of the movement and followed a huge rise in insurgency and violence thereafter.

While the term "Naxalite" comes from the place of birth of the movement the term "Maoist" is used due to the Maoist ideologies that many of these rebel groups adhere too. The CPI-Maoists for example claim to be committed to a "democratic revolution" through "a protracted people's war with the armed seizure of power remaining as its central and principal task" (Iyer, 2009). While the conflict as a whole has been termed as a Maoist/Naxalite conflict, the movement itself is hardly homogeneous. It has in fact, always had a fragmented structure with multiple groups operating without a centralized movement organization. At present, the CPI-(Maoist) is the biggest operating group.

There have been more than 5000 lives lost (civilians, rebels and security personnel) and more than 12000 incidents of violence in the period of 2004-2010 due to the Naxalite conflict. The number of people displaced in Chhattisgarh alone was more than 43000 (Sundar [2008]), end of 2006. The geographical spread of the conflict has seen a phenomenal increase from 55 districts in 8 states in 2003 to 194 districts spread across 18 states in 2007 [Iyer, 2009]. "While it is difficult to put an exact figure to the number of rebels, there is an estimated 10000 to 20000 full time fighters with countless thousands of village militias controlling particularly remote jungle areas where the state is hardly present." (The Economist).

Bhatia [2005] focusing on Central Bihar, identifies three distinct albeit inter-related classes of reasons behind the Naxalite movement viz. (1) Economic rights (2) Social rights (3) Political rights. The economic issues that Bhatia [2005] mentions are, land rights; minimum wages; common property resources; and housing. The Government of India in its own study recognizes land related factors, displacement and forced evictions, livelihood, and social oppression as the main socio economic reasons behind the discontent of people and support to Naxalism.[13] Guha [2007] on the other hand argues that the exclusion of the tribals from the development in India is crucial to the conflict.

As far as the participation is concerned, "... the social base of the movement ... consists overwhelmingly of the landless, small peasants with marginal landholdings, and to lesser extent middle peasants. In caste terms, the base of the movement consists of lower and intermediate castes" (Bhatia, 2005). The fight against the social oppression that the dalits [lower castes] and the lower among the OBCs [other backward castes] have been regularly subjected to is perhaps the most significant among the issues used by the Naxalite movement" (GOI, 2008). Moreover, there is a huge tribal participation in the movement. The Scheduled Tribes are in fact economically one of the worst performing

[13]See Government of India [2008]. Apart from the socio economic reasons issues arising out of bad governance and policing are also mentioned.

groups in India (even behind the Scheduled Castes) and exclusion of the tribes from the growth that the Indian mainstream has been experiencing is seen as one of the key driving forces of the Maoist movement.[14]

1.3 The Main Hypotheses

The literature speaks of a variety of different factors that might lead to civil conflicts. "Civil Wars are more likely to occur in countries that are poor, are subject to negative income shocks, have weak state institutions, have sparsely populated peripheral regions and possess mountainous terrain" [Blattman and Miguel, 2010]. Some of the other factors that have been mentioned in the literature are, ethnic and religious diversity and fragmentation, lack of democracy and civil liberties, linguistic and religious discrimination, inequality, new states and political instability, non-contiguous territory, population pressure, colonial occupation, etc.[15]

In this study we try to identify which factors might be more relevant in the context of the Naxalite conflict in India. Thus, combining the findings from the previous literature and our understanding of the conflict we test for five different hypotheses.

1.3.1 Hypothesis 1: Land inequality increases conflict.

Land rights are one of the most important issues taken up by the Naxalites: "*Khet par adhikar ke liye ladho, desh mai janwad ke liye badho*" (Fight for land rights, march towards democracy in the country - Liberation (a Naxalite group) slogan) [Bhatia, 2005]. In fact, regardless of the ideologies of the different Maoist factions in practice land redistribution has remained one of the main goals of the movement. This is evident from the failed peace talks between the Andhra Pradesh government and the PWG in 2004 where this was one of the main issues [Iyer, 2009].[16] There is empirical evidence on the importance of land distribution on conflicts from elsewhere in the world as well [Andre and Platteau, 1996, Macours, 2011, Verwimp, 2003].

There are several potential ways in which the land distribution could affect conflict outcomes. A highly skewed land distribution reflects higher disparities in the social and economic lives of the people and thus a higher potential for grievance. Moreover,

[14] See Guha [2007]

[15] See Fearon and Laitin [2003] for a discussion.

[16] In a recent hostage incident when a government collector was kidnapped, two of the primary demands of the Maoists were as follows: grant of land rights to tribal people in scheduled areas; minimum displacement of tribals while making space for industries and mining (The Hindu, Feb 22, 2011: Orissa accepts eight Maoist demands: http://www.hindu.com/2011/02/22/stories/2011022264251300.htm).

if the distribution is too unequal and dominated by a few large landlords there is an additional source of problem. In case of land acquisition by the government while the entire community is adversely affected the compensation mostly or fully goes to the big landlords who to start with enjoy a higher socio-economic status, while the vast majority of the affected are left uncompensated. Moreover, it is also common that tribal people in remote places have been living on a certain plot of land for generations but do not have any formal title deeds to the land resulting in no compensation whatsoever in case of acquisition by the government. Thus, the Naxalites are fighting for land issues against both big landlords and the state.

The land distribution is crucial not only for the Maoist conflict but for the overall health of the rural economy. The land reforms in India in the post-independence period has been associated with significant poverty reduction [Besley and Burgess, 2000]. However, implementation was hardly homogeneous across states. These heterogeneities exist even within states which have witnessed widespread implementation. "The 3 extreme Maoist districts in West Bengal are West Midnapur, Puruliya, Bankura-where land reforms on the scale affected by the CPI-(M) in other parts of the state haven't taken place." [Chakravarti, 2008] Thus, we will use a measure of land inequality at the district level and try to verify if indeed more land inequality is associated with more conflict.

1.3.2 Hypothesis 2: Underdevelopment leads to more conflict.

The relation between income and civil conflict is one of the most robust relations in the empirical literature on conflicts (Collier and Hoeffler, 1998, 2004 and Fearon and Laitin, 2003). Collier and Hoeffler [2004] argue that the opportunity cost of fighting is lower when incomes are low and thus making it easier to recruit rebels. Fearon and Laitin [2003] on the other hand argues that lower incomes reflect limited state capacity to put the rebellion down. The Maoist conflict is also concentrated in some of the most impoverished regions of the country. "Even if there was no truth behind the accusations that poor tribes are being exploited and are being illegally disposed of their lands, it is fairly certain that the Naxalites are feeding on the festering discontent of the impoverished and marginalized tribal communities. According to the 2001 census for example, about three quarters of Dantewada's 1,220 villages are almost wholly tribal; 1,161 have no medical facilities; 214 have no primary school; the literacy rate is 29% for men and 14% for women." (The Economist, 2006).

Apart from the levels of income, negative shocks to income have also been found to be important. The theoretical idea goes back to Becker [1968] where he argues that rise in returns to crime induces more workers to the criminal sector. Miguel et al. [2004]

demonstrate how negative income shocks lead to higher levels of conflict. Dube and Vargas [2008] argue that lower wages increase conflict (lower international coffee prices lead to a negative shock to certain regions of Colombia leading to higher conflict levels).[17]

Thus, a priori both income levels and changes in incomes could potentially affect the conflict. We will explicitly check in our analysis whether districts with lower per capita incomes, and lower rates of growth indeed experience more conflict.

1.3.3 Hypothesis 3: The exclusion of disadvantaged groups (Scheduled Castes and Scheduled Tribes) leads to more conflict.

In India, the Scheduled Castes (SC) and the Scheduled Tribes (ST) are two groups that have been historically disadvantaged and find themselves at the bottom of the social hierarchy.[18] The Scheduled Castes are at the bottom of the Hindu caste system. The Scheduled Tribes on the other hand are tribal populations who are outside the Hindu caste system. As discussed in the previous section the main support and the recruits for the conflict are supposed to be mostly from the lower caste and tribal populations.[19] While these disadvantaged minorities have enjoyed affirmative action in the post independence India, they are still not economically and socially at par with others. "There are over 200,000 pending cases of atrocities against lower castes in India, and the conviction rate is just a little over 2 percent. There are an estimated 162 million untouchables in rural India, according to the National federation of Dalit Land rights movement. 70% of them don't own land" [Chakravarti, 2008]. Guha [2007] stresses that in the last 60 years after independence the tribals have been the group that has been left out and even suffered in the growth story of India. In fact, the predominantly tribal areas like Chhattisgarh have experienced very high levels of conflict.

Thus, caste and tribal identities are important issues that could exacerbate the conflict. It would be interesting to see once we control for incomes, if presence of lower castes and tribals still makes a difference. Moreover, heterogeneity in incomes across these ethnic lines could also be important. "...'[H]orizontal' inequality- inequality that coincides with ethnic or other politically salient cleavages is a particular important driver of civil conflict." (Blattman and Miguel [2010]) Thus, we try to verify whether grievances arising out of feelings of exclusion of the disadvantaged groups impact the conflict by controlling

[17] This is the opportunity cost channel. Dube and Vargas [2008] also find that a positive shock to the natural resource price increases conflict via the rapacity channel.

[18] "Historically disadvantaged groups are commonly defined as groups which have been systematically excluded from institutions and cultural practices that provide skills and resources."(Pande [2003])

[19] "The main support for the Naxalite movement comes from dalits (SC) and adivasis (ST) (Government of India [2008])."

for both the presence of different groups and differences in incomes and growth rates across groups.

1.3.4 Hypothesis 4: Historical land institutions directly impact the conflict.

There is a huge body of literature showing that historical institutions have a persistent effect on current economic performance.[20] In the context of civil conflicts, focussing on the Hindu-Muslim communal conflicts in India, Jha [2008] finds that, in the presence of a historical complementarity between the two groups the probability of conflict is lower.[21] Banerjee and Iyer [2005] have highlighted the importance of historical land institutions in the context of India. They have shown that districts in which property rights in land were historically given to landlords by the British have ended up with worse economic outcomes in the post-independence period.

We argue that the historical land institutions could affect the Maoist conflict, through their effects on economic outcomes, and directly through their effects on social relations among people. "Areas most associated with Maoist uprisings are WB [West Bengal], Bihar & Srikakulam district of AP [Andhra Pradesh]- all landlord areas. Paul R. Brass (1994 pg. 326-327) argues explicitly that these peasant movements have their roots in the history of exploitation and oppression of peasants by landlords."(Banerjee and Iyer [2005]) Also, the allocation of the responsibility of collecting land revenue to the landlords (which translates into a de facto property right over the land) gave birth to a reason for perpetual conflict between the peasants and landlords. "Elsewhere, the colonial state directly collected the land revenue from the cultivator, thereby avoiding this particular source of internecine conflict." (Banerjee and Somanathan [2007])[22]

Moreover, land acquisition for mining purposes, building dams, or for private industry and the resulting displacement and loss of livelihood of people is an important issue pursued by the Maoists. Such land acquisition often leads to large scale. Given adequate compensation people might be less discontent with the displacement induced by land acquisition. In this context, Duflo and Pande [2007] show how the landlord districts do worse than the non-landlord districts as far as effects of dams are concerned. They argue that since the social relation in the landlord districts somehow renders collective action

[20] Acemoglu et al. [2001, 2002], La Porta et al. [1998, 1999, 2000], Engerman and Sokoloff [1997, 2002]

[21] Besley and Persson [2008] highlight the importance of political institutions i.e. whether the country is a parliamentary democracy, or has a system of strong checks and balances, in explaining conflict outcomes.

[22] The effect of these institutions on land inequality is unlikely to be a problem since, " districts with worse land distribution historically have also seen more land reforms in the post-independence period. This makes it unlikely that the persistence of the landlord effect is mainly through its effect on contemporaneous land distribution." Banerjee and Iyer [2005])

difficult it leads to inadequate compensation. Following the same argument, if land is taken away in the landlord districts for industrial, mining or the purposes of building dams, historically people in such districts have less potential for collective action leading to inadequate compensation, which in turn leads to more grievances. This makes for easier Maoist recruits.

Thus, there are reasons to believe that the land institutions have an effect on the conflict over and above the effect through the contemporaneous land inequality and underdevelopment. In this paper we empirically verify if these institutions indeed have an impact on the conflict.

1.3.5 Hypothesis 5: Suitable geographic conditions impact the conflict.

Guerrilla warfare is usually concentrated in remote rural areas where the rebels get suitable conditions to pursue such insurgencies. The favourable conditions that such regions provide include superior knowledge of local conditions and territory by rebels than by government forces, rough mountainous territory or forests providing suitable hideouts to insurgents etc. Even for nations with strong military capabilities such conditions could turn out to be daunting (e.g. the US in Vietnam, British in the initial days in Northern Ireland),[23] for poorer nations with weak military capabilities and corrupt bureaucratic set ups it could be hopeless.

Geography determines the cost, time and tactics necessary to curtail and check such activity. Fearon and Laitin [2003] for example find that conditions favouring insurgency like rough terrain increases the likelihood of conflict. We thus control for the geographical factors like percentage of forests, steep sloping land, sandy land, barren rocky land etc. in the district.

Apart from the variables mentioned in the above hypotheses we control for a host of other variables that the literature has found to be important. We control for income inequality and size variables like population density, area etc.[24]

1.4 Empirical Analysis

Districts are the smallest administrative units in India for which coherent reliable data is available. The idea of the study is to exploit the sub-national micro structure of the

[23]Examples from Fearon and Laitin [2003]

[24]Population might either indicate more pressure on existing resources like land or more potential recruits.

data and thus all regressions are at the district level. We use data from the 16 main states in India [25] which covers about 90% of the Indian population. The data is pooled from diverse sources and a complete description is provided in the data appendix.

We have annual data on the conflict variables for the period 1979-2009. However, more than 90% of the data is concentrated in the years 2000 onwards and 70% in the years 2005 onwards. We divide the data into three distinct time periods due to the nature of our income data.

District-specific indicators of income or expenditure are not available for India. Instead, per capita consumption expenditure is calculated from the National Sample Survey (hereafter NSS) data. These data are not only nationally representative but also representative at the district level. We thus use mean per capita consumption expenditure at the district level from the NSS as our proxy for income at the district level. Since we have only three rounds of NSS data viz. 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st), we club the conflict data to match these three different NSS periods. The district names are first mapped to the districts that existed in 1987. Then all the conflict data from the period 1988-1999 are collapsed and clubbed together and matched to the 1987-88 NSS data, all the conflict data from the period 2000-2004 are clubbed together and matched to the 1999-2000 NSS data, all the conflict data from the period 2005-2009 are clubbed together and matched to the 2004-2005 NSS data. Since the conflict is concentrated completely in the rural areas only rural data from the NSS is used. Thus we have data on around 360 districts for three time periods.[26]

There are primarily two variables of interest that we try to explain viz. 'Probability of Conflict' and 'Intensity of Conflict'. In order to explain the probability of conflict we create a 0-1 binary outcome variable that takes the value '1' if a district has seen any Maoist activity in the relevant period or the value '0' otherwise. While it is extremely important to understand which factors increase conflict probabilities it is equally important to understand what factors lead to higher intensity of conflict. Intensity is a latent variable that is measured using two variables viz. "Total number of incidents" in a particular district and "Total number of deaths + wounded" in the district in the relevant periods. The econometric specification is straightforward:

$$(Conflict)_{j,t} = \alpha(Conflict)_{j,t-1} + \beta X_{j,t-1} + \gamma G_j + \alpha_s + \delta_t + \epsilon_{j,t} \tag{1.1}$$

[25] Districts of Chattisgarh, Jharkhand and Uttaranchal are considered to be part of Madhya Pradesh, Bihar and Uttar Pradesh states respectively, i.e. their old state before the creation of the new ones.

[26] While running the regressions we sometimes have lesser number of districts due to the unavailability of data on certain covariates for some or all periods for certain districts.

The $(Conflict)_{j,t}$ variable gives the conflict (presence or intensity) in district "j" in round "t". The explanatory variables $X_{j,t}$ include economic variables like the mean per capita consumption expenditure (MPC) and the gini coefficient of income inequality. "G" includes all variables that do not change over time: land inequality, demographic variables, presence of marginalized factions like Scheduled Tribes and Castes and geographic variables like barren and rocky land, steep sloping land, percentage forests etc. "α_s" is the state dummy, while "δ_t" is the time dummy.[27]

For the 0-1 probability of conflict variable we use Probit specifications. The intensity variables are both count variables by nature taking integer values from zero upwards. The Poisson model is the standard model used in such cases. However, the Poisson model assumes the mean and variance to be equal. In the presence of many zeros in the data like in our case there is over-dispersion and thus the equi-dispersion assumption of the Poisson model does not hold true. The number of zero counts in our data is more than 70%. The standard parametric model to account for over-dispersion is the Negative Binomial model (Cameron and Trivedi, 2005). Thus, we use the Negative Binomial model (Poisson is a special case of it) for explaining intensity. All the regressions use cluster robust standard errors, clustered at the state level.

Our identification strategy relies on using data on the explanatory variables from the pre-conflict period. In particular, we use data on consumption expenditure (our proxy for income) from the beginning of the period. Thus, there is no serious risk of endogeneity arising out of reverse causality. Further, we always control for past conflict levels in all the specifications. However, there is still potential for endogeneity arising out of omitted variables. We thus try to control for all potential variables that might be important in our specification. However, in order to be completely sure we also run the specifications using an instrumental variables approach using lagged consumption expenditure as an instrument for current consumption expenditure.

1.4.1 Benchmark

Table 1.1 is our benchmark table. In columns 1-3, we try to understand what variables explain the presence of conflict using Probit regressions. In Columns 4-6, we run Negative Binomial regressions to explain the intensity of conflict. The dependent variable in these columns is the number of dead and wounded people in the district.

[27]The summary statistics and the correlation matrix for all the variables used in the analysis are provided in the appendix C in tables C.1 and C.2

In column 1, which is our baseline specification to explain the probability of conflict, we see that lower incomes and higher land inequality both significantly increase the probability of conflict. A higher Scheduled Caste population does not affect the probability of conflict. On the other hand, a higher tribal population significantly increases the probability of conflict. However, this relationship is not monotonic. This is evident from the fact that the square term of the proportion Scheduled Tribe variable is negatively significant.

Table 1.1: The determinants of the Maoist Conflict in India

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
consumption per capita	-1.323***	-1.434***	-1.561**	-3.014**	-1.530	-4.239**
	(0.425)	(0.528)	(0.734)	(1.478)	(1.272)	(1.922)
land inequality	3.617***	3.566***	4.435***	15.01***	13.52***	14.68***
	(0.781)	(1.202)	(0.899)	(2.934)	(3.083)	(3.647)
proportion sandy	-1.524	-0.603	0.332	-31.58	-42.51**	-10.25
	(2.645)	(2.687)	(1.393)	(21.01)	(21.55)	(8.838)
log state capital distance	-0.0823	0.0521	-0.0201	-0.795***	-0.472**	-0.303
	(0.147)	(0.146)	(0.178)	(0.216)	(0.190)	(0.219)
proportion barrenrocky	10.50***	15.93***	13.12***	40.97**	47.68***	26.93***
	(2.734)	(3.510)	(4.414)	(16.73)	(16.70)	(4.656)
proportion steepsloping	-67.35***	-86.49***	-126.6***	-270.3***	-242.1***	-346.3***
	(21.16)	(32.92)	(32.94)	(36.59)	(56.58)	(72.58)
proportion forest cover	0.889	1.568**	2.397***	2.669	2.965	6.060***
	(0.708)	(0.775)	(0.597)	(2.040)	(2.075)	(1.757)
%Scheduled Castes	-0.629	0.0471	2.922**	-4.485	-5.038*	7.018*
	(1.349)	(1.517)	(1.291)	(4.095)	(2.794)	(4.046)
%Scheduled Tribes	4.383***	4.229**	2.882***	10.04**	10.01***	6.231***
	(1.651)	(1.872)	(0.677)	(4.432)	(3.621)	(2.243)
%Scheduled Tribes square	-5.630**	-4.826**	-2.062	-12.15**	-13.49***	-3.831
	(2.805)	(2.309)	(1.361)	(6.174)	(4.954)	(2.832)
population density	0.293	0.618*	0.329	-0.150	-0.00329	-0.0382
	(0.230)	(0.329)	(0.297)	(0.562)	(0.515)	(0.535)
log area	0.671***	0.749***	0.681***	2.416***	2.380***	2.388***
	(0.216)	(0.275)	(0.142)	(0.408)	(0.394)	(0.396)
income inequality	-0.131	-3.545*	1.658	-7.344	-14.11**	-1.936
	(2.057)	(1.849)	(1.500)	(5.025)	(5.544)	(3.095)
initial consumption per capita		-0.491			-3.749***	
		(0.615)			(0.778)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	996	656	996	996	656	996
ll	-269.1	-181.9	-192.9	-1013.1	-829.0	-913.4

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The marginal effects are provided in Table A.10. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

In column 2 we consider data from the latest 2 rounds instead of all 3 rounds and control for the initial period income using data from the first round. In column 3 we control for the state dummies in the baseline specification of column 1. Our results are robust to the different specifications.

Column 4, which is our baseline specification to explain the intensity of conflict, shows that lower incomes and higher land inequality significantly increase the number of dead and wounded people. Moreover, the presence of tribals also has a significant effect.

Like in column 2, in column 5 we consider data from the latest 2 rounds instead of all 3 and control for the initial period income using data from the first round. In column 6 we control for the state dummies in the baseline specification of column 4. The results remain unaltered.

From the above it is clear that, poorer regions experience more conflict. This is in line with the previous literature and is in fact one of the most robust results in the conflict literature. In terms of average marginal effects, a 1 standard deviation increase in the log of mean per capita consumption expenditure increases the probability of conflict in the next period by 5%, and results in 50 more dead and wounded people in a year.

On the other hand, while income inequality does not seem to play any role in the context of the Maoist conflict, land inequality is a highly significant and robust predictor of conflict in all the specifications. Land inequality also significantly increases the intensity of conflict and this is robust to different controls. This gives support to the grievance arising out of land inequities being a key reason behind the conflict. In terms of average marginal effects if we look at our baseline specification, a 1 standard deviation increase in land inequality increases the probability of conflict in the next period by 9%, and results in 157 more dead and wounded people a year.

As mentioned above the percentage of Scheduled Tribes in the district has a non monotonic impact on conflict. In Figure 1, we notice that the marginal effect increases till about 40% tribal presence in the district and then it starts falling. At around 80% tribal presence the marginal effect actually becomes negative. We see that districts which have around 40% more Scheduled Tribes face around 12% higher probability of conflict and around 120 more dead and wounded people in a year.

As far as the geography variables are concerned, they turn out to be quite important. We see that barren and rocky areas experience more conflict. This is in agreement with our expectations. In a primarily agrarian economy, a higher proportion of barren and rocky land indicates lack of economic opportunity. This supports the opportunity cost story.

The percentage of forest cover variable comes out to be significantly increasing the probability and intensity of conflict in many of the specifications (but not all). One

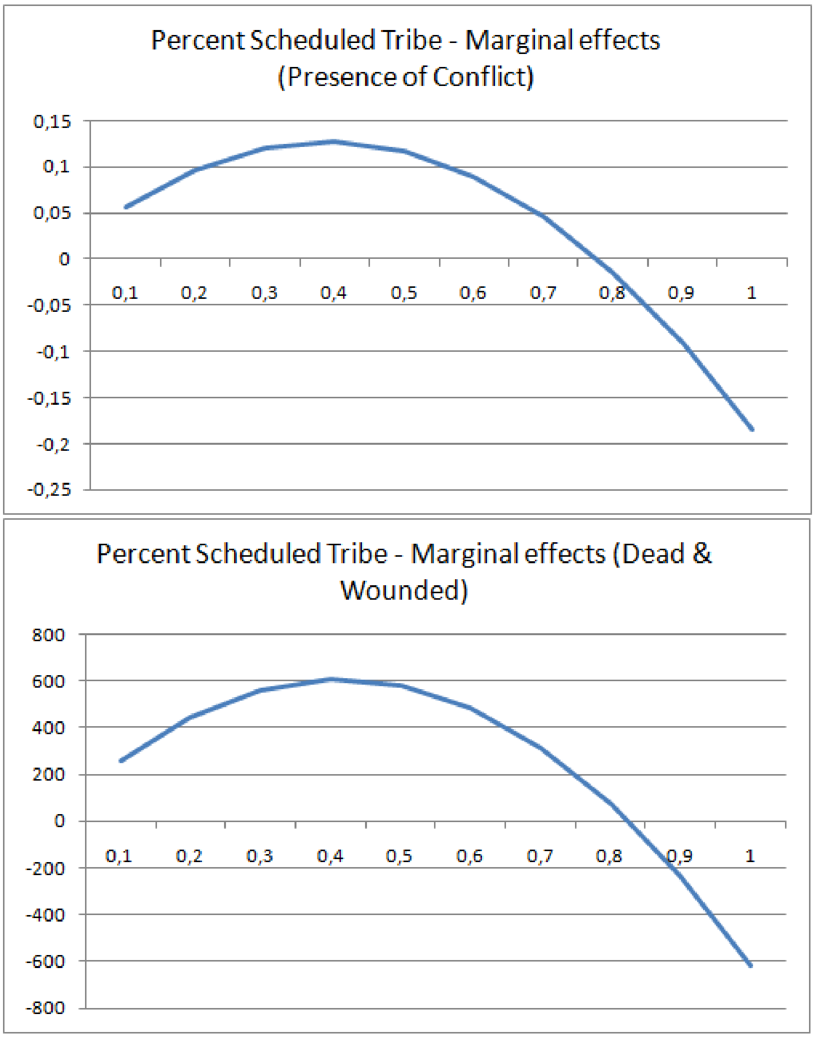

Figure 1.1: Effects of Scheduled Tribe presence

explanation could be that large patches of forest cover provides the perfect hiding and fighting conditions for rebels and makes it quite difficult for security forces to keep up with them. Also, we see that districts with higher geographical areas experience both a higher probability and intensity of conflict.

The percentage steep sloping land variable on the other hand comes out to be negatively significant. This stands in contrast to the previous literature which has found mountainous land as being more suitable for conflict. In the context of India, the proportion of the district which is steep sloping probably proxies some other variable like the proportion of land (not) under agriculture or the like and thus it has a negative sign. However, due to data constraints we are unable to point out what exactly this variable is picking up.

In Table 1.2 we add a control for historical institutions to the otherwise identical specifications of Table 1.1. The relevant variable gives the proportion of the district that was not under the control of the landlords. The historical institutions variable has a significant and quite robust impact on the conflict. However, consumption per capita becomes less robust while land inequality on the other hand continues to be significant and robust. The presence of Scheduled Tribes does not seem to matter anymore. However, by adding the control for historical institutions we lose a third of the observations due to the unavailability of data on the institutions variable for all districts.[28] Thus, sample change might be the cause of the change in significance of the other variables. In fact, this is what seems to be the reason behind the Scheduled Tribe and the consumption per capita variables turning insignificant in Table 1.2.[29]

In Table 1.3, we instrument current consumption expenditure with its own lagged values. Otherwise, we have the identical specifications of Table 1.1. We use two stage instrumental variable regressions. In the first stage we notice that the instrument is highly significant.

In the second stage we notice a few changes from Table 1. We find that consumption expenditure per capita and land inequality continue to be significant in explaining the probability and the intensity of conflict. However, we notice that the presence of Scheduled Tribes is no longer significant. We also report the coefficients of the proportion Scheduled Tribe variable in the first stage. We see that the proportion Scheduled Tribe significantly reduces the per capita consumption expenditure. Thus, in the second stage once we control for this effect, the proportion Scheduled Tribe variable does not matter anymore.

[28]Data on the land institutions variable is available only for the districts which were directly under British control.

[29]See Table A.3 in appendix C where we have the same sample as in Table 1.2.

Table 1.2: The role of historical institutions

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
consumption per capita	-0.862*	-0.955*	-0.942	-1.724	-1.378	-3.115
	(0.470)	(0.556)	(0.891)	(1.294)	(1.131)	(2.809)
land inequality	2.438***	2.170**	3.917***	11.20***	9.867***	13.36***
	(0.546)	(0.852)	(0.901)	(2.189)	(2.359)	(2.998)
proportion Non landlord	-0.935***	-1.041***	-1.188**	-2.496***	-2.088***	-1.402
	(0.276)	(0.249)	(0.476)	(0.679)	(0.609)	(0.859)
proportion sandy	17.66**	20.83***	32.15**	44.76	-17.15	42.28
	(7.832)	(6.724)	(15.35)	(78.21)	(38.28)	(55.40)
log state capital distance	0.190	0.243	0.163	-0.0216	0.134	-0.127
	(0.195)	(0.219)	(0.208)	(0.470)	(0.352)	(0.393)
proportion barrenrocky	19.02***	26.28***	22.95***	54.94**	51.73***	39.80***
	(3.395)	(4.609)	(6.081)	(25.72)	(15.01)	(9.432)
proportion steepsloping	-53.55*	-95.53***	-132.3***	-207.3***	-331.6***	-366.8***
	(27.34)	(35.33)	(43.57)	(74.18)	(116.1)	(101.9)
proportion forest cover	1.543**	2.109***	2.727***	5.319***	5.458***	7.333***
	(0.692)	(0.695)	(0.622)	(1.922)	(1.842)	(1.904)
%Scheduled Castes	-0.969	-0.592	4.252***	-4.802	-6.595	10.62***
	(1.398)	(1.520)	(1.310)	(4.316)	(4.504)	(3.125)
%Scheduled Tribes	2.749	4.273*	1.438	-3.397	-1.640	-0.123
	(2.542)	(2.596)	(1.160)	(6.369)	(7.405)	(3.261)
%Scheduled Tribes square	-5.565	-8.690**	-1.199	-2.575	-8.935	2.205
	(4.337)	(4.050)	(2.020)	(10.20)	(12.91)	(4.054)
population density	0.484**	0.762**	0.413	0.199	0.0307	0.134
	(0.203)	(0.341)	(0.317)	(0.597)	(0.651)	(0.642)
log area	0.846***	1.008***	0.902***	3.316***	3.029***	2.782***
	(0.200)	(0.317)	(0.138)	(0.587)	(0.498)	(0.326)
income inequality	-2.807	-5.845***	-1.116	-15.47***	-20.54***	-3.583
	(1.865)	(1.724)	(1.937)	(5.244)	(6.346)	(6.438)
initial consumption per capita		0.242			-2.315***	
		(0.811)			(0.765)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	655	431	655	655	431	655
ll	-177.9	-131.6	-138.2	-642.2	-535.5	-590.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table is identical to Table 1.1 but also controls for historical institutions. The table reports coefficients and not marginal effects. The marginal effects are provided in Table A.11. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table 1.3: The determinants of the Maoist Conflict in India - IV estimation

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
consumption per capita	-2.642***	-3.033***	-4.556***	-6.971***	-4.813**	-10.34***
	(0.857)	(1.164)	(1.160)	(1.651)	(1.988)	(3.006)
land inequality	3.171***	3.183**	3.768***	12.78***	13.48***	7.892***
	(1.103)	(1.367)	(0.992)	(3.440)	(3.620)	(2.894)
proportion sandy	0.516	0.443	0.397	-25.68*	-35.76**	-12.84
	(2.530)	(2.926)	(1.567)	(15.35)	(16.76)	(11.73)
log state capital distance	0.0129	0.0316	0.0567	-0.616***	-0.546***	-0.123
	(0.153)	(0.141)	(0.158)	(0.180)	(0.177)	(0.220)
proportion barrenrocky	15.11***	16.40***	14.98**	49.91***	52.07***	32.64***
	(3.832)	(3.592)	(6.119)	(14.19)	(14.49)	(8.903)
proportion steepsloping	-48.50**	-81.30**	-91.21**	-174.1***	-232.2***	-232.2***
	(22.01)	(33.72)	(39.60)	(25.66)	(62.55)	(62.66)
proportion forest cover	1.642*	1.597**	2.872***	3.855*	3.562	7.583***
	(0.855)	(0.804)	(0.854)	(2.217)	(2.180)	(1.974)
%Scheduled Castes	-0.249	-0.430	2.022*	-6.754**	-6.538**	0.573
	(1.635)	(1.677)	(1.084)	(2.914)	(2.847)	(3.741)
%Scheduled Tribes	2.809	2.486	-0.766	2.963	5.844	-1.733
	(2.196)	(2.293)	(2.136)	(2.874)	(3.693)	(2.291)
%Scheduled Tribes sq	-3.444	-3.043	1.693	-5.680	-9.201**	1.831
	(2.819)	(2.816)	(2.703)	(3.765)	(4.620)	(2.841)
population density	0.577	0.535	0.258	-0.357	-0.303	0.315
	(0.356)	(0.371)	(0.382)	(0.628)	(0.639)	(0.711)
log area	0.628**	0.608**	0.464**	1.793***	2.112***	1.816***
	(0.311)	(0.297)	(0.207)	(0.358)	(0.343)	(0.415)
income inequality	-1.296	-0.516	5.412**	-6.500	-10.22*	7.342
	(2.200)	(2.595)	(2.418)	(5.021)	(5.981)	(5.337)
initial consumption per capita	No	Yes	No	No	Yes	No
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
First Stage						
consumption per capita lagged	**0.602*** **	**0.461*** **	**0.359*** **	**0.601*** **	**0.461*** **	**0.358*** **
	(0.0642)	(0.0608)	(0.0520)	(0.065)	(0.065)	(0.052)
%Scheduled Tribes	-0.679***	-0.674***	-0.684***	-0.679***	-0.678***	-0.680***
	(0.166)	(0.170)	(0.136)	(0.168)	(0.171)	(0.140)
%Scheduled Tribes square	0.636***	0.656***	0.678***	0.636***	0.665***	0.669***
	(0.158)	(0.165)	(0.174)	(0.166)	(0.170)	(0.180)
Observations	640	638	640	640	638	640
ll	84.31	91.60	236.0	-834.1	-826.5	-762.2

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using IV-Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using OLS regressions in the first stage and Negative Binomial regressions in the second stage. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds. Previous period consumption expenditure instruments following period consumption expenditure.

1.4.2 Testing for the "Exclusion" hypothesis

In this section we check specifically for the "Exclusion" story. In other words we try to identify if horizontal inequalities in incomes or growth i.e. differences in incomes or the growth of incomes across different ethnic groups have any significant impact on the conflict.

TABLE 1.4: The growth in income of different ethnic groups

	(1) Probability	(2) Probability	(3) Intensity	(4) Intensity	(5) Intensity	(6) Intensity
General Castes growth	-0.0263**	-0.0325**	0.00549	-0.00239	0.00901	-0.00426
	(0.0129)	(0.0149)	(0.0168)	(0.0160)	(0.00972)	(0.00835)
Scheduled Castes growth	-0.0142	0.00539	0.0377	0.0386*	0.0142	0.0244
	(0.0137)	(0.0105)	(0.0247)	(0.0227)	(0.0121)	(0.0155)
Scheduled Tribes growth	-0.0156	-0.0201	**-0.0930*** **	**-0.0695***	**-0.0595*** **	**-0.0438** **
	(0.0187)	(0.0276)	(0.0290)	(0.0355)	(0.0231)	(0.0220)
proportion sandy	-4.741	-0.0125	-84.78***	-39.39**	-23.26**	-13.07***
	(5.887)	(2.860)	(32.08)	(17.89)	(10.71)	(4.514)
log state capital distance	0.118	0.129	-0.347	0.0230	-0.303	0.00305
	(0.104)	(0.211)	(0.220)	(0.323)	(0.227)	(0.263)
proportion barrenrocky	15.97***	21.55***	30.53	27.05***	21.27	21.06**
	(5.749)	(8.341)	(26.10)	(9.680)	(14.81)	(9.082)
proportion steepsloping	-185.2***	-251.8***	-442.7***	-352.4***	-241.6***	-198.3***
	(53.76)	(95.98)	(129.8)	(130.7)	(72.69)	(74.77)
proportion forest cover	2.461***	4.882***	-1.162	4.502	1.365	5.412***
	(0.788)	(1.418)	(1.552)	(2.788)	(1.539)	(1.923)
%Scheduled Castes	0.950	3.568	-7.556*	1.666	-1.812	3.120
	(1.353)	(2.990)	(3.952)	(5.108)	(3.485)	(4.505)
%Scheduled Tribes	1.548	0.638	8.996	6.967***	4.568	2.847
	(2.319)	(1.816)	(5.636)	(2.675)	(3.725)	(1.978)
%Scheduled Tribes square	-1.006	0.231	-12.87*	-8.309***	-5.967	-1.821
	(2.683)	(2.076)	(6.687)	(2.856)	(4.325)	(2.252)
population density	0.192	0.264	-0.701	0.0926	-0.0870	0.663
	(0.438)	(0.490)	(0.749)	(0.921)	(0.628)	(0.710)
log area	1.242***	0.996***	3.705***	3.152***	2.657***	2.370***
	(0.319)	(0.255)	(0.478)	(0.542)	(0.619)	(0.500)
income inequality	-5.516**	-6.117***	-15.73***	-14.37***	-11.84***	-9.873**
	(2.189)	(1.658)	(4.707)	(5.475)	(3.053)	(4.345)
land inequality	7.679***	7.624***	19.09***	12.58***	12.09***	7.446***
	(1.436)	(0.894)	(3.899)	(4.120)	(2.270)	(2.569)
initial consumption per capita	No	Yes	No	Yes	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	No	Yes	No	Yes	No	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	290	284	290	290	290	290
ll	-74.40	-57.41	-544.7	-511.9	-493.8	-453.0

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The marginal effects are provided in Table A.12. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1 and 2 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 3, 4, 5 and 6 explain the intensity of conflict (columns 3 and 4 - the no. of dead & wounded in the district; columns 5 and 6 - the number of incidents in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The growth rates correspond to the 2 latter rounds. The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

In Table 1.4, we study the impact of differences in growth rates of the different groups. We see that as far as the presence of conflict is concerned (i.e. columns 1 and 2), it is the growth rate of the incomes of the General Castes that matter but the marginal effects are small and are similar in magnitude to the effects of growth rates of other ethnic groups. A 1 standard deviation increase in the growth rate for the General Castes leads to a 4% lower probability of conflict while a similar increase for the Scheduled Castes and Tribes is not significant in the baseline specification of column 1.

For the specifications explaining intensity (columns 3-6), the growth rates of the General Castes is not significant and even changes its sign in some of the specifications. On the other hand a growth in incomes of the Scheduled Tribes significantly and robustly reduces the intensity of conflict in all the specifications.[30] In terms of marginal effects, a 1 standard deviation fall in the income growth of the Scheduled Tribes leads to an increase in the number of dead and wounded people in the district by around 155 people in the next period. This means that some of the participation in the conflict by the tribals and/or support for the Maoists among them might actually be driven by the lower growth in incomes of the Scheduled Tribes as they feel excluded from the growth that the rest of society is facing.[31] The rate of growth of incomes of the Scheduled Castes comes out to be insignificant.

Looking at the levels of income Table A.4, and Table A.5, we do not find any robust results. The lower castes and the tribes are ethnically two very different groups of people. However, both these groups are seriously disadvantaged socially and economically. In order to be able to use more observations, we club the Scheduled Castes and Tribes together and treat them as one group (See Table A.6). We find that the incomes of both groups matter. However, the coefficient for the incomes of the Scheduled Castes and Tribes is almost always greater than that of General Castes. Moreover, in almost all the specifications we have percentage of Scheduled Castes and Tribes in the district, significantly increasing the conflict presence and intensity. While this is not conclusive, it does give some confirmation of the participation from the lower castes and tribes.

One limitation with this analysis is that the number of Scheduled Tribal people surveyed is low in many of the districts owing to the low percentage of Scheduled Tribes living in those districts. We include only districts which have at least a weighted population of 15 tribal people surveyed.[32] This significantly reduces our sample size. However, the comparison of incomes across ethnic groups is suitable only in districts which have at

[30]In columns 3 and 4 the dependent variable is the number of dead and wounded people and in columns 5 and 6, it is the number of incidents in the district.

[31]The overall growth rate of income in the district does not have any significant impact on the conflict. See Table A.7 and Table A.8

[32]Including districts which have even fewer or more Scheduled Tribal people surveyed makes no significant difference in the results.

least some residents of each of the groups we would like to compare, actually residing in those districts. Thus, we proceed with our analysis with the reduced sample size.

1.4.3 Temporal variation in the conflict

In the previous tables we have found some interesting insights about the causes of the Naxalite conflict in India. However, it will be interesting to check whether there is a change in the nature and causes of the conflict over time. In Table 1.5, we thus regress the conflict variables over the 3 rounds separately. This table gives us some more interesting insights. Columns 1-3 show the change in causes of the presence of conflict over the three rounds, while columns 4-6 show the corresponding change in the causes of intensity of conflict. We see that land inequality is important across all periods. But its relative importance varies over time, particularly for the intensity of the conflict. We see that to start with land inequality is a very important factor but over time the magnitude of the coefficient falls, suggesting that its relative importance is declining over time.

On the other hand we see that to start with Consumption expenditure is not that important, but in the first half of decade of 2000s (round 2) the conflict shifts to poorer areas. And finally we see that the conflict actually in the last five years (round 3) moves to areas where there are more tribal people. However, as in Table 1, this relationship is non-monotonic. We also see how in the first round the proportion of barren and rocky land does not matter but it does in the next rounds.

Interestingly we see that initially having a higher proportion of tribals in the district does not matter. In rounds 1 and 2 the proportion of Scheduled Tribes in the district is not significant. However, in round 3 we notice that having a higher proportion of tribals in the district significantly increases the conflict. This perhaps goes to show that in recent times the conflict has been spreading to tribal areas.

We also try to split the incomes of the separate groups and try to identify the temporal differences in their effects. We do not find any significant results. This is perhaps driven by the fact that we lose a lot of observations in the rounds 2 and 3 which are in fact the rounds where most of the conflict is concentrated.[33]

[33] See Table A.9

TABLE 1.5: Change in causes of conflict over time

	Round 1 Probability	Round 2 Probability	Round 3 Probability	Round 1 Intensity	Round 2 Intensity	Round 3 Intensity
consumption per capita	0.698	-3.436***	-0.990	0.875	-6.042***	-2.333
	(0.663)	(0.590)	(0.673)	(3.301)	(1.602)	(1.509)
land inequality	4.796***	3.726***	3.605**	20.19***	18.29***	10.32***
	(1.290)	(1.221)	(1.447)	(2.987)	(4.070)	(2.835)
%Scheduled Tribes	3.219	-0.652	8.317***	10.16	7.630*	10.66**
	(3.656)	(2.438)	(2.126)	(11.80)	(4.375)	(5.202)
%Scheduled Tribes square	-5.395	0.930	-10.31***	-16.05	-7.126	-17.65**
	(6.475)	(3.454)	(2.776)	(14.10)	(5.571)	(7.297)
%Scheduled Castes	-2.400	-3.382**	1.916	-5.941	-4.721	-4.695
	(2.472)	(1.601)	(1.923)	(7.436)	(3.660)	(3.622)
proportion sandy	-27.77	0.125	0.547	-98.65*	-150.7***	-12.81
	(19.41)	(2.847)	(2.561)	(59.65)	(38.47)	(17.52)
proportion barrenrocky	-0.203	22.24***	10.42*	56.20	32.23**	41.43**
	(4.344)	(4.492)	(5.641)	(44.11)	(14.71)	(16.58)
proportion steepsloping	-139.3***	-134.4***	-36.40*	-829.2***	-202.2***	-173.5***
	(48.60)	(40.72)	(19.22)	(230.3)	(44.76)	(30.76)
log state capital distance	-0.330*	0.178	-0.0338	-0.748*	-0.690***	-0.376*
	(0.176)	(0.157)	(0.257)	(0.384)	(0.261)	(0.222)
proportion forest cover	-0.172	1.866*	1.368	3.163	-0.458	3.479
	(1.315)	(1.087)	(1.125)	(2.656)	(1.138)	(2.500)
log area	1.208***	0.927***	0.574**	6.459***	2.569***	1.773***
	(0.335)	(0.356)	(0.276)	(1.085)	(0.480)	(0.426)
population density	-0.316	0.679*	0.623	1.239	-1.219	0.542
	(0.474)	(0.391)	(0.421)	(1.168)	(0.831)	(0.546)
income inequality	5.712*	-1.404	-4.271**	0.758	-15.05	-9.497*
	(3.365)	(1.258)	(2.118)	(10.35)	(9.751)	(5.252)
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	336	321	339	336	321	339
ll	-61.32	-85.11	-88.77	-147.0	-306.9	-522.6

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

1.5 Conclusion

This paper studies the political economy of the Maoist conflict in India in-depth using a district level panel. It also contributes to the civil conflict literature by adding to a growing number of studies that use sub-national micro data to study civil conflicts. Making use of a newly constructed district level conflict database the paper provides some interesting insights on the causes of the Maoist conflict in India.

The main finding of the paper is that the story behind the Maoist Conflict in India is a story of grievances arising out of feelings of exclusion of various forms. We see how the underdeveloped districts and districts with higher land inequality are more prone to conflict. Moreover, we have evidence in favour of participation from the Scheduled Tribes in the conflict. We see how low growth rates in the incomes of Scheduled Tribes leads to more conflict. Also, we see how social divisions created by historical institutions play a role in the conflict. All these findings indicate that certain sections of society feel left behind in the growth story of India leading to more grievances and social tensions which help Maoists by creating more sympathizers and boosting their recruitment efforts.

There is a huge scope for future research on the Maoist Conflict in India. One interesting exercise would be controlling for land reforms at the district level and verifying its effects on the conflict outcomes. Also, establishing causal relations between the different variables of interest and the conflict outcomes would be the other avenue of future research. Finding adequate instruments for the income and land inequality variables would allow us to further ensure that our results are not affected by endogeneity and thus establish causality. Moreover, a lot of work remains to be done in terms of data collection from the households of the perpetrators and victims in order to further pin down both the causes and consequences of the Maoist conflict at the household/individual level.

Chapter 2

Religious diversity, Intolerance and Civil conflict

2.1 Introduction

Does religious diversity affect the probability of civil conflict? If we are to take seriously the popular perception supported by the views of political scientists like Samuel Huntington then the answer to this question should be in the affirmative. Huntington (1993a,1993b,1998), in his well-known *Clash of Civilizations* hypothesis, proposes that people's cultural and religious identities will be the primary source of conflict in the post-cold war period. Brahm [2005] points out that, "[i]n virtually every heterogeneous society, religious difference serves as a source of potential conflict". Surprisingly, the few studies that try to empirically answer this question suggest otherwise. We resolve this apparent contradiction by addressing some of the major shortcomings of the existing empirical literature. We argue that the groupings used so far in the literature to calculate religious diversity are unsatisfactory. For example, while calculating measures of religious diversity, the existing literature considers the cleavage between Catholics and Protestants to be identical to that between Catholics and Muslims. However, since Protestants and Catholics are both sub-sects of Christianity the cleavage between them might be less problematic than that between Catholics and Muslims. We incorporate this insight in the diversity calculations. Moreover, we show that one cannot ignore religious intolerance while investigating the effects of religious diversity on civil conflict. Using newly constructed measures of religious diversity and religious intolerance we find that both religious diversity and religious intolerance are important correlates of civil conflict.

Empirically, there are numerous studies that have tried to pin down the relation between ethnic diversity and civil conflicts.[1] However, most of these studies focus on ethno-linguistic diversity and very few of them rigorously study the relation between religious diversity and civil conflict. The few papers that have actually controlled for religious diversity while investigating the correlates of civil conflict have found it to be insignificant [Fearon and Laitin, 2003, Montalvo and Reynal-Querol, 2005].[2] This is surprising since conflicts such as the civil wars in Afghanistan, the former Yugoslavia and the Sudan, the peace process in Israel and the conflict in Northern Ireland are essentially all conflicts between ethnic groups of different religions [Fox, 1997]. However, the accepted paradigm in this literature is that ethno-linguistic diversity is relevant for civil conflicts while religious diversity is not.

There are several reasons why we revisit the relationship between religious diversity and civil conflicts. First, we argue that the definition of religious groups used in the literature so far has been highly unsatisfactory and unclear which has led to the erroneous finding that religious diversity does not matter for civil conflicts. Religious diversity is always calculated using the currently existing religious sub-groups like Protestants, Catholics, Shias and Sunnis etc. as the relevant groupings.[3] However, it is hardly obvious why these groupings should be more relevant than broader groupings of Christians, Muslims and Hindus etc. Also, Christianity and Islam both share the same origin (Abraham), whereas Hinduism is an Indian religion. Thus the difference between Hindus and Muslims might arguably be more relevant than that between Christians and Muslims. Furthermore if there is any truth in the *Clash of Civilizations* idea of Huntington (1993a,1993b,1998), all sects of Christianity belong to the same civilization, whereas Christianity and Islam clearly belong to different civilizations. Thus, a conflict between the different sects of Christianity might be less likely than that between Christians and Muslims. We thus calculate indices of religious diversity at three different levels. At Level 3 we consider all the existing sub-sects of religions (e.g. Shias and Sunnis), at Level 2 we consider the parent religions of these sects (e.g. Islam and Christianity) and at Level 1 we consider the broad religious traditions from which these religions come (e.g. Abrahamic and Indian).

Theoretically, it is clear that there could be differences in the diversity indices calculated using different group definitions but does it make a difference empirically? To illustrate our approach let us consider a comparison of two countries - India and Switzerland.

[1] Fearon and Laitin (2003), Miguel et al. [2004], Collier (2001), Collier and Hoeffler (2004), Fearon (2005), Montalvo and Reynal-Querol [2005].

[2] Fox [2004] is the only paper that finds evidence in favour of religion being important for conflicts. He finds that not only can religion influence conflict, its influence has been increasing.

[3] For instance the religious diversity measures used in Fearon and Laitin [2003], Miguel et al. [2004], Montalvo and Reynal-Querol [2005]

The religious composition of India is: Muslims -13.4%, Christians - 2.3%, Hindus- 80.5%, Sikhs -1.9%, other religions - 1.8% (including some other Indian religions like Buddhism, Jainism etc.), and none - 0.1%. Switzerland's religious composition is: Muslims - 4.3%, Christians - 78.5% (Roman Catholics- 41%, Protestants - 35.3%, Orthodox 1.8%, Other Christians -0.4 %), other religions - 1%, and none -15.4.%. Evidently Switzerland has high religious diversity if we consider all the sub-sects of Christianity along with its 4.3% Muslim population. However, not only do all the sub-sects of Christianity share the same origins, Christianity and Islam themselves are both Abrahamic religions and thus share the same origin. On the other hand, in India, not only are the three biggest groups of Hindus, Muslims and Christians culturally more dissimilar than the different denominations of Christianity that are present in Switzerland, but more importantly, Hinduism and Sikhism on the one hand and Islam and Christianity on the other hand represent completely different civilizations. Hinduism and Sikhism are both Indian religions whereas Islam and Christianity originated from Abraham.

Not surprisingly, calculating religious polarization (fractionalization) for India at the level of existing religious sub-groups or sects[4] its ranking is 138 (139) which is quite low. However, as soon as we move up levels of aggregation to take into account the origins/cultural similarity of the religions, its ranking changes to 78 (70) at level 2, and to 57 (56) at level 1. Thus, from level 3 to level 2 its ranking moves up 60 places. India is a country which has indeed experienced several violent riots between the Hindus and the Muslims right from the pre-independence period to the present times. Looking at the diversity index calculated at the most disaggregated level, India looks like a below average religiously diverse country. However, once we move up levels it's religiously diversity ranking is quite high. On the other hand if we look at Switzerland, it is one of the most religiously diverse countries calculating diversity at the most disaggregated level. Its religious polarization (fractionalization) ranking is 9 (76) at level 3. However once we move up levels its religious polarization (fractionalization) ranking changes to 128 (128) at level 2, and to 156 (156) at level 1. Switzerland is indeed one of the most peaceful countries in the world.[5]

Second, the issue of religious intolerance has been entirely ignored in the literature. Since, religious intolerance could lead to both lower religious diversity and higher conflict, not controlling for it would lead to meaningless results. Let us consider some illustrative examples. Afghanistan is one of the least religiously diverse countries in the world with 99% of the population being Muslims. However, it is the 5^{th} most intolerant country in our dataset and it has faced years of violent domestic conflict. Moreover, religious intolerance might itself directly lead to lower religious diversity but a higher

[4]This is what is done in all the existing studies that calculate religious diversity.

[5]More examples are provided in the data section.

probability of civil conflict. For example, Pakistan is the third most intolerant country in our dataset. During its partition from India and later on right through to the present day there has been a mass movement of Hindus and Sikhs from Pakistan to India, leading to a fall in religious diversity.[6] On the other hand Pakistan has continuously experienced conflict throughout the years. Thus, without controlling for intolerance Pakistan appears to be a not so diverse country with significant civil violence. In order to correctly identify the effects of religious diversity on civil conflict we need to control for religious intolerance. We thus argue that the finding that religious diversity is irrelevant for conflict while ethno-linguistic diversity is important for it is as much a consequence of not controlling for religious intolerance as it is for constructing religious diversity measures at an erroneous level of aggregation.

We have several novel findings in this paper. First, we find that religious diversity is a significant and robust correlate of the incidence of civil conflict but not so robust correlate of the onset of civil conflict. While religious fractionalization has a significantly negative correlation with the incidence of conflict, religious polarization has a significantly positive correlation with it. This is in line with the theory that points out that there is a non-monotonic relation between diversity and conflict. High polarization represents a large ethnic minority facing off an ethnic majority increasing the probability of conflict, while fractionalization on the other hand increases coordination problems and reduces the probability of conflict (Horowitz [1985], Montalvo and Reynal-Querol [2005]). Contrary to the findings in the existing literature, religious diversity remains a significant correlate of conflict even after controlling for ethno-linguistic diversity.[7]

We find that the religious diversity measures at the second level of aggregation are the more robust correlates of conflict than at other levels. This implies that the cleavage between Hindus, Muslims, and Christians etc. is more relevant than that between Abrahamic and Indian religions or that between different denominations of Christians - like Protestants and Catholics, or of Muslims - like Shias and Sunnis. This result further indicates that aggregating the data at different levels is crucial. Following Huntington's hypothesis, Hinduism, Christianity and Islam all represent different civilizations and that explains the potential for clash among them.

We also find that religious intolerance is a significant and robust predictor of civil conflict. Since our measure of religious intolerance is composed of several components we

[6]Hindu refugees continue coming to India as recently as 2012: http://indiatoday.intoday.in/story/hindu-refugees-from-pakistan-continue-to-reach-india/1/214086.html

[7]Furthermore, the correlations between our measures of religious diversity and the different measures of ethno-linguistic diversity are very low which further ensures that we are not picking up the effects of ethno-linguistic diversity and that religious diversity is an important correlate of civil conflict in its own right.

investigate which specific aspects of intolerance are more important than others. We find that intolerance arising out of social and government regulation of religion significantly lead to more conflict. Government favouritism on the other hand is not a significant predictor of conflict. This is not surprising since social and government regulation of religion are arguably related to the more fundamental right of freedom to worship or to practice a religion of one's own choice.

Finally, we also try to verify whether some religions are more conflict-prone than others and if having more religious people in the population has any impact on civil conflict.[8] We find no evidence in favour of the popular perception that some religions (specifically looking at Islam and Christianity) are more violent than others. We do find some evidence that having more Christians or Non-Religious people in the country reduces the incidence of conflict. This is however not the case for the onset of conflict. We argue that what matters is having a more polarized society and that this polarization comes from groups which are culturally dissimilar, like Christians, Hindus, and Muslims etc. The actual combination of religions that leads to more polarization does not matter. More importantly, the intolerance of the government and society of a country is more relevant for conflict than the presence of any particular religion.

Our results are robust to the use of alternate datasets and specifications, viz. Desmet et al. [2012] and Montalvo and Reynal-Querol [2005]. We then add a host of additional controls including ethnic fractionalization, ethnic polarization, percentage of different religious groups including Muslims, Christians and even Atheists and Non-religious. Our results remain qualitatively similar.

In this paper we seek to make a four-fold contribution to the literature. As discussed above most of the existing literature has focused on the relation between ethno-linguistic diversity and civil conflict. This is a serious gap in the literature. As Huntington [1993b] highlights, "In the modern world, religion is a central, perhaps the central, force that motivates and mobilizes people." Thus, our first contribution is that we rigorously investigate the relation between religious diversity and civil conflict by calculating indices of diversity at three different levels of aggregation. We then let the data tell us what level of aggregation matters for civil conflict.

The second contribution is that we highlight the importance of intolerance in the debate on diversity and conflict. Religious diversity may or may not be important in predicting conflict depending on how tolerant or intolerant society is towards other religions. Moreover, both diversity and conflict might be correlated to intolerance. Thus, it is impossible to over emphasize the importance of intolerance.

[8]Previous research has found religious beliefs to have an effect on crime rates [Shariff and Rhemtulla, 2012].

Thirdly, one often heard argument is that some religions are more peace-loving or more violent than others. Thus one could argue that it is not religious diversity that matters per se but some religions by virtue of being more violent than others lead to more civil conflicts. Controlling for the percentage of Christians, Muslims and Atheists/Non-religious populations we are partly able to answer this question. We find no evidence of any particular religion being more violent than others.

Our final contribution is in terms of the new dataset that we create. We construct six different measures of religious diversity (fractionalization and polarization) at three different levels of aggregation corresponding to different historical depths of cleavages. Moreover, we generate a completely new index of religious intolerance.

To the best of our knowledge no other study has done such a rigorous analysis of the relation between religious diversity, religious intolerance and civil conflict. Moreover, no data on such detailed measures of religious diversity and religious intolerance currently exist for such an exhaustive list of countries.

The rest of the paper is organized as follows. In Section 2 we describe our data sources and explain the construction of our measures in detail. In Section 3 we report our results and in Section 4 we conclude.

2.2 Data & Methodology

2.2.1 Religious diversity

In order to construct our measures of religious diversity we follow the methodology of Desmet et al. [2012]. They demonstrate that the degree of coarseness of ethno-linguistic classifications has profound implications for inference on the role of diversity. They compute ethno-linguistic diversity measures at different levels of aggregation by exploiting the information of language trees. They refer to this as a phylogenetic approach, since tree diagrams describe the family structure of world languages. Depending on how finely or coarsely groups are defined the measure of diversity will be different.[9]

This approach has two advantages. Firstly, it allows the classification of diversity at different levels of aggregation. Secondly and perhaps more interestingly this approach gives a historical dimension to the analysis. Coarse divisions, obtained at high levels of aggregation, describe cleavages that go back thousands of years. In contrast, finer divisions, obtained at low levels of aggregation, are the result of more recent cleavages.

[9]They find that less aggregate measures matter more for public good provision, whereas for civil conflicts deeper cleavages are more relevant than shallower ones.

Moreover, calculating our diversity measure at three different levels we are able to introduce in our indices a measure of cultural dissimilarity between religions. Hindus and Christians are culturally more dissimilar than Protestants and Catholics. This cultural dissimilarity aspect is a crucial point in the *Clash of Civilizations* hypothesis.

The data on religious diversity comes from three distinct sources. We primarily use the CIA World Factbook,[10] and the Alesina et al. [2003] data from Encyclopedia Britannica (EB). Both of these datasets give the proportion of adherents to different religions in the different countries of the world. This data is supplemented by data from http://www.worldstatesmen.org/[11] in case of missing values or lack of detail for some country. Our criteria was to have the most detailed data possible on sub-categories of religions which would allow us to construct meaningful indices at the different levels. For example for Papua New Guinea, the different groups following the CIA World factbook are Baha'i, Indigenous religions, Roman Catholic, Evangelical Lutheran, United Church, Seven day Adventist, Pentecostal, Evangelical Alliance, Anglican, Baptist and Other Protestant. Whereas following the EB there are only the four following groups: Protestant, Roman Catholic, Anglican, and Others. Thus, in this case we prefer to use the CIA data instead of the EB data.

Finally using this data on the percentage of followers of the different religions in each country, we construct six different measures of religious diversity (three of fractionalization and three of polarization) following the below explained methodology. All religions in the world can be classified into several broad groups owing to their origins or cultural traditions. For example, Christianity and Islam are both Abrahamic religions, while Hinduism and Buddhism are both Indian religions. Again, Protestants and Catholics are two sects of Christianity, while Sunnis and Shias are two sects of Islam. For the purposes of this paper we represent this information as tree diagram as given in Figure 2.1. As evident in Figure 2.1, sects like Protestants, Catholics, Shias and Sunnis form our Level 3, which is the most disaggregated level. Then at Level 2 come the parent religions of these sects (Christianity and Islam in this case). And finally at the highest level i.e. Level 1 we have the broad groupings like Abrahamic and Indian religions. Our final data comprises of 118 religious groups at the third level, 45 groups at the second level and 5 groups at the first level (excluding Atheists and Non-religious).[12]

As evident in Figure 2.1, classifying the above broad groups and their corresponding divisions and subdivisions as a tree diagram we have three different levels at which we can measure religious diversity. We thus construct three indices of religious fractionalization

[10] Fearon and Laitin [2003] use a similar dataset based on estimates derived using the CIA Factbook by R. Quinn Mecham.

[11] Allan Drazen also uses data from "World Statesmen" but on some different variables. See: *http : //econweb.umd.edu/ drazen/Data_Sets/Appendix_Composition_and_Elections_revision*22012.*pdf*

[12] The entire list of the divisions can be found in the appendix Table B.16.

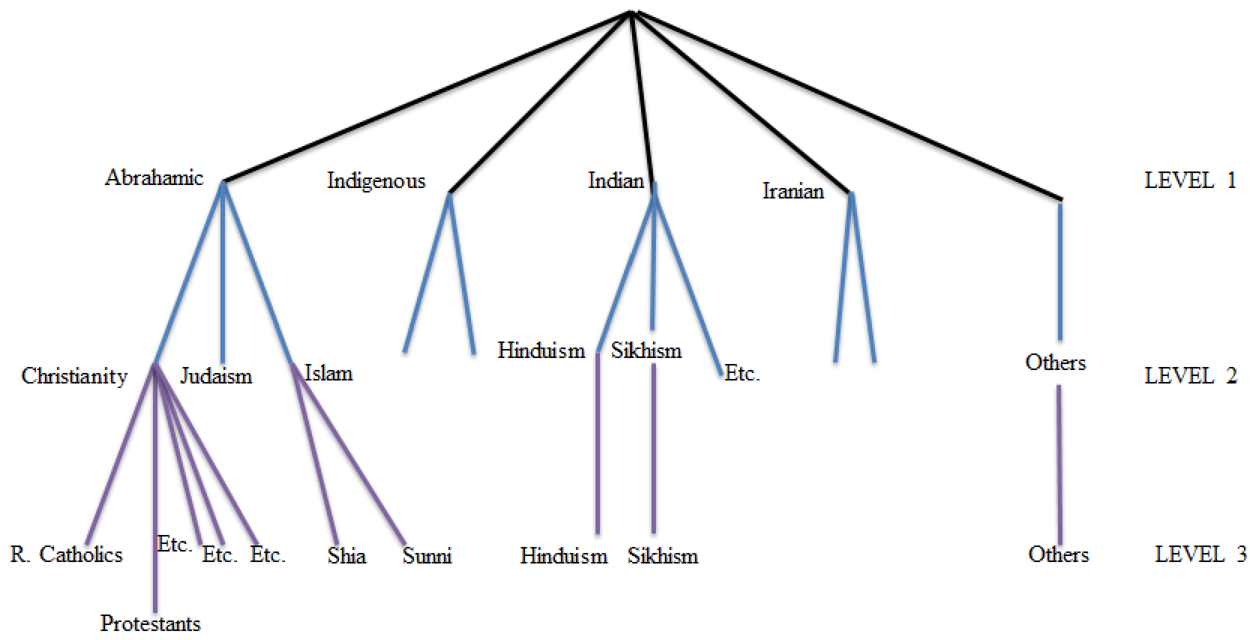

FIGURE 2.1: The religion tree

(rfrac1, rfrac2, and rfrac3) and three indices of religious polarization (rpol1, rpol2, and rpol3) corresponding to the three different levels of aggregation. Rfrac1 (rpol1) corresponds to the highest level of aggregation i.e. it is the most aggregated. Rfrac2 (rpol2) corresponds to the second level of aggregation. And rfrac3 (rpol3) corresponds to the lowest level of aggregation i.e. it is the least aggregated.

The idea behind the measures at each level is identical to the measures of ethno-linguistic diversity (ELF) in Desmet et al. [2012]. The different measures of fractionalization and polarization are constructed as follows:

Fractionalization:

$$rfrac(j) = 1 - \Sigma[S_{i(j)}]^2. \tag{2.1}$$

Polarization:

$$rpol(j) = 4\Sigma[S_{i(j)}]^2[1 - S_{i(j)}]. \tag{2.2}$$

where $S_{i(j)}$ is the proportion of the population pertaining to religious group i at level of aggregation j.[13] The fractionalization measure $rfrac(j)$ gives the probability that two randomly selected individuals from a given country belong to different religious groups. The polarization measure $rpol(j)$ on the other hand measures how far the distribution of the religious groups is from the bipolar distribution (i.e. the (1/2, 0, 0, ... , 0, 1/2)

[13]In case of ethno-linguistic diversity as in Desmet et al. [2012], $S_{i(j)}$ refers to the share of population speaking a particular language i at level of aggregation j.

distribution) which represents the highest level of polarization (Montalvo and Reynal-Querol [2005]). The fractionalization index is maximized when each individual in the country belongs to a different religious group, while the polarization index is maximized when there are only two groups in the country and they are equally sized. The reader is directed to Montalvo and Reynal-Querol [2005] for a detailed discussion and comparison of the two measures.

In order to better illustrate the importance of aggregating at the three different levels let us consider a few countries that have experienced civil conflicts in the past few decades. Consider Angola for example. It is a highly religiously polarized country at any of the three levels of aggregation. At aggregation level 3, it is the 13^{th} most polarized country in the world. However, once we move up levels, it comes out to be the most and second most polarized country in the world considering the 2^{nd} and 1^{st} levels of aggregation respectively. Moreover, if we consider countries like India, Nepal or Indonesia, their religious diversity rankings change by about 60 places moving from the third to the second level of aggregation. The movement from the 3rd to the 2nd level and that from the 2nd to the 1^{st} level need not always be in the same direction. For instance for India, the ranking keeps going up if we move from the 3^{rd} to the 2^{nd} level or the 2^{nd} to the 1^{st} level. However, for Indonesia and Nepal, the ranking goes up from the 3rd to the 2^{nd} level, and falls while moving from the 2^{nd} to the 1^{st} level. These examples help illustrate how changes in the level of aggregation could lead to non-trivial changes in the rankings according to religious diversity.[14]

At any given level, religious fractionalization and religious polarization are highly correlated. While moving from one level to another in many cases both religious fractionalization and polarization seem to move in the same direction. However, the relative changes in the rankings are often different. Consider Nigeria for example. Its polarization ranking is 96 at level 3 making it not a very polarized country. Its fractionalization ranking is 28, making it highly fractionalized country. But when we move to level 2, its fractionalization ranking goes up 19 places to 7, while its polarization ranking goes up 65 places to 31. Moreover, there are cases when the rankings by fractionalization and polarization move in opposite directions while moving from one level to the other. For example, Lebanon is a highly fractionalized country with its fractionalization ranking being 16 at level 3. But it is not a very polarized country placed at rank 116 at level 3. Once we move up one level to level 2, its fractionalization ranking goes down by a

[14] In view of the recent happenings in the Arab world, it is interesting in its own right to look at the religious diversity indices of the Arab Spring countries. While none of these countries have very high levels of religious diversity, their rankings go up significantly while moving up from the 3^{rd} to the 2^{nd} level of aggregation.

marginal 4 places, placing it at 20. On the other hand, its polarization ranking shoots up by 90 places taking it to a rank of 26.[15]

In Tables 2.1 and 2.2 we provide the summary statistics and the correlations between the six different measures of religious diversity. In the appendix Table B.2 we provide the correlations between our measures of religious diversity and the ethnic and religious diversity measures of Montalvo and Reynal-Querol [2005]. In appendix Table B.3 we provide the correlations of our measures with the ethno-linguistic diversity measures of Desmet et al. [2012] respectively.[16] We notice that there is not a very high correlation between our measures and those of either Montalvo and Reynal-Querol [2005] (most are below 0.5 and the highest is 0.7 which is the correlation between their relfrac and our rfrac2) or Desmet et al. [2012] (most are below 0.2 and the highest is 0.37 which is the correlation between their elf10 and our rfrac2).[17]

Table 2.1: Summary statistics for the religious diversity indices

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Religious fractionalization at level 1	0.156	0.178	0	0.643
Religious fractionalization at level 2	0.239	0.206	0	0.703
Religious fractionalization at level 3	0.432	0.25	0	0.891
Religious polarization at level 1	0.295	0.325	0	1
Religious polarization at level 2	0.412	0.321	0	0.996
Religious polarization at level 3	0.570	0.259	0	0.992
N		222		

Table 2.2: Correlation between the religious diversity indices

	rfrac1	rpol1	rfrac2	rpol2	rfrac3
rpol1	0.9906				
rfrac2	0.8235	0.8132			
rpol2	0.7818	0.7906	0.9669		
rfrac3	0.5472	0.5504	0.5845	0.5721	
rpol3	0.5022	0.5146	0.5864	0.6281	0.7821

2.2.2 Religious Intolerance

One obvious concern with any study analysing the effects of religious diversity on conflict is the possible endogeneity of religious diversity. Societies that are more tolerant towards other religions are likely to sustain more religions and thus experience more religious diversity on the one hand, and less civil conflict on the other. Thus if we are to say anything interesting about the effects of religious diversity on civil conflict we must take

[15]See maps in appendix figures A.6 to A.11 to visualize how the rankings change across countries moving from one level to another.

[16]The religious diversity measures of Montalvo and Reynal-Querol [2005] are calculated using the existing religious groups which theoretically correspond to our third level of aggregation and are based on data from *L' Etat des religions dans le monde* and the World Christian Encyclopedia.

[17]A list of all the 222 countries along with their corresponding rankings according to the different rfrac and rpol values is provided in the appendix Tables B.18 and B.19.

into account how tolerant the society is. If we do not control for religious intolerance we would be facing the risk of endogeneity arising from the omitted variable bias.

Measuring religious intolerance is not an easy task since getting reliable data is a big challenge. We use the cross-national, International Religious freedom data, from the Association of Religious Data Archives (ARDA). The specific dataset used is the "International Religious Freedom Data, Aggregate File (2001-2005)."[18]

Each year (since 1999) the U.S. State Department releases International Religious Freedom Reports on approximately 196 countries or territories.[19] Based on the text in these reports, ARDA researchers systematically coded the measures using a survey questionnaire for the years 2001, 2003, and 2005. The most immediate goal was to develop measures for religious regulation and favouritism. For all variables, the coders were asked to make substantive observations of the qualitative data and to base their codes on empirical observations of actions or patterns of behaviour that were documented in the reports.

The three different years of coding are not three discrete measures, but rather represent trend information that continues to be reported for several years running. Thus, ARDA advises researchers to not treat the data as separate measures from which time lines are developed since it may be possible that later years report newly arising problems in addition to old ones. The aggregate dataset for the three years of coding contains the mean score of each ordinal variable across the three years. ARDA suggests that those using the data for social scientific modelling and analysis use the aggregate data set, which has the benefit of greater variation in the variables and lesser error since random errors from one year will be attenuated in the aggregate data. We thus use this aggregate dataset which contains the different indices measured as averages of the three years 2001, 2003 and 2005.[20]

In order to construct our measures of religious intolerance we take into account three different broad level indices which are related to religious intolerance.

1. *Government Regulation Index* (GRI): This index takes into account the following factors: whether foreign or other missionaries are allowed to operate; if proselytizing, public preaching, or conversion is limited or restricted; if the government interferes with an individual's right to worship; how freedom of religion is described in the report; and, if the Introduction section of the Report mentions that the government "generally respects" the right (to religious freedom) in practice.

[18]http://www.thearda.com/Archive/Files/Descriptions/IRFAGG.asp
[19]http://www.state.gov/g/drl/irf/
[20]The reader is directed to the ARDA website for a more detailed description of the data.

2. *Social Regulation Index* (SRI): This index takes into account the following factors: the societal attitudes toward other or non-traditional religions; social attitudes towards conversions to other religions; if traditional attitudes and/or edicts of the clerical establishment strongly discourage proselytizing [trying to win converts]; if established or existing religions try to shut out other religions in any way; and the situation regarding social movements in relation to religious brands in the country.

3. *Government Favouritism Index* (GFI): This index takes into account the following factors: What is the balance of government funding (including 'in kind' such as funding buildings) to the religious sector; how does the government subsidize religion (including 'in kind' to organizations run by religions, e.g., hospitals, schools, etc.); and if the government funds some things related to religion.

Making use of the above indices we construct our measure of religious intolerance via a principal component analysis for 197 countries. Religious intolerance is defined as the first principal component of the three variables, GRI, SRI and GFI. This allows us not only to reduce the dimensionality i.e. have one measure of religious intolerance instead of multiple ones, but also since we use the first principal component we are able to explain about 74% of the orthogonal variation in the data with our measure of religious intolerance.

It is of course possible that our measure of religious intolerance is itself endogenous to conflict. If individuals of any religion experience more conflict with individuals of other religions they might become more intolerant towards other religions and thus the possibility of reverse causality. This is very much a realistic possibility and given our data we partly solve this problem by using a time invariant measure of intolerance. Moreover, we leave out other available variables that also indicate religious intolerance but are more prone to endogeneity. For example variables like, ESTIMAAG - estimated number of people who were physically abused or displaced due to religion and PERSECAG - estimated number of people who were physically abused, displaced from home, imprisoned, or killed due to religion, are left out. Government and Social regulation of religion are variables that are relatively stable over long periods of time.[21]

In Table 2.3, we provide the summary statistics of the religious intolerance variable and its components. Higher values of the variable indicate more intolerance. In Table B.20 of the appendix there is a list of all countries with the corresponding value of religious intolerance of that country. The ten most intolerant countries in our sample are Saudi

[21]As a robustness check we do include both ESTIMAAG and PERSECAG in the calculation of our intolerance index. But due to their potential endogeneity we leave them out from our final calculations. However, results remain qualitatively unchanged to their inclusion. Results are not provided, but are available upon request.

Arabia, Iran, Pakistan, Burma (or Myanmar), Afghanistan, Egypt, Iraq, Uzbekistan, Kuwait and the Maldives, in that order.[22]

In appendix Table B.1 we provide the correlations of religious intolerance with the measures of diversity calculated at different aggregation levels. We notice that as expected religious intolerance is negatively correlated with religious diversity at all levels of aggregation. The correlations are not very high, the highest correlation being of about -0.4 between religious intolerance and religious fractionalization at the third level of aggregation.

TABLE 2.3: Summary statistics - Religious intolerance

Variable	Mean	Std. Dev.	Min.	Max.
Government Regulation Index	3.293	3.076	0	9.722
Social Regulation Index	3.605	2.928	0	10
Government Favouritism Index	4.837	2.778	0	9.388
Religious intolerance	-0.008	1.489	-2.299	3.258
N		197		

2.2.3 Specification

We use the above constructed measures to study the effects of religious diversity and intolerance on the onset and incidence of civil conflict. Our baseline econometric specification given by equation 3, follows Desmet et al. [2012] who in turn borrow it from the baseline specification of Fearon and Laitin [2003] and augment it with a number of additional control variables.

$$y_{it} = \alpha + \delta D_i(j) + \gamma I_i + \beta X_{it} + \epsilon_{it} \tag{2.3}$$

where, y_{it} is either the onset or the incidence of civil conflict in country i in year t. While onset refers to a new conflict starting in a particular year, incidence refers to whether a country is experiencing a civil conflict in that particular year. $D_i(j)$ is a time invariant measure of religious diversity at aggregation level j in country i, I_i is the time invariant religious intolerance in country i, α is the constant term and ϵ_{it} the error term. The vector of controls X_{it} come from major contributions in the literature. They include, lagged civil war, the log of per capita GDP (lagged), the percentage of the country that is mountainous, non-contiguous state dummy, oil exporter dummy, new state dummy, instability dummy, democracy lagged (polity2), continent dummy variables for Sub-Saharan Africa, East and Southeast Asia, Latin America and the Caribbean, and legal origin dummies. The conflict data and the corresponding data on the covariates span

[22]Appendix Figure A.5 gives a world map for religious intolerance.

from 1945 to 1999 and were constructed by Fearon and Laitin [2003]. The legal origin dummies come from La Porta et al. [1999].

Our dependent variable y_{it} representing either the onset or incidence of conflict is a 0-1 binary variable. On the other hand our primary variables of interest, religious diversity and intolerance are time invariant. Thus, we use a pooled Logit approach.[23] The standard errors are always clustered at the country level.

Since we want to study the partial effects of religious diversity and religious intolerance on civil conflict, δ and γ are the main coefficients of interest.[24]

2.3 Results

2.3.1 Civil Conflict

First, using the data and estimation method of Desmet et al. [2012], we examine how religious diversity and religious intolerance affect the incidence and onset of civil conflicts. The difference is that instead of using their measures of ethno-linguistic diversity we use our measures of religious diversity in addition to controlling for religious intolerance in some of the specifications. Also, following Montalvo and Reynal-Querol [2005] we include both fractionalization and polarization in the same specification.

Table 2.4 and Table 2.5 give our baseline results. The dependent variable in Table 2.4 is the incidence of civil war, while the dependent variable in the Table 2.5 is the onset of civil war. In both tables, Columns 1 to 3 each correspond to a different level of aggregation in the calculation of religious diversity. Column 1 corresponds to the highest level of aggregation while column 3 to the lowest level. Columns 4 to 6 are identical to the specifications of columns 1 to 3, but in these three columns we also control for religious intolerance.

In the first three columns of Table 2.4 we notice that while religious polarization is significant at the second and third levels of aggregation, religious fractionalization is significant at all levels of aggregation. In columns 4 to 6, where we control for religious intolerance in the specifications of columns 1 to 3, the results change substantially. We see that religious intolerance is associated with more civil conflict and this relation is significant in two of the three specifications. Religious diversity on the other hand

[23]We could have also use a pooled Probit approach, but following Fearon and Laitin [2003] and Desmet et al. [2012] we stick to the Logit model

[24]Causality is not the main focus of this paper and thus caution should be exercised when interpreting δ and γ causally.

Table 2.4: Correlates of Incidence of Civil wars

	(1) Level1	(2) Level2	(3) Level3	(4) Level1	(5) Level2	(6) Level3
Religious fractionalization	**-7.911***	**-3.197***	**-1.574***	**-9.910****	**-4.252****	-1.123
	(4.302)	(1.873)	(0.892)	(4.303)	(1.877)	(0.919)
Religious polarization	3.819	**1.965***	**1.576****	**4.948****	**2.605****	**1.328***
	(2.392)	(1.188)	(0.697)	(2.375)	(1.196)	(0.705)
Lagged civil war	6.258***	6.271***	6.253***	6.196***	6.206***	6.202***
	(0.216)	(0.215)	(0.211)	(0.211)	(0.210)	(0.207)
Log lagged GDP/cap	-0.307**	-0.302**	-0.230*	-0.194	-0.182	-0.174
	(0.146)	(0.146)	(0.139)	(0.144)	(0.143)	(0.139)
Log lagged population	0.346***	0.337***	0.317***	0.317***	0.300***	0.292***
	(0.0700)	(0.0628)	(0.0654)	(0.0779)	(0.0717)	(0.0762)
% mountainous	0.0104***	0.00927**	0.0105***	0.00993**	0.00843**	0.00963**
	(0.00383)	(0.00398)	(0.00387)	(0.00389)	(0.00411)	(0.00384)
Noncontiguos state dummy	0.393	0.399	0.496	0.512	0.512	0.586*
	(0.348)	(0.339)	(0.336)	(0.346)	(0.333)	(0.335)
Oil exporter dummy	0.236	0.223	0.198	0.0289	0.00142	0.0543
	(0.275)	(0.260)	(0.254)	(0.302)	(0.280)	(0.282)
New state dummy	1.818***	1.831***	1.831***	1.762***	1.781***	1.789***
	(0.381)	(0.383)	(0.382)	(0.394)	(0.395)	(0.392)
Instability dummy	-0.0165	-0.0125	0.0127	-0.0200	-0.0143	0.00889
	(0.280)	(0.277)	(0.280)	(0.282)	(0.280)	(0.281)
Democracy lagged (Polity 2)	0.0125	0.0122	0.0162	0.0185	0.0192	0.0206
	(0.0203)	(0.0201)	(0.0203)	(0.0208)	(0.0205)	(0.0208)
French legal origin dummy	2.422***	2.835***	2.955***	2.250***	2.666***	2.776***
	(0.532)	(0.547)	(0.594)	(0.578)	(0.600)	(0.638)
UK legal origin dummy	2.316***	2.697***	2.832***	2.085***	2.497***	2.577***
	(0.539)	(0.542)	(0.610)	(0.603)	(0.614)	(0.681)
Socialist legal origin dummy	2.064***	2.171***	2.223***	1.990***	2.075***	2.103***
	(0.566)	(0.540)	(0.617)	(0.603)	(0.588)	(0.657)
Latin America and Carribean Dummy	0.162	0.104	-0.0321	0.474	0.439	0.212
	(0.342)	(0.328)	(0.361)	(0.371)	(0.363)	(0.390)
Sub-Saharan Africa dummy	0.555*	0.477	0.594*	0.984***	0.980**	0.853**
	(0.309)	(0.330)	(0.359)	(0.376)	(0.394)	(0.391)
East and Southeast Asia Dummy	0.710**	0.482	0.486	0.902**	0.694**	0.545
	(0.358)	(0.316)	(0.321)	(0.391)	(0.338)	(0.347)
Religious intolerance				**0.192***	**0.207****	0.149
				(0.0996)	(0.0960)	(0.107)
Constant	-7.810***	-8.171***	-8.901***	-8.464***	-8.802***	-9.072***
	(1.444)	(1.407)	(1.450)	(1.497)	(1.469)	(1.536)
Observations	5733	5733	5733	5678	5678	5678
Pseudo R^2	0.744	0.743	0.744	0.745	0.744	0.744
ll	-593.8	-594.7	-592.6	-589.4	-590.0	-589.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors, clustered at the level of countries, in parentheses.

The dependent variable is the incidence of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

becomes highly significant at the first and second level of aggregation and less so at the third level of aggregation.

We find that religious polarization increases the probability of conflict while fractionalization reduces it. This is in line with the previous literature on ethnicity and conflict which argues that the relationship between ethnic diversity and conflict is not monotonic. Societies where a large ethnic minority face an ethnic majority experience more conflict. On the other hand, highly heterogeneous or homogenous societies face less

Table 2.5: Correlates of Onset of Civil Conflict

	(1) Level1	(2) Level2	(3) Level3	(4) Level1	(5) Level2	(6) Level3
Religious fractionalization	-8.280	-2.859	-1.639	**-10.27***	**-3.791***	-1.077
	(5.297)	(2.100)	(1.257)	(5.805)	(2.154)	(1.283)
Religious polarization	4.068	1.595	**1.755***	5.255	**2.240***	1.473
	(2.987)	(1.314)	(0.974)	(3.208)	(1.339)	(0.958)
Lagged civil war	-0.872***	-0.892***	-0.899***	-0.931***	-0.956***	-0.928***
	(0.262)	(0.267)	(0.260)	(0.249)	(0.247)	(0.245)
Log lagged GDP/capita	-0.603***	-0.607***	-0.553***	-0.512***	-0.510***	-0.496***
	(0.148)	(0.147)	(0.151)	(0.158)	(0.156)	(0.158)
Log lagged population	0.336***	0.319***	0.301***	0.275***	0.254***	0.238***
	(0.0827)	(0.0766)	(0.0736)	(0.0937)	(0.0862)	(0.0846)
% mountainous	0.00931*	0.00805	0.00757	0.00848	0.00652	0.00640
	(0.00481)	(0.00497)	(0.00511)	(0.00548)	(0.00578)	(0.00565)
Noncontiguos state dummy	0.373	0.417	0.441	0.478	0.527	0.568
	(0.370)	(0.368)	(0.353)	(0.365)	(0.357)	(0.347)
Oil exporter dummy	0.683***	0.707***	0.710***	0.467*	0.494**	0.538**
	(0.239)	(0.238)	(0.230)	(0.253)	(0.243)	(0.239)
New state dummy	1.771***	1.781***	1.793***	1.713***	1.730***	1.746***
	(0.371)	(0.371)	(0.377)	(0.382)	(0.383)	(0.383)
Instability dummy	0.606***	0.626***	0.646***	0.602***	0.619***	0.641***
	(0.217)	(0.215)	(0.217)	(0.218)	(0.216)	(0.219)
Democracy lagged (Polity 2)	0.0200	0.0194	0.0231	0.0274	0.0279	0.0295
	(0.0209)	(0.0209)	(0.0207)	(0.0206)	(0.0205)	(0.0207)
French legal origin dummy	1.258*	1.547**	1.842**	1.082	1.373*	1.612*
	(0.679)	(0.706)	(0.799)	(0.749)	(0.770)	(0.858)
UK legal origin dummy	1.027	1.308*	1.601**	0.724	1.040	1.198
	(0.669)	(0.685)	(0.797)	(0.750)	(0.766)	(0.871)
Socialist legal origin dummy	1.289*	1.347**	1.445*	1.224	1.257*	1.276
	(0.708)	(0.684)	(0.783)	(0.761)	(0.745)	(0.833)
Latin America and Carribean Dummy	0.172	0.112	-0.0117	0.581	0.566	0.370
	(0.404)	(0.397)	(0.431)	(0.393)	(0.394)	(0.422)
Sub-Saharan Africa dummy	0.295	0.288	0.235	0.770*	0.834*	0.614
	(0.401)	(0.479)	(0.481)	(0.435)	(0.491)	(0.473)
East and Southeast Asia Dummy	0.600*	0.404	0.344	0.812**	0.620*	0.440
	(0.354)	(0.349)	(0.336)	(0.387)	(0.366)	(0.357)
Rel intolerance				**0.240****	**0.250****	**0.219****
				(0.102)	(0.104)	(0.108)
Constant	-4.491***	-4.553***	-5.359***	-4.728***	-4.815***	-5.232***
	(1.732)	(1.685)	(1.797)	(1.788)	(1.739)	(1.869)
Observations	5733	5733	5733	5678	5678	5678
Pseudo R^2	0.100	0.098	0.101	0.105	0.103	0.104
ll	-453.6	-454.5	-453.3	-450.3	-451.0	-450.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors, clustered at the level of countries, in parentheses.

The dependent variable is the onset of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

conflict. While Polarization captures the former concept, fractionalization captures the latter concept.[25]

In Table 2.5 we use the same specifications as in Table 2.4 , but instead of using the incidence of civil wars we use the onset of civil wars as our dependent variable. We notice that our results are qualitatively similar. Religious intolerance is highly significant

[25]See Horowitz [1985], Esteban and Ray [1999], Montalvo and Reynal-Querol [2005]. See Montalvo and Reynal-Querol [2005] for a discussion.

and associated with more civil conflicts. Religious diversity on the other hand remains significant only at the second level of aggregation.[26]

We notice that the level of aggregation matters for understanding the relation between religious diversity and civil conflicts. In particular, the robustness of the measures of religious diversity constructed at the second level of aggregation implies that the cleavage between Hindus, Muslims, and Christians etc. is more relevant than a higher or lower level of aggregation. This gives further support to our hypothesis that the level of aggregation at which the measures of diversity are constructed is important.

2.3.2 Components of intolerance

Our measure of religious intolerance is constructed using three different components viz. government regulation, social regulation and government favouritism. Next we analyse which of these specific components of religious intolerance are more important for civil conflicts. In Table 2.6 we use the identical specifications of the last three columns of Table 2.4, but in place of religious intolerance we control for its different components separately in the 3 panels. The columns 1 to 3 (and 4 to 6) each correspond to a different level of aggregation in the calculation of religious diversity. The dependent variable in the first three columns is the Incidence of conflict, whereas in the columns 4 to 6 the dependent variable is the Onset of conflict.

In Table 2.6, we see that government and social regulation of religion are significant and robust correlates of both the Incidence and Onset of civil conflict. On the other hand, government favouritism of religion does not seem to be relevant for civil conflicts. This result is not surprising since government and social regulation of religion are arguably more fundamental types of intolerance as they relate to the more fundamental right of freedom to religion, while government favouritism is less so.

We also notice that when we control for each component of religious intolerance separately our religious diversity measures continue to remain significant and robust predictors of the Incidence of civil conflict particularly at the second level of aggregation. As far as the Onset of conflict is concerned as before religious diversity does not come out to be relevant.

[26]When religious fractionalization and polarization enter the specifications separately, they are not significant. Results in appendix Tables B.5, B.6, B.7 and B.8.

Table 2.6: Components of intolerance and Civil wars

	Incidence			**Onset**		
	(1) Level1	(2) Level2	(3) Level3	(4) Level1	(5) Level2	(6) Level3
Religious fractionalization	**-10.01****	**-4.040****	-1.210	**-10.02***	**-3.728***	-1.178
	(4.214)	(1.870)	(0.917)	(5.739)	(2.119)	(1.274)
Religious polarization	**4.906****	**2.395****	**1.329***	4.990	2.094	1.434
	(2.333)	(1.196)	(0.715)	(3.157)	(1.302)	(0.981)
Government Regulation Index	**0.0905***	**0.0886***	0.0550	**0.108****	**0.110****	0.0849
	(0.0477)	(0.0473)	(0.0540)	(0.0442)	(0.0463)	(0.0521)
Observations	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.744	0.744	0.744	0.104	0.102	0.102
ll	-589.5	-590.6	-590.3	-450.6	-451.5	-451.5
Religious fractionalization	**-9.121****	**-3.974****	-1.131	-8.915	-3.246	-0.920
	(4.417)	(1.867)	(0.865)	(6.210)	(2.126)	(1.190)
Religious polarization	**4.566***	**2.382****	**1.278***	4.612	1.896	1.313
	(2.420)	(1.206)	(0.681)	(3.363)	(1.333)	(0.893)
Social Regulation Index	**0.124****	**0.134*****	**0.112****	**0.163*****	**0.170*****	**0.164*****
	(0.0492)	(0.0461)	(0.0514)	(0.0530)	(0.0525)	(0.0559)
Observations	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.745	0.745	0.745	0.109	0.108	0.109
ll	-587.6	-588.0	-587.8	-448.2	-448.5	-448.2
Religious fractionalization	**-8.151***	**-3.478***	-1.366	-8.501	-3.036	-1.425
	(4.387)	(1.895)	(0.909)	(5.412)	(2.134)	(1.275)
Religious polarization	3.960	**2.160***	**1.470****	4.190	1.716	**1.662***
	(2.422)	(1.196)	(0.691)	(3.055)	(1.344)	(0.967)
Government Favoritism Index	0.0127	0.0215	0.00744	0.0110	0.0145	0.0146
	(0.0434)	(0.0436)	(0.0449)	(0.0505)	(0.0517)	(0.0517)
Observations	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.743	0.743	0.744	0.100	0.098	0.100
ll	-591.8	-592.6	-591.2	-452.8	-453.7	-452.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust Standard errors clustered at the country level in parentheses.

The dependent variable in the first three columns is the Onset of Civil Conflict, and in the last three columns it is the Incidence of civil conflict. Column 1 (4), 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

2.3.3 Are some religions more problematic than others?

One often heard argument is that some religions are more violent than others. Thus, one could argue that it is not religious diversity that matters per se but some religions by virtue of being more violent than others lead to more civil conflicts. In this section we investigate this claim and find no evidence whatsoever in its favour. Since Christians and Muslims are the biggest religious groups in the world and are widely distributed across countries we consider these two religions. We also consider the presence of Atheists/Agnostics and Non-religious populations.

In Table 2.7 we have religious intolerance and religious diversity at the three different levels of aggregation as in our baseline specification of Table 2.4. We however, also control for the percentage of Muslims, Christians, and Non-religious etc. entering in different combinations. We first notice that our measures of religious diversity continue to remain significant predictors of the conflict. Religious intolerance however, becomes

Table 2.7: Incidence of Civil wars - Percentages of different groups

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6	(7) est7
Religious fractionalization	-7.633*	-9.831**	-9.552**	-6.924*	-7.660*	-9.503**	-6.190
	(4.261)	(4.308)	(4.189)	(4.120)	(4.250)	(4.174)	(4.082)
Religious polarization	3.735	4.950**	4.599**	3.167	3.783	4.447*	2.519
	(2.352)	(2.368)	(2.318)	(2.290)	(2.340)	(2.295)	(2.288)
Religious intolerance	0.156	0.165	0.134	0.0833	0.137	0.154	0.108
	(0.102)	(0.109)	(0.104)	(0.109)	(0.110)	(0.107)	(0.109)
Percentage of Non Religious/Atheists	-0.0122			-0.0147**	-0.0118		-0.0185**
	(0.00751)			(0.00736)	(0.00763)		(0.00812)
Percentage of Muslims		0.00213			0.00163	-0.00326	-0.00633
		(0.00322)			(0.00326)	(0.00462)	(0.00504)
Percentage of Christians			-0.00642**	-0.00727**		-0.00883*	-0.0122**
			(0.00309)	(0.00309)		(0.00477)	(0.00516)
Observations	5678	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.745	0.745	0.745	0.746	0.745	0.745	0.746
ll	-588.4	-589.2	-588.0	-586.5	-588.3	-587.7	-585.7
Religious fractionalization	-4.288**	-4.208**	-4.389**	-4.446**	-4.248**	-4.439**	-4.542**
	(1.849)	(1.880)	(1.851)	(1.814)	(1.856)	(1.874)	(1.831)
Religious polarization	2.816**	2.674**	2.650**	2.874**	2.872**	2.606**	2.797**
	(1.181)	(1.194)	(1.178)	(1.155)	(1.176)	(1.178)	(1.139)
Religious intolerance	0.163	0.171	0.160	0.107	0.132	0.173	0.130
	(0.1000)	(0.105)	(0.101)	(0.106)	(0.107)	(0.106)	(0.108)
Percentage of Non Religious/Atheists	-0.0173**			-0.0188**	-0.0170**		-0.0198**
	(0.00792)			(0.00787)	(0.00797)		(0.00822)
Percentage of Muslims		0.00273			0.00246	-0.00181	-0.00359
		(0.00351)			(0.00345)	(0.00572)	(0.00574)
Percentage of Christians			-0.00565*	-0.00633**		-0.00691	-0.00886
			(0.00320)	(0.00319)		(0.00549)	(0.00556)
Observations	5678	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.745	0.744	0.745	0.746	0.745	0.745	0.746
ll	-587.9	-589.7	-588.8	-586.4	-587.7	-588.8	-586.2
Religious fractionalization	-4.288**	-4.208**	-4.389**	-4.446**	-4.248**	-4.439**	-4.542**
	(1.849)	(1.880)	(1.851)	(1.814)	(1.856)	(1.874)	(1.831)
Religious polarization	2.816**	2.674**	2.650**	2.874**	2.872**	2.606**	2.797**
	(1.181)	(1.194)	(1.178)	(1.155)	(1.176)	(1.178)	(1.139)
Religious intolerance	0.163	0.171	0.160	0.107	0.132	0.173	0.130
	(0.1000)	(0.105)	(0.101)	(0.106)	(0.107)	(0.106)	(0.108)
Percentage of Non Religious/Atheists	-0.0173**			-0.0188**	-0.0170**		-0.0198**
	(0.00792)			(0.00787)	(0.00797)		(0.00822)
Percentage of Muslims		0.00273			0.00246	-0.00181	-0.00359
		(0.00351)			(0.00345)	(0.00572)	(0.00574)
Percentage of Christians			-0.00565*	-0.00633**		-0.00691	-0.00886
			(0.00320)	(0.00319)		(0.00549)	(0.00556)
Observations	5678	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.745	0.744	0.745	0.746	0.745	0.745	0.746
ll	-587.9	-589.7	-588.8	-586.4	-587.7	-588.8	-586.2

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust Standard errors clustered at the country level in parentheses.

The dependent variable is the Incidence of civil conflict. Panel 1, 2 and correspond to religious diversity calculated at levels 1, 2 and 3 respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section. The other controls are: a constant term, lagged civil war, the log of per capita GDP (lagged), the percentage of the country that is mountainous, non-contiguous state dummy, oil exporter dummy, new state dummy, Instability dummy, democracy lagged (polity2), continent dummy variables for Sub-Saharan Africa, East and Southeast Asia, Latin America and the Caribbean, and legal origin dummies from La Porta et al. [1999].

insignificant. We notice that the percentages of Christians and Non-religious/Atheist population reduce the incidence of civil conflict. The percentage of Muslims in the

TABLE 2.8: Incidence of Civil wars - Percentages of different groups

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6	(7) est7
Religious fractionalization	-5.767	-8.391**	-8.246**	-6.167	-6.502	-8.111**	-5.255
	(4.332)	(4.237)	(4.182)	(4.176)	(4.240)	(4.107)	(4.082)
Religious polarization	2.698	4.204*	3.802	2.700	3.189	3.659	1.978
	(2.407)	(2.333)	(2.331)	(2.335)	(2.336)	(2.270)	(2.298)
Percentage of Non Religious/Atheists	-0.0141*			-0.0149**	-0.0119		-0.0187**
	(0.00805)			(0.00747)	(0.00812)		(0.00825)
Percentage of Muslims		0.00499*			0.00404	-0.00163	-0.00537
		(0.00297)			(0.00304)	(0.00449)	(0.00498)
Percentage of Christians			-0.00902***	-0.00912***		-0.0104**	-0.0136**
			(0.00306)	(0.00298)		(0.00487)	(0.00519)
Observations	5733	5733	5733	5733	5733	5733	5733
Pseudo R^2	0.744	0.744	0.745	0.746	0.744	0.745	0.746
ll	-592.3	-592.5	-590.6	-589.0	-591.5	-590.5	-588.4
Religious fractionalization	-3.420*	-3.402*	-3.701**	-3.953**	-3.578*	-3.697**	-3.967**
	(1.835)	(1.892)	(1.857)	(1.816)	(1.854)	(1.867)	(1.813)
Religious polarization	2.308**	2.294*	2.209*	2.563**	2.549**	2.243*	2.509**
	(1.171)	(1.199)	(1.170)	(1.146)	(1.174)	(1.177)	(1.131)
Percentage of Non Religious/Atheists	-0.0185**			-0.0190**	-0.0167**		-0.0197**
	(0.00825)			(0.00795)	(0.00835)		(0.00843)
Percentage of Muslims		0.00589*			0.00496	0.000873	-0.00171
		(0.00324)			(0.00326)	(0.00549)	(0.00564)
Percentage of Christians			-0.00835***	-0.00837***		-0.00766	-0.00971*
			(0.00314)	(0.00307)		(0.00558)	(0.00565)
Observations	5733	5733	5733	5733	5733	5733	5733
Pseudo R^2	0.744	0.744	0.744	0.745	0.745	0.744	0.745
ll	-592.2	-593.1	-592.0	-589.4	-591.1	-592.0	-589.3
Religious fractionalization	-3.420*	-3.402*	-3.701**	-3.953**	-3.578*	-3.697**	-3.967**
	(1.835)	(1.892)	(1.857)	(1.816)	(1.854)	(1.867)	(1.813)
Religious polarization	2.308**	2.294*	2.209*	2.563**	2.549**	2.243*	2.509**
	(1.171)	(1.199)	(1.170)	(1.146)	(1.174)	(1.177)	(1.131)
Percentage of Non Religious/Atheists	-0.0185**			-0.0190**	-0.0167**		-0.0197**
	(0.00825)			(0.00795)	(0.00835)		(0.00843)
Percentage of Muslims		0.00589*			0.00496	0.000873	-0.00171
		(0.00324)			(0.00326)	(0.00549)	(0.00564)
Percentage of Christians			-0.00835***	-0.00837***		-0.00766	-0.00971*
			(0.00314)	(0.00307)		(0.00558)	(0.00565)
Observations	5733	5733	5733	5733	5733	5733	5733
Pseudo R^2	0.744	0.744	0.744	0.745	0.745	0.744	0.745
ll	-592.2	-593.1	-592.0	-589.4	-591.1	-592.0	-589.3

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust Standard errors clustered at the country level in parentheses.

The dependent variable is the Incidence of civil conflict. Panel 1, 2 and correspond to religious diversity calculated at levels 1, 2 and 3 respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section. The other controls are: a constant term, lagged civil war, the log of per capita GDP (lagged), the percentage of the country that is mountainous, non-contiguous state dummy, oil exporter dummy, new state dummy, Instability dummy, democracy lagged (polity2), continent dummy variables for Sub-Saharan Africa, East and Southeast Asia, Latin America and the Caribbean, and legal origin dummies from La Porta et al. [1999].

country on the other hand has no effect on the incidence of civil conflicts.[27]

Looking at the list of countries ranked by our measure of religious intolerance in appendix Table B.20, one could argue that since most of the Muslim majority countries figure high up on the list, our insignificance of the percentage of Muslims in the data is driven

[27] In appendix Tables B.9 (with intolerance) and B.10(without intolerance) we see that no religion can be particularly blamed for the onset of conflict. Religious intolerance continues to be significant for the onset of conflict.

by the inclusion of the religious intolerance variable. In Table 2.8 we use the same specification of Table 2.7 but leave the religious intolerance variable out. Our results remain qualitatively similar. We notice that percentages of Christians and Atheists/Non-religious continue to reduce the incidence of the conflict. The percentage of Muslims in marginally significant as long as we do not control for the percentage of other groups, but as soon as we control for the other groups the percentage of Muslims becomes insignificant. Thus, the insignificance of percentage of Muslim variable is not driven by the inclusion of religious intolerance in the specification.[28]

In the diversity measures we have been using so far, each religious group existing at a particular level enters as a separate entity at that level. One might argue that if a particular religion is more problematic than others then the relevant conflict inducing cleavage is the one between that problematic religion and all other religions. Thus, a diversity index which includes all religions separately might not be the best one to pick this effect up. In order to verify this, we construct three new measures of religious diversity dividing the population of each country into only two mutually exclusive and exhaustive groups for each of the three measures. In the first one, we consider only Muslims and Non-Muslims as the relevant groups, in the second one we consider only Christians and Non-Christians as the relevant groups and finally, in the third one we consider only Religious and Atheists/ Non-Religious as the relevant groups. Since, in this case there are only two groups entering the calculation of diversity, both polarization and fractionalization yield the same ranking of countries. We use fractionalization without loss of generality.[29] Neither of these new measures of diversity are significant with or without the inclusion of religious intolerance. Religious intolerance continues to have a significant and robust effect on the onset of civil conflict and less so on the incidence of civil conflict.[30]

The insignificance of the diversity measures that include only the division between Muslims and Non-Muslims or that between Christians and Non-Christians further supports our previous finding that neither Christianity nor Islam is particularly problematic. The insignificance of the diversity index that includes only Religious and Atheists/ Non-Religious as the relevant groups, indicates that the cleavage between the religious and Non-religious people in the country is not relevant in predicting civil conflict.

[28]As seen in the Table B.4 the correlation between religious intolerance and the percentage of different groups is not especially high.

[29]Using polarization instead would produce identical results. For only two groups polarization = 2*fractionalization (see Montalvo and Reynal-Querol [2005] page 798 for a discussion.)

[30]Results are not provided and are available upon request. In the first two of these indices we have tried including the atheists and Non-religious in the calculation. The inclusion or non-inclusion of the atheists and Non-religious has no qualitative effect on the results.

2.3.4 Robustness checks

Next, we try to ensure that our results are robust to other datasets and specifications. In order to do so we look specifically at the dataset and specification of Montalvo and Reynal-Querol [2005]. They use a sample of 138 countries for the 1960-1999 period and divide the sample into 5 five-year periods. The data comes from the Peace Research Institute of Oslo (PRIO) dataset for civil wars and their basic endogenous variable corresponds to the incidence of civil wars following the definition of PRIO which includes intermediate and high-intensity armed conflicts (PRIOCW). [31]

TABLE 2.9: Montalvo and Reynal-Querol [2005] specification with Religious diversity & intolerance

	(1) Level1	(2) Level2	(3) Level3	(4) Level1	(5) Level2	(6) Level3
Religious fractionalization	**-33.14*****	**-11.39*****	**-3.563***	**-52.84****	**-14.19*****	-2.273
	(12.55)	(3.765)	(1.923)	(23.68)	(4.379)	(2.167)
Religious polarization	**15.68****	**6.924*****	**3.809****	**26.03****	**8.882*****	**3.392***
	(6.803)	(2.296)	(1.806)	(12.46)	(2.897)	(1.930)
LGDPC	-0.584**	-0.415	-0.427*	-0.400*	-0.243	-0.309
	(0.237)	(0.253)	(0.225)	(0.227)	(0.235)	(0.213)
LPOP	0.549***	0.479***	0.412**	0.410**	0.327*	0.238
	(0.180)	(0.151)	(0.207)	(0.202)	(0.167)	(0.222)
PRIMEXP	0.0697	-0.457	-0.569	-0.863	-1.167	-1.830
	(2.068)	(1.778)	(1.694)	(2.126)	(1.679)	(1.910)
MOUNTAINS	-0.00478	-0.00391	-0.00582	-0.00660	-0.00659	-0.00562
	(0.00925)	(0.00921)	(0.0105)	(0.0103)	(0.0105)	(0.0100)
NONCONT	0.106	0.000484	0.330	0.185	0.0433	0.404
	(0.581)	(0.536)	(0.622)	(0.635)	(0.550)	(0.670)
DEMOCRACY	0.0675	0.0332	0.115	0.303	0.315	0.225
	(0.348)	(0.357)	(0.354)	(0.398)	(0.397)	(0.370)
ETHPOL	2.175**	2.276**	2.360**	1.896	1.804	1.885
	(1.088)	(1.109)	(1.132)	(1.155)	(1.149)	(1.180)
ETHFRAC	0.257	0.526	0.270	1.096	1.476	0.647
	(0.920)	(0.968)	(1.004)	(1.036)	(1.007)	(1.054)
Religious intolerance				**0.395***	**0.448****	**0.407****
				(0.212)	(0.195)	(0.192)
CONSTANT	-7.160**	-7.859***	-7.234**	-6.728**	-7.151**	-5.542
	(3.284)	(2.862)	(3.584)	(3.343)	(2.930)	(3.691)
N	846	846	846	838	838	838
pseudo R^2	0.178	0.154	0.157	0.215	0.202	0.190
ll	-294.1	-302.7	-301.5	-279.9	-284.5	-288.6

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The dependent variable is the incidence of civil wars from PRIO following the definition which includes intermediate and high-intensity armed conflicts (PRIOCW). Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Montalvo and Reynal-Querol [2005] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

Their main finding is that ethnic polarization has a positive and statistically significant effect on the incidence of civil wars. Then in some of their specifications they also look at the effects of religious heterogeneity. They find that neither religious fractionalization nor religious polarization have a significant effect on conflict when they enter separately.

[31] See Montalvo and Reynal-Querol [2005] for more details.

On the other hand, in the basic logit regressions using both religious fractionalization and religious polarization, they find that religious fractionalization is marginally insignificant, while religious polarization is statistically significant. However, once ethnic polarization is included only ethnic polarization is significant and all the other diversity measures become insignificant. They argue that "It seems clear that ethnic polarization has a robust and powerful explanatory power on civil wars in the presence of other indices of fractionalization and polarization, while the statistical relevance of religious polarization depends on the particular specification." Thus, we try to verify if this is indeed true or does our measure of religious diversity still have a significant effect on civil conflict once we control for ethnic fractionalization and polarization. At the same time we also ensure that our results are robust to using the PRIO dataset.

In Table 2.9 we use the data and specification of Montalvo and Reynal-Querol [2005] and add to it our measures of religious diversity instead of theirs.[32] Columns 1 to 3 correspond to the 3 different levels of aggregation. Columns 4 to 6 are identical to columns 1 to 3 but also control for religious intolerance. Unlike them we find that religious fractionalization and religious polarization continue to be highly significant in predicting civil conflict even after controlling for ethnic fractionalization and polarization. While this result holds at almost all levels of aggregation, it is the second level of aggregation that is the most significant once we control for religious intolerance. Also, religious intolerance is highly significant in all the specifications. Moreover, while our measures of religious diversity continue to be significant, ethnic polarization becomes insignificant once we control for religious intolerance. Most of the literature has so far found ethnic polarization to be significant and religious diversity to be insignificant in explaining civil conflict. But our finding indicates that the result was driven by the non-inclusion religious intolerance in the specifications.

In Table B.11, we re-investigate which of the three components of religious intolerance are more relevant. This is similar to Table 2.6, but while in Table 2.6 we use the data and specification of Desmet et al. [2012], here we use the data and specification of Montalvo and Reynal-Querol [2005]. In the first three columns we do not control for any component of intolerance whereas in the last three we do control for each of the three components in the three different panels. Again we clearly notice that government and social regulation of religion are highly significant in explaining civil conflict. Moreover, religious diversity continues to be highly significant and robust.

We finally subject our analysis to some more robustness checks. We control explicitly for ethnic fractionalization and polarization from Montalvo and Reynal-Querol [2005] in the Desmet et al. [2012] data and specification (Appendix Table B.12 and Table B.13).

[32] This corresponds to Table 1, Column 8 of Montalvo and Reynal-Querol [2005]

We also added the percentage of Muslims, Christians and Non-religious/atheists in the countries, both entering together and separately in of Montalvo and Reynal-Querol [2005] (Appendix Table B.14). Our results remain qualitatively unchanged.

In all the above analyses we do not consider the Non-religious/Atheists/Agnostics as a relevant group in the calculation of the diversity indices. As a further robustness test we re-calculated our religious diversity measures including these groups. Our results remain qualitatively unchanged but less significant. We find that when we do not control for religious intolerance, and religious fractionalization and polarization enter the specifications separately, both religious polarization and fractionalization are significant (only) at the highest level of aggregation (for both incidence and onset). But the sign is negative i.e. religious diversity (both fractionalization and polarization) at the highest level of aggregation seems to reduce conflict. This significance however, disappears for polarization once we control for religious intolerance for the incidence of conflict and also fractionalization for the onset of conflict. Further, unlike our other specifications if religious fractionalization and polarization enter the specifications together, neither is significant.[33]

We argue that "since what we are trying to capture is religious interaction, it is reasonable not to treat the no-religion group as other religions because the only things that people in this group have in common is the fact they do not belong to any religious group. Therefore, there are not specific common interests that permit to identify them as a collective and that distinguish them from the interest of all the other groups. This means that from a political point of view there is no common point of reference that keeps them together. Moreover, the non-religious group does not have the necessity to reaffirm its identity because, as a group, it has no identity. This means that social friction caused by religious differences with other groups will not be present" (Montalvo and Reynal-Querol [2000]).

2.4 Conclusion

In this paper we create measures of religious diversity at three different levels of aggregation corresponding to different historical depths of cleavages. We also construct a new measure of religious intolerance. Using our newly constructed measures we do an in-depth empirical analysis of the relation between religious diversity, intolerance and the probabilities of onset and incidence civil conflict.

[33]These results are not provided and are available upon request.

Through our empirical analysis, we find that religious diversity is an important correlate of civil conflict. The relationship is more robust for the incidence of conflict and less so for the onset of conflict. Religious fractionalization reduces conflict while religious polarization increases it. The relationship is a lot weaker for the onset of conflict. Moreover, religious intolerance is a significant and robust correlate of civil conflict, more so for the onset of conflict. In particular intolerance arising out of social and government regulation of religion significantly leads to more conflict. While we do find some evidence that a higher percentage of Non-religious/Atheists and Christians reduce the incidence of conflict we find no evidence in favour of the perception that some religions are more violent than others. Moreover, having more Muslims, Christians or more religious people in the country has no significant effect on the onset of conflict.

We also find that religious diversity measured at the second level of aggregation is the most robust one. In other words, the cleavage between Hindus, Muslims, and Christians etc. is more relevant than that between Abrahamic and Indian religions or that between different denominations of Christians - like Protestants and Catholics, or of Muslims - like Shias and Sunnis. Thinking in terms of Huntington [1993a], the relevant groups that define civilizations which potentially clash are the groups like Hindus, Muslims, and Christians etc. as defined by the second level of aggregation. On the other hand, thinking in terms of Caselli and Coleman [2013], these religious identies or groups are seperated by an ethnic distance which imposes a high enough cost on individuals of one group to pass themselves off as members of the other.

Our results are robust to a host of specifications, data and controls including controls for other forms of ethnic diversity. We conclude from the above empirical analysis that both religious diversity and intolerance are important predictors of civil conflict and must be taken into account in any analysis investigating the correlates of civil conflict.

Chapter 3

The Health Costs of Ethnic Distance: Evidence from Sub-Saharan Africa

3.1 Introduction

Nineteen thousand children die worldwide every day before reaching the age of five. The highest rates of child mortality are still concentrated in Sub-Saharan Africa, where 1 in 9 children die before reaching the age of five, which is not only more than 16 times the average for developed regions (1 in 152) but also a lot higher than South Asia (1 in 16) which has the second highest rates of child mortality [UNICEF, 2012]. The striking feature of child mortality rates in Africa is that while all of Africa is poor, there is a huge disparity in child mortality rates across ethnic groups. We argue that ethnic distances could explain the existence of such disparities in child mortality rates across ethnic groups in Africa. The ethnic distance between any two ethnic groups is measured by how different the languages that the two groups speak are. We show that children of mothers who are ethnically distant to their neighbours have higher mortality rates. One possible explanation for our finding is that information does not flow smoothly across ethnic lines and individuals who are ethnically distant to their neighbours lose out.

Child mortality or under-five mortality is an important measure of development and reducing it is in fact a Millennium Developmental Goal (MDG no. 4). Our main focus is to explain why the child mortality rates vary across ethnic groups in Africa. The literature so far has attributed such differences in child mortality rates across ethnic groups in Africa to different practices by different ethnic groups which affect the demographic behaviour and cultural status of women; the geographical location of different groups, i.e. whether the groups live close to cities, or in environmentally better regions with better climate that supports better crops, and less diseases like malaria or if they live in economically important regions etc. Gyimah [2002] for example, particularly stresses the differences in socio-cultural practices such as dietary taboos and food avoidances on mothers and infants, as well as perceptions of disease aetiology and treatment patterns across ethnic groups.[1]

Using high quality individual level micro data from the Demographic and Health Surveys (DHS) combined with a novel dataset on the spatial distribution of ethnic groups at the level of approximately 1 km^2 for fourteen Sub-Saharan African countries we highlight the importance of ethnic distance in explaining the disparities in child health outcomes between ethnic groups while controlling for the commonly provided explanations including geography, location, cultural differences between ethnic groups etc. The ethnic distance variable is calculated using the average linguistic distance of the mother from individuals living around her in circles of different radii.

There are two primary findings of our study. First, children of mothers who are ethnically distant from their neighbours have a higher probability of dying as children. This result is robust to the inclusion of the commonly used measures of ethnic diversity (fractionalization or polarization), as well as several individual specific controls, apart from ethnicity, region, and country-time fixed effects. This finding holds as long as ethnic distances are calculated using circles of radii ranging from 25 km to 125 km around the mother, but not in bigger circles. Our second key finding is that children of mothers living in more ethnically fractionalized places have a lower probability of child death.[2]

In terms of average marginal effects, if we consider a circle of 75km around the mother, a one standard deviation (0.267) increase in the linguistic distance of the mother from her neighbours increases the probability of her children dying by around 0.75%, which is about 3.3% of all the child deaths in our sample. By itself this marginal effect looks small. But to put things in perspective, a one year increase in the mother's education leads to only a 0.5% decrease in the probability of child death. And moving to live one standard deviation closer to the capital decreases the probability of death by only 0.6%.

While there are several possible alternative explanations, we use the recent insights of Ashraf and Galor [2013] to interpret our empirical findings. Ashraf and Galor [2013]

[1] "Among the Mole-Dagbani groups in northern Ghana, for instance, women are denied eggs and other protein food during pregnancy which is likely to affect their nutritional status hence the birth weight of the child. Similarly, pregnant Akan women are encouraged to avoid rich food such as mangoes and ripe plantain for fear of miscarriages, particularly in the early months of pregnancy" [Gyimah, 2002].

[2] Controlling for polarization instead does not change the results. Fractionalization gives the probability that two randomly selected individuals from a given country speak two different languages. Polarization measures how far the distribution of the linguistic groups is from the bipolar distribution (1/2, 0, 0, ... , 0, 1/2).

point out that diversity could have both positive and negative impacts on economic outcomes. In the same vein, we argue that ethnic diversity leads to a higher stock of knowledge in society about how to rear one's children and thus improves health outcomes of children, whereas individual ethnic distances act as barriers to accessing such knowledge and thus leads to worse health outcomes.[3]

Since we control for ethnicity specific fixed effects our results are not driven by heterogeneity in unobservable characteristics across ethnic groups. This lets us abstract from previously provided explanations of differences in child mortality rates between ethnic groups including differences in cultural practices between different ethnic groups etc. Also, the use of region and country-time fixed effects allow us to discard the location and geography explanations. However, there still remain concerns about endogeneity of our results due to unobserved differences between individuals who live in places where they are ethnically more or less distant. In order to address such endogeneity we control for a host of variables that would reduce the possibilities of omitted variable bias. We also explicitly control for migration in some of the specifications. Then, using recently developed methods by Altonji et al. [2005] we show that our results are not driven by selection on unobservables.

Apart from our two main results we also find some evidence that wealthier mothers can mitigate some of the negative effects of linguistic distance. But we find no heterogeneity in the effects of linguistic distance by education, gender, place of residence (urban or rural), or distance from the capital. Neither do we find any evidence of nonlinearities in the effects of either linguistic distance or ethnic diversity. We do several robustness checks including the use of the linear probability model instead of probit regressions, and using infant mortality as our dependant variable instead of child mortality. Our results remain qualitatively unchanged.

Our paper contributes to several different strands of the literature. One of the primary strands is the literature that finds ethnic diversity to have a negative effect on the provision of public goods including health outcomes.[4] However, most of the literature is at the cross country level. Interestingly, Platas [2010] finds that, but for Africa ethnic diversity is either insignificant or has a positive and significant effect on some health outcomes including child mortality (e.g. Appendix Table 12 of Platas [2010]). The literature has yet to provide an explanation for this seemingly puzzling finding. In contrast take this literature to an individual level analysis with a rich set of controls

[3]We discuss some of the other possible explanations in the next section.

[4]Miguel and Gugerty [2005], Alesina et al. [1999], Vigdor [2004], Habyarimana et al. [2007], Alesina et al. [2003], Desmet et al. [2012], La Porta et al. [1999], Ahlerup [2009], Ghobarah et al. [2004], Lieberman [2007], Platas [2010]

and point out that while ethnic diversity might indeed have a positive effect on child mortality rates, ethnic distance has a negative effect.

The second strand that we contribute to is the literature that views child mortality as an indicator of individual welfare or development and tries to understand its determinants.[5] Some of this literature particularly focuses on ethnic favouritism ([Franck and Rainer, 2012, Kudamatsu, 2009]. However, these papers usually attempt to identify the effect of the ethnicity of the countries' leader on mortality rates in different groups. So far none of these papers have looked at the effects of individual ethnic distance on child health outcomes.

The third strand to which we contribute is the small but growing literature that emphasizes the role of ethnic distances in explaining different socio-economic outcomes.[6] We take this literature a step forward. To the best of our knowledge, we are the first to relate individual level health outcomes to the ethnic distance of the individual from her neighbours.[7]

Finally, we contribute to the literature that tries to explain the existence of ethnic inequality in health outcomes. Ethnic inequality defined as the inequality in well-being across ethnic groups that coexist, is bad for economic growth [Alesina et al., 2012], provision of public goods [Baldwin and Huber, 2010], and can lead to civil conflicts [Gomes, 2012, Mitra and Ray, 2010] etc. Most of the literature studying the effects of ethnic inequality on different socio-economic outcomes has taken such inequality to be exogenously given. In contrast, we show how ethnic distances might lead to disparities in health outcomes between ethnic groups. There are some papers [Brockerhoff and Hewett, 2000, Gyimah, 2002] that have tried to explain ethnic inequality in child mortality rates, but we are the first to underscore the importance of ethnic distance.[8]

The rest of the paper is organized as follows. Section 2 explains why ethnic distance and diversity might matter for child mortality. Section 3 discusses the data sources and how different variables are constructed. Section 4 gives the econometric specification. Section 5 gives the results and section 6 concludes.

---

[5] Besley and Kudamatsu [2006], Kudamatsu [2009], and Kudamatsu et al. [2012], Franck and Rainer [2012].

[6] Desmet et al. [2012], Desmet et al. [2009], Gomes [2013], Spolaore and Wacziarg [2009], Esteban et al. [2012a], Esteban et al. [2012b].

[7] Kumar et al. [2012] find geographic distance to be an important barrier to good maternal and child health outcomes in India. We on the other hand investigate the effects of ethnic distance on child health outcomes while controlling for geographic distance.

[8] Ethnic inequality is not just a feature of Africa. Banerjee and Somanathan [2007] point out how the Scheduled Tribes are falling behind Scheduled Castes in India while to start with both the groups were historically disadvantaged.

3.2 Why Does Ethnic Distance Matter?

Our empirical findings show that the more ethnically distant a mother is from her neighbours the higher the probability of her children dying before reaching the age of five. On the other hand children of mothers living in more ethnically fractionalized places have a lower probability of dying. There are several possible alternative explanations to our empirical findings. We use the recent insights of Ashraf and Galor [2013] who underscore that diversity could have both beneficial and detrimental effects on productivity. On the one hand diversity enhances knowledge creation and accumulation and fosters technological progress in the economy. On the other hand diversity leads to more inefficiency by increasing the possibilities of disarray and mistrust and thus leading to reduced cooperation and disrupting socioeconomic order.

Using the Ashraf and Galor [2013] framework with minor modifications we argue that diversity reflects a higher stock of knowledge and information about how to rear one's children. Overall child mortality is thus lower in diverse localities. However, such knowledge does not flow smoothly across ethnic groups and especially it does not flow to groups which are ethnically very distant and thus such groups lose out. Formally, we consider an economy where the level of ethnic diversity affects the level of productivity. The level of technology is given by $A = A(z, \omega)$, where z denotes the institutional, geographical and human capital factors and $\omega = [0, 1]$, which is the degree of ethnic diversity, has a positive but diminishing effect on the level of technology.[9] However, δ_i, which is the ethnic distance of an individual to her neighbours, reduces the individual production. Let x be the individual labour input. Thus, the individual health production function is given by,

$$y_i = (1 - \delta_i)A(z, \omega)f(x_i) \tag{3.1}$$

The individual ethnic distance, δ_i, represents a barrier to knowledge and impedes individual access to information about good health.[10] On the other hand overall diversity actually improves health outcomes via the technology term technology $A = A(z, \omega)$.

The role of information for healthcare cannot be over emphasized. Malhotra [2012] for example points out how lack of information on feeding practices or nutritional knowledge amongst families plays a key role in the persistence of chronic child malnutrition in India.

[9] $A(z; \omega) > 0$, $A_\omega(z; \omega) > 0$, and $A_{\omega\omega}(z; \omega) < 0$ for all $\omega \in [0; 1]$.

[10] If we take $\delta_i = 0$, then we get a simplified version of the model in Ashraf and Galor [2013]. In Ashraf and Galor [2013] additionally "a fraction, $\alpha\omega$, of the economy's potential productivity, $A(z; \omega)$, is lost due to lack of cooperation and resultant inefficiencies in the production process." Hence, they show that $y = (1 - \alpha\omega)A(z, \omega)f(x) \equiv y(x, z, \omega)$ is a strictly concave hump-shaped function of ω.

More than a third of under-five deaths worldwide are attributable to under nutrition [UNICEF, 2012]. Again, the Audience Scape National Surveys highlight the role of information for health care for several African countries. For example, they find that in Zambia people who have received information about HIV/AIDS, malaria, or family planning within the month prior to the survey were more likely to be in better health [Zhou, 2010]. "Word of mouth" is found to be important for health information, with friends or family members acting as key channels of such information. The top three most trustworthy sources for health issues were found to be medical doctors, radio, and friends or family members.[11]

There is also evidence in the literature of how ethnicity might act as a barrier to information and thus affect health care. Pongou [2009] for example points out that information circulates more easily within ethnic groups than across and highlights the implications for HIV/AIDS in Africa. Again, in Kenya there is clear evidence on targeting of spread of health information via maximum language use, Kiswahili being the language of the majority [Bowen, 2010]. However, this makes it harder for fringe groups to get access to such information. Again, Singleton and Krause [2009] points out how Spanish speaking patients face barriers to accessing health care even in the US.[12] We take this literature a step further and show how ethnic distances might have a negative impact on child mortality.

Our framework can be easily modified and related to the literature on ethnic diversity and public goods. If like the Alesina et al. [1999] model we assume that the median voter decides which public goods get provided, then as an individual's distance from the median voter (the average person around her) increases, the worse off she is. Again, individuals who are very different from others can be easily identified and thus can be easily discriminated against reducing their access to public goods [Blattman and Miguel, 2010, Caselli and Coleman, 2013, Fearon and Laitin, 1996, Miguel and Gugerty, 2005]. However, using these models we cannot explain why ethnic diversity has a positive effect on health outcomes while ethnic distances have a negative effect. We thus stick to our simple framework based on Ashraf and Galor [2013].

[11] Similar surveys with similar findings exist for other countries as well. See Montez [2011] for Tanzania, for example.

[12] In Mexico indigenous people don't go to the hospital in fear that their language and customs will not be understood and due to lack of trust between groups. http://www.nytimes.com/video/2013/08/13/world/americas/100000002373842/a-chiapas-medicine-man.html

3.3 Data

3.3.1 Spatial Distribution of Ethnic groups

In order to construct the ethnic distance of the mother from people living around her, we need the distribution of ethnic groups across space. Until very recently there was no comprehensive database on the spatial distribution of ethnic groups available.[13] Desmet et al. [2013] (work in progress) fill this gap by constructing the most comprehensive database on the spatial distribution of ethnic groups for the whole world at a resolution of approximately $1km^2$. In order to do so they use two different sources of data. For the spatial distribution of population they use the Landscan data. Landscan is the finest resolution global population distribution data available for the entire world. The resolution is 30 arc seconds by 30 arc seconds, which is approximately 1 km^2 at the equator.[14] For the information on ethnic groups they use the 15^{th} edition of Ethnologue which maps over 7600 linguistic groups for the whole world. The linguistic groups are represented in the form of polygons across space where each polygon represents the homeland of a particular linguistic group. Areas where multiple languages are spoken are represented via overlapping polygons. Moreover, the total population pertaining to a particular linguistic polygon within a particular political boundary is also provided.[15]

Figure 3.1 gives the distribution of all African languages along with their corresponding populations. Polygons of different colours represent different languages. Areas where multiple languages are spoken are represented by overlapping polygons. But the overlapping polygons cannot be seen in this map. The blank areas have no information on languages but are almost always sparsely populated (or unpopulated) desert areas. If there is some population living in these areas then they are assigned to the language of the nearest polygon. Figure 3.2 gives the population distribution at the 1 km^2 level coming from LandScan which gives us the number of people living in each square kilometre of the world. Desmet et al. (2013) overlay these two maps on each other via some programming in ArcGIS, Python, Matlab and Stata to construct a distribution of linguistic groups for each of the Landscan cells. The final data on the spatial distribution of ethnic groups gives us for each square km of the world the languages spoken and how many people speak each of those languages in that square km.

[13] Alesina and Zhuravskaya [2011] construct a database of ethnic diversity at the sub-national level but their database only goes down to the district level and only for 92 countries.

[14] For details see http://web.ornl.gov/sci/landscan/

[15] Alesina et al. [2013] use a similar methodology using the same data sources and calculate district and country level averages of historical plough use by ancestors of different ethnic groups.

Figure 3.1: African Languages

For example, let us consider the case of Ethiopia whose population distribution can be seen in the Figure 3.3.[16] Figure 3.4 gives the languages in Ethiopia once it is overlaid on the population distribution. Then depending on the number of language polygons overlying a particular population cell, the total language population, and the total population of the country, each cell gets a particular distribution of languages. This is repeated for the whole world. We use the data for Africa.

3.3.2 Linguistic distance

In this paper we measure ethnic distances using distance between the languages which the different ethnic groups speak. There are several different ways of measuring linguistic distances. We follow Fearon [2003], Desmet et al. [2009], Desmet et al. [2012] and several other recent papers which use linguistic tree diagrams to measure distance between languages. The distance between two languages j and k using this approach is defined as:

$$\tau_{jk} = 1 - \left(\frac{l}{m}\right)^{\delta} \tag{3.2}$$

[16]If we zoom in into the capital Addis Ababa we can clearly notice how the population is concentrated around the capital, see in Figure C.1.

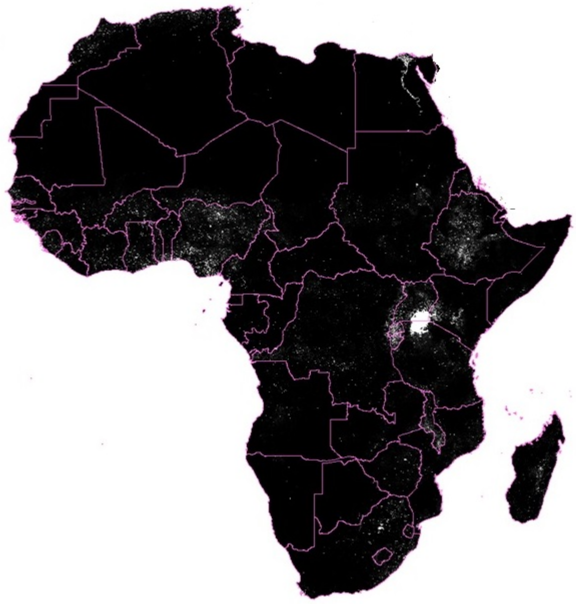

Figure 3.2: Africa Population

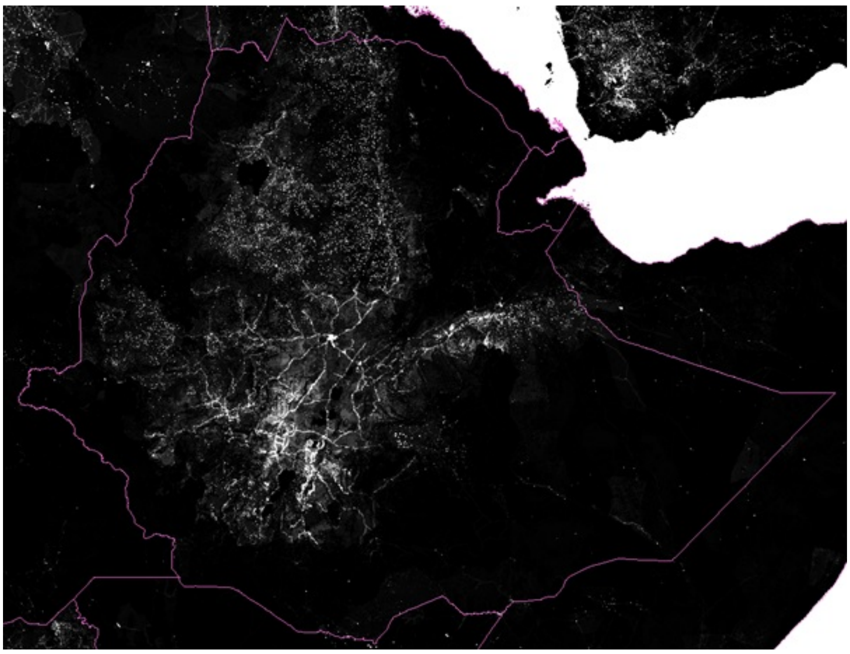

Figure 3.3: Ethiopia Population

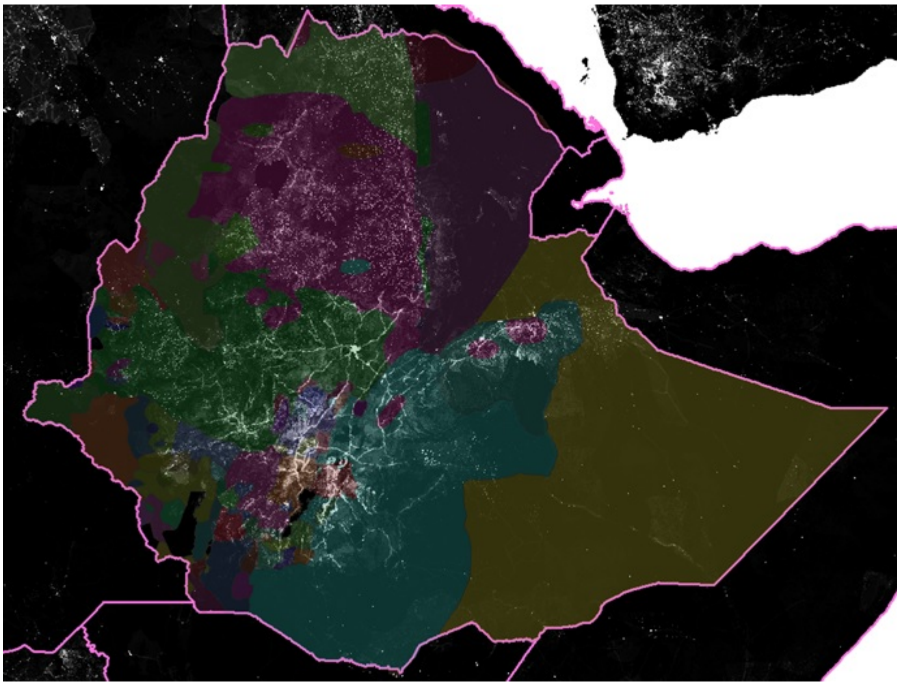

FIGURE 3.4: Ethiopia languages

where l is the number of shared branches between j and k, m is the maximum number of branches between any two languages, and δ is the decay factor, which is a parameter that determines how fast the distance declines as the number of shared branches increases. Data on language trees come from the Ethnologue data.

For our final analysis we need to calculate the average linguistic distance of each mother in our sample to all other individuals living around her in circles of different radii. The linguistic distance for each mother j (who speaks language j) to all other individuals in the circle is given by $\sum_{k=1}^{n} \tau_{jk}$, where there are n individuals living in the circle and k represents the language of each of those n individuals. The function τ_{jk} is defined by the formula 3.2.

There are other ways of measuring linguistic distances. For example, Dyen et al. (1992) measure linguistic distances based on lexicostatistical studies which focus on the proportion of cognates in any two languages.[17] The distance between any two languages is usually defined as one minus the proportion of cognates they have. Again, Isphording [2013] uses not only cognates but also the number of sounds that need to be changed between two words that have the same meaning (say, Tu and You) in two different languages. However, distance calculated using language tree diagrams are more useful since the data is a lot more comprehensive and exists for all countries.[18]

[17]Cognates are words in different languages that sound similar and mean the same thing.
[18]See Desmet et al. [2009] for a discussion.

The decay factor δ measures, "how much more distant should we consider two languages from different families to be relative to languages that belong to the same family" [Desmet et al., 2009]. Following Desmet et al. [2009] we use a δ of 0.5. Fearon [2003] on the other hand uses a δ of 0.05. Let us consider the two Indo-European languages Greek and Italian. Following the language tree from Ethnologue, these two languages share one common branch and with a δ of 0.5, the distance between them is 0.74. Again, if we consider Chinese and Italian which belong to completely different families and thus share no branches in common, the distance between them is one. On the other hand if like Fearon (2003), we take a δ of 0.5, the distance between Greek and Italian becomes 0.13, whereas that between Chinese and Italian continues to be one.[19]

3.3.3 Linguistic Diversity

Our primary measure of ethnic diversity is the commonly used measure of ethnic fractionalization. Ethnic fractionalization has been found to be bad for a host of socio economic outcomes [Alesina et al., 2003] and has often been blamed for Africa's poor economic performance [Easterly and Levine, 1997]. However, recent literature has emphasized that for certain outcomes like civil conflicts etc. ethnic polarization rather than fractionalization is more relevant [Montalvo and Reynal-Querol, 2005]. In some of our specifications we also control for polarization. If the level of analysis is the country level, then there is a distinction between these two measures of diversity. However, at very fine levels of disaggregation, as is our case, both fractionalization and polarization are highly correlated with a correlation of above 0.82, and yield similar results.

The fractionalization measure $frac(j)$ gives the probability that two randomly selected individuals from a given country speak two different languages. The polarization measure $pol(j)$ on the other hand measures how far the distribution of the linguistic groups is from the bipolar distribution (i.e. the (1/2, 0, 0, ... , 0, 1/2) distribution) which represents the highest level of polarization [Montalvo and Reynal-Querol, 2005]. The fractionalization index is maximized when each individual in the country belongs to a different linguistic group, while the polarization index is maximized when there are only two groups in the country and they are equally sized.[20] Formally, the two measures are defined as follows:

$$\text{Fractionalization: } frac(j) = 1 - \Sigma[S_{i(j)}]^2. \tag{3.3}$$

[19] Example from Desmet et al. [2009].

[20] The reader is directed to Montalvo and Reynal-Querol [2005] for a detailed discussion and comparison of the two measures.

$$\text{Polarization: } pol(j) = 4\Sigma[S_{i(j)}]^2[1 - S_{i(j)}]. \tag{3.4}$$

where $S_{i(j)}$ is the proportion of the population speaking language i at geographic region j. The disaggregated nature of our data allows us to calculate diversity and different levels of aggregation. We thus calculate our measures of diversity at both the circle and district levels. Geographic region in the different specifications could refer to either circles or districts. However, as we will see later this does not affect our results.

3.3.4 Child Mortality

Child mortality is the death of a child before reaching the age of five. If the child dies before reaching the age of one then it is termed as infant mortality while if the child does not survive for a month then it is termed as neo-natal mortality. The Demographic and Health surveys (DHS) make available data on child mortality at the individual level for many developing countries from across the world. Funded by the the U.S. Agency for International Development (USAID), the DHS has been conducting surveys in several developing countries since the 1980s. By interviewing a nationally representative sample of women of child bearing age (15 to 49), the DHS collects data on all the children they have ever given birth to in the past including the children who did not survive till the time of the interview. The standardized components of the DHS questionanaire can be used to compile cross country micro datasets. Each survey provides information on life and health outcomes of individuals which allow us to construct measures of child mortality and several other indiviudal level variables including the child's gender, birth-order, their mother's weight, stature, years of education, and occupation, wealth level etc. Moreover, in many of the surveys the mother's ethnicity is also provided. Also, using a GPS receiver, the geographic coordinate of each geographic cluster (village or town) is also collected. In Figure 3.5 for example, we see the the spatial location of all DHS clusters in Ethiopia. In Figure 3.6 we draw circles of 25 km radius around these locations in order to calculate average ethnic distances.

We use fourteen countries in our analysis viz. Kenya, Uganda, Ethiopia, Burkina Faso, Malawi, Senegal, Zambia, Sierra Leone, Mali, Guinea, Ghana, Benin, Namibia and Niger. To start with we consider all Sub-Saharan African countries for which DHS data is available. Our analysis requires along with the regular DHS data on health outcomes, the ethnicity and GPS coordinates of the mother's location. Thus, countries which have either of these two information missing have to be discarded. Then in order to calculate the linguistic distance we need to be able to map the ethnicities to the languages. For the matching between ethnicity and languages we closely follow Fearon [2003]. But our

Figure 3.5: Ethiopia DHS example

matching is a lot more comprehensive than Fearon [2003]. He provides the matching for the major ethnic groups in the country whereas we construct the matching for each and every ethnic group in the country regardless of its size. We have also had to discard countries for which this mapping is not good enough.[21]

3.3.5 Geographic Distance

The geographic isolation and distance from the capital are calculated by using the formula for the great circle distance. The geographic distance between any two points in space ℓ and k, denoted by $|\ell, k|$, is computed as the great circle distance:

$$|\ell, k| = r_E \arccos(\sin(lat_\ell)\sin(lat_k) + \cos(lat_\ell)\cos(lat_k)\cos(long_\ell - long_k) \qquad (3.5)$$

Using the above formula we calculate the geographic isolation of any individual as the average geographic distance betweeen that individual and all the other people in the country. The highest distance between any two individuals in the country is normalized to 1. The distance to the capital is simply the distance of the individual to the capital

[21] We have cross-checked our data using Fearon [2003].

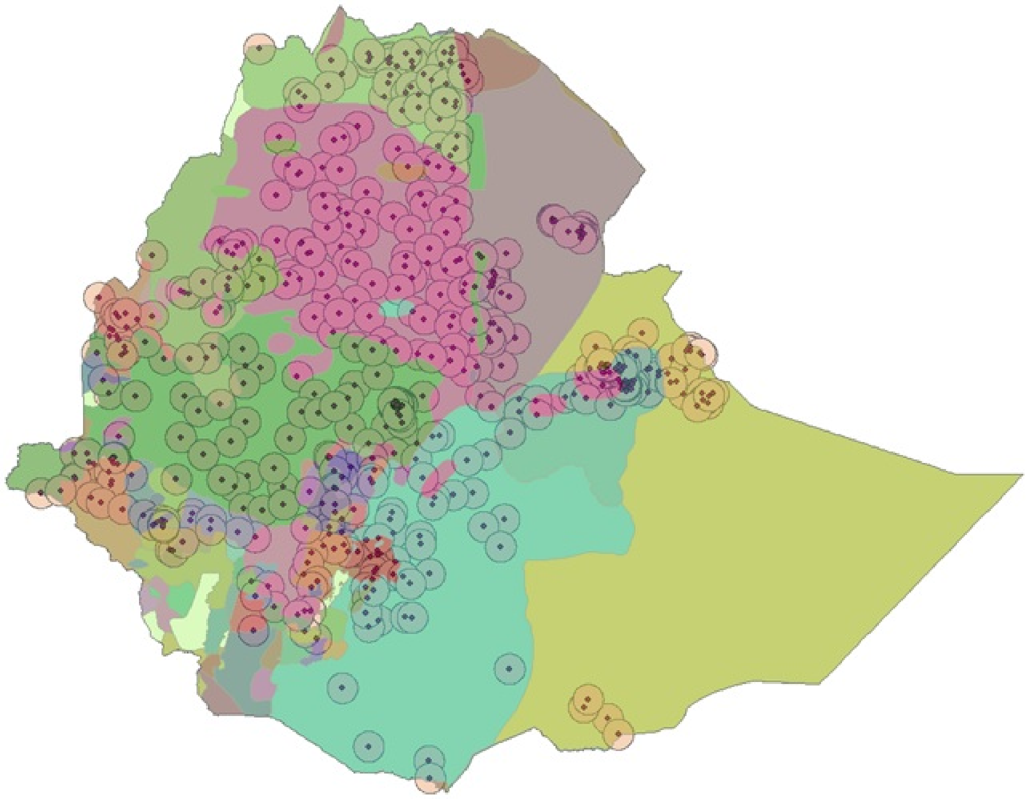

FIGURE 3.6: Ethiopia distance example

calculated using formula 3.5, where we take ℓ to be the individual's location and k to be the location of the capital.

3.3.6 Summary Statistics

Figure C.2 in the appendix gives the maps of all the countries included in the analysis. We have fourteen countries and a total of thirty surveys with information on the births and deaths of over 825,000 children. However, we can consider only the children who have already reached the age of five by the day of the sampling, since we do not know if the others are going to survive till the age of five or not. Thus, as can be seen from Table C.1, in the child mortality sample we have information on the births of 658,755 children out of who about 23% do not survive till their fifth birthday. About 12% do not survive till their first birthday. There is huge variability in the data over time and space. From Table C.2 we can see that in Niger 35% of the children die before reaching the age of five, whereas in Kenya the corresponding number is 13%. The linguistic distance and fractionalization variables all lie between 0 and 1. 22% of the sample is urban and 49% of the sample is female.

3.4 Econometric Specification

Our primary relationship of interest is that between child mortality and the ethno-linguistic distance of the mother from her neighbours. We would also like to understand how ethnic diversity of the neighbourhood affects child mortality. The baseline specification is given by equation (3.6). In our baseline specification we have child mortality on the left hand side and our primary variables of interest on the right hand side along with a host of control variables that have been found to be important for child mortality.

$$\begin{aligned} Prob(death_{iet}) = \alpha_R \;+\; \alpha_{ethnicity} \;+\; \alpha_{religion} \;+\; \alpha_C * \alpha_t \;+\; \beta_1\, ethnic\_distance_{ie} \\ +\; \beta_2\, ELF_i \;+\; \beta_3\, X_{it} \;+\; \beta_4\, X_i \;+\; \epsilon_{iet} \end{aligned} \tag{3.6}$$

where $Prob(death_{iet})$ is the probability of death of child 'i' born to mother belonging to ethnicity 'e' in year 't'. The $ethnic\_distance_{ie}$ variable is our primary variable of interest and it gives the linguistic distance of the mother of child 'i' belonging to ethnicity 'e' from people living within circles of different radii around her. The ELF_i variable gives the ethno-linguistic fractionalization in the circles of different radii around the mother. For calculating both the linguistic distance and the ELF we have used circles of different radii, viz. 25, 50, 75, 100, 125, 150, 175, 200, 250 km around the mother. In some specifications we also calculate ELF at the district level.

The variables X_{it} and X_i come from the literature on child mortality and have been found to be important for child mortality.[22] X_{it} includes birth specific variables viz. female child dummy, age at birth, age at birth squared, multiple birth, birth order, birth order squared, short birth space prior, short birth space post. X_i includes mother specific variables viz. urban dummy, education years, and dummies for the wealth index. We also control for the geographic distance, which gives the geographic isolation of the mother from everybody else in the country and the distance of the mother's location from the capital, which are both included in X_i.

We include region fixed effects α_R apart from country specific time effects $\alpha_C * \alpha_t$ which allows for the non-parametric evolution of time effects differently for each country. By including ethnicity fixed effects, $\alpha_{ethnicity}$, we control for unobserved heterogeneity across ethnic groups, which allows us to identify the effect of ethnic distance on child mortality that is not driven by ethnicity specific characteristics like ethnic dominance of certain groups, cultural differences leading to differences in health practices between different

[22]See Kudamatsu [2009], Baird et al. [2011], Franck and Rainer [2012] for example.

groups, etc. Likewise $\alpha_{religion}$, the religion dummy controls for differences in religious beliefs and practices across different individuals.

Since child mortality is a 0-1 binary variable we estimate it via pooled probit regressions.[23] β_1 is our coefficient of interest since it gives the effect of linguistic distance of the mother on the probability of death of the child. Given the possibilities of endogeneity giving a causal interpretation to β_1 is not straightforward. Moreover, we cannot use mother specific fixed effects since the ethnic distance variable does not vary across time for the same mother. However, we are able to control for a host of maternal and birth characteristics which alleviate endogeneity concerns to a great extent. Moreover, we later do some analysis to gauge how much selection on unobservables is taking place and this increases our faith in the causal interpretation of β_1. The standard errors are clustered at the Survey-Country level.

3.5 Results

3.5.1 Ethnic distance and Child mortality

Most studies look at how ethnic diversity affects different economic outcomes including infant/child mortality at the aggregate country/ district/ or town level. In this paper we study the effects of individual level ethnic distance on child mortality at the individual level while controlling for ethnic diversity at the aggregate level. We are able to construct measures of ethnic diversity taking into account the exact location of the individual. Also, Child mortality is an actual outcome variable and thus instead of just provision it represents actual access to health services. Our focus is thus on access and not merely provision.

Table 3.1 gives our baseline specification. In Table 3.1 we try to explain how the linguistic distance of the mother from people living around her, affects the probability of her child dying before reaching the age of five. In this baseline specification we consider the average distance of the mother from all individuals living in a radius of 75 km around her.

In the different columns of Table 3.1 we keep adding different control variables. In column 1 we do not control for any other variables apart from the linguistic distance variable. In column 2, we add a control for linguistic fractionalization. We notice that without any other controls neither linguistic distance nor fractionalization is significant. In column 3 we add individual level controls including wealth and education of the

[23] As a robustness check we also use linear probability models, but as we will see later our results remain unchanged.

TABLE 3.1: Child mortality: Baseline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
linguistic distance 75	0.0502	0.0511	0.0862*	0.0326	0.102***	0.103***	0.0779**
	(0.0559)	(0.0580)	(0.0470)	(0.0236)	(0.0340)	(0.0342)	(0.0392)
fractionalization 75		-0.0113	-0.0646	-0.0149	-0.0375*	-0.0442**	-0.0233
		(0.0741)	(0.0642)	(0.0210)	(0.0197)	(0.0217)	(0.0244)
urban			-0.0903***	-0.0692***	-0.0719***	-0.0690***	-0.0474*
			(0.0317)	(0.0154)	(0.0149)	(0.0149)	(0.0246)
female			-0.0654***	-0.0674***	-0.0678***	-0.0678***	-0.0701***
			(0.00435)	(0.00433)	(0.00433)	(0.00433)	(0.00660)
education_years			-0.0363***	-0.0221***	-0.0212***	-0.0211***	-0.0180***
			(0.00532)	(0.00253)	(0.00248)	(0.00247)	(0.00280)
wealth index 2			0.0313	0.0107	0.00733	0.00732	0.0253
			(0.0190)	(0.0144)	(0.0141)	(0.0141)	(0.0195)
wealth index 3			0.0129	-0.00593	-0.0106	-0.0103	0.0131
			(0.0210)	(0.0219)	(0.0212)	(0.0213)	(0.0283)
wealth index 4			-0.0235	-0.0552**	-0.0592**	-0.0587**	-0.0218
			(0.0251)	(0.0260)	(0.0250)	(0.0250)	(0.0313)
wealth index 5			-0.0996***	-0.152***	-0.156***	-0.152***	-0.105***
			(0.0365)	(0.0274)	(0.0255)	(0.0251)	(0.0317)
log distance to capital						0.0172**	0.00699
						(0.00774)	(0.0164)
ln_geog_dist						-0.0577	-0.0895*
						(0.0413)	(0.0507)
Individual controls	No	No	Yes	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes	Yes
N	658755	658755	658505	649182	648468	648468	263730
pseudo R^2	0.000	0.000	0.062	0.084	0.086	0.086	0.091

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated. The first 6 columns use the full sample i.e. all the births in the maternal history of the mother. In column 7 we restrict the sample to only births occurring within 10 years prior to the survey. The individual controls include age at birth, age at birth squared, multiple births (twins, triplets etc.), birth order, birth order squared, short birth space prior to birth, and short birth space post birth.

mother, sex of the child, location of the mother, and other individual specific variables that affect child mortality like birth order etc.[24] In column 4 we add region fixed effects and country-time fixed effects. In column 5 we add religion and ethnicity fixed effects and in column 6 we add controls for geographic isolation.

Column 6 is our most complete and preferred specification. The inclusion of ethnicity fixed effects allows us to control for heterogeneity in unobservable characteristics across ethnic groups including heterogeneity in health outcomes across ethnic groups, and differences in cultural practices between different groups etc. We notice that once religion and ethnicity fixed effects are controlled for (in both columns 5 and 6), our linguistic distance variable becomes highly significant. The linguistic distance of the mother from people living around her significantly increases the probability of her child dying before

[24]The wealth index variable is an ordinal index that takes five different values with a higher value indicating a higher level of wealth.

reaching the age of five. Contrary to previous findings in the literature, the fractionalization variable on the other hand significantly reduces the probability of a child dying before reaching the age of five.

The rest of the variables all have the expected signs and effects. Female children, urban children and children whose mothers are more educated have a lower probability of dying. Again, children whose mothers belong to either of the two highest wealth quintiles have a lower probability of dying. Geographic distance to the capital increases the probability of dying, but geographic isolation does not significantly affect the probability of death.

In the first six columns of Table 3.1 we have included all the births in the maternal history of the mother. The DHS are a rich source of data on individual level health outcomes. However, like any other data sources there are some shortcomings of the data which have been discussed in Baird et al. [2011] for example. One of the primary concerns in using retrospective data is recall bias which stems from the fact that women might be less likely to accurately remember more distant births and deaths. To minimize recall bias in some of our specification we use births and deaths occuring in the preceding ten years from the date of the survey. Both Baird et al. [2011] and Kudamatsu et al. [2012] resort to this strategy as well.

Moreover, since some of the variables like wealth level are available only for the survey year, they might more accurately represent current conditions than conditions long ago in the past. As a robustness check in columns 7 we restrict the sample to only births occurring within 10 years prior to the survey. The results become slightly less significant but overall they remain qualitatively unchanged. The linguistic distance variable continues to significantly increase the probability of child death. Fractionalization on the other hand becomes insignificant.[25]

In the tables above, so far we have considered the coefficients from the regressions. It is clear that the coefficient for the linguistic distance variable is statistically significant, but is it economically significant as well? In order to gauge the magnitude of the effect of linguistic distance on the probability of child death we now present the average marginal effects. The average marginal effect of an increase in linguistic distance of the mother from her neighbours within a circle of 75 km is of 2.8%. A one standard deviation (0.267) increase in the linguistic distance of the mother from her neighbours increases the probability of her children dying by around 0.75%. If we consider only the births and deaths that took place within the last 10 years of the survey the marginal effects are somewhat smaller. Considering the births that took place within the last 10 years, the average marginal effect of an increase in linguistic distance of the mother from her

[25] In appendix Table C.3 we replicate the first six columns of Table 3.1 using only the births occurring in the last 10 years

neighbours within a circle of 75 km is of 2.1%. In other words, a one standard deviation increase in linguistic distance leads to a 0.57% increase in the child's probability of dying.

Looking at the average marginal effects of fractionalization in the radius of 75 km we see that if we move from the completely homogenous (fractionalization = 0) to a completely heterogeneous (fractionalization = 1) location, the probability of child death falls by 1.1%. And a one standard deviation increase in fractionalization leads to a fall of 0.3% in the probability of child death. Again considering only the births from the last 10 years prior to the survey, the average marginal effect of fractionalization is of 0.4% only. And a one standard deviation increase in fractionalization leads to a fall of 0.1 % in the probability of child death.

By themselves these marginal effects look very small. But to put them in perspective let us compare these average marginal effects with those of some of the other important variables. For instance a one year increase in the mother's education leads to only a 0.5% decrease in the probability of child death. Moving from a rural to urban location reduces the probability of child death by only 1.8%. Moving to live one standard deviation closer to the capital decreases the probability of death by only 0.6%. Thus we notice that relatively speaking the average marginal effect of linguistic distance is economically important when compared to some of the other variables. The marginal effect of fractionalization on the other hand is much smaller.

In all the analyses up to now we have only considered the radius of 75 km to calculate the mother's ethnic distance. Next in Table 3.2 we verify if our results are robust for alternate radii or not. In the different panels of Table 3.2 we have exactly the identical specifications as in Table 3.1 except that in each panel we have a different radius of the circle around the mother. To start with in panel 1 we investigate the effects of linguistic distance of the mother from people living in a radius of 25 km around her. Subsequently in panel 2, 3, 4, 5, 6, 7, 8, 9 the radii are 50, 75, 100, 125, 150, 175, 200, 250 km respectively.

As in the previous table we see that the linguistic distance of the mother from her neighbours significantly increases the probability of death of the child before reaching the age of five. This holds true for distance calculated in circles ranging from radius 25 km to 125 km, beyond which the linguistic distance variable is not significant. The results are robust to the inclusion of region, ethnicity, country-time fixed effects, individual specific controls, geographic isolation and the popularly used measure of ethno-linguistic fractionalization. Fractionalization whenever it is significant continues to reduce the

Table 3.2: Child mortality: Alternative radii

	(1)	(2)	(3)	(4)	(5)	(6)
linguistic distance 25	0.0439	0.0437	0.0833*	0.0290	0.0648***	0.0648***
	(0.0536)	(0.0573)	(0.0472)	(0.0223)	(0.0222)	(0.0219)
fractionalization 25		0.00163	-0.0490	-0.0173	-0.0329**	-0.0337**
		(0.0514)	(0.0443)	(0.0159)	(0.0146)	(0.0149)
linguistic distance 50	0.0476	0.0473	0.0856*	0.0315	0.0860***	0.0866***
	(0.0555)	(0.0584)	(0.0476)	(0.0231)	(0.0276)	(0.0274)
fractionalization 50		0.00275	-0.0544	-0.0112	-0.0290*	-0.0325*
		(0.0630)	(0.0547)	(0.0183)	(0.0170)	(0.0180)
linguistic distance 75	0.0502	0.0511	0.0862*	0.0326	0.102***	0.103***
	(0.0559)	(0.0580)	(0.0470)	(0.0236)	(0.0340)	(0.0342)
fractionalization 75		-0.0113	-0.0646	-0.0149	-0.0375*	-0.0442**
		(0.0741)	(0.0642)	(0.0210)	(0.0197)	(0.0217)
linguistic distance 100	0.0519	0.0533	0.0846*	0.0302	0.102***	0.105***
	(0.0562)	(0.0576)	(0.0465)	(0.0237)	(0.0353)	(0.0358)
fractionalization 100		-0.0225	-0.0748	-0.0162	-0.0432**	-0.0525***
		(0.0865)	(0.0733)	(0.0208)	(0.0171)	(0.0195)
linguistic distance 125	0.0526	0.0545	0.0826*	0.0289	0.102**	0.104**
	(0.0557)	(0.0567)	(0.0456)	(0.0235)	(0.0429)	(0.0434)
fractionalization 125		-0.0419	-0.0873	-0.0129	-0.0415**	-0.0526***
		(0.100)	(0.0843)	(0.0246)	(0.0178)	(0.0198)
linguistic distance 150	0.0510	0.0530	0.0798*	0.0257	0.0832	0.0856
	(0.0560)	(0.0566)	(0.0454)	(0.0235)	(0.0530)	(0.0535)
fractionalization 150		-0.0537	-0.0971	0.00527	-0.0246	-0.0366
		(0.117)	(0.0975)	(0.0279)	(0.0214)	(0.0235)
linguistic distance 175	0.0496	0.0520	0.0779*	0.0236	0.0759	0.0779
	(0.0563)	(0.0565)	(0.0450)	(0.0233)	(0.0634)	(0.0642)
fractionalization 175		-0.0695	-0.109	0.0215	-0.0111	-0.0222
		(0.137)	(0.113)	(0.0296)	(0.0250)	(0.0266)
linguistic distance 200	0.0489	0.0515	0.0768*	0.0219	0.0690	0.0700
	(0.0569)	(0.0566)	(0.0451)	(0.0230)	(0.0730)	(0.0739)
fractionalization 200		-0.0835	-0.124	0.0439	0.00843	-0.00197
		(0.159)	(0.130)	(0.0334)	(0.0305)	(0.0315)
linguistic distance 250	0.0476	0.0487	0.0720	0.0228	0.0574	0.0566
	(0.0591)	(0.0575)	(0.0459)	(0.0231)	(0.0827)	(0.0837)
fractionalization 250		-0.123	-0.183	0.0142	-0.0399	-0.0486
		(0.203)	(0.162)	(0.0553)	(0.0578)	(0.0558)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	658755	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

probability of child death.[26]

From the above analysis it is clear that the linguistic distance of the mother from people living around her is a significant and robust predictor of child mortality. Contrary to the findings in the previous literature, fractionalization on the other hand is not a robust determinant of child mortality. Moreover, in specifications where the fractionalization variable is significant it actually reduces the probability of child death.

[26]In appendix Table C.4 we restrict our sample to only births that took place in the last 10 years prior to the interview. Our results remain qualitatively unchanged for the linguistic distance variable. Fractionalization becomes insignificant.

3.5.2 Fractionalization

Ethnic diversity usually measured by ELF or Ethno-linguistic fractionalization has received a lot of attention in the literature. Such fractionalization has often been found to have a negative effect on different socio-economic outcomes.[27] However, as seen from the previous sections, in our data, fractionalization if anything has a positive effect on child mortality, i.e. it reduces child mortality. We relate this to the recent findings by Ashraf and Galor [2013], who point out that diversity could have both beneficial and detrimental effects. Following them we argue that that diversity could potentially reflect a higher stock of knowledge and information, and thus lead to better health outcomes. In this section we focus specifically on verifying whether this is a robust result. In Table 3.3 we use the identical specifications from the previous section but do not control for linguistic distance. In all but the last panel of Table 3.3 we control for fractionalization calculated for circles of different radii. As a robustness test in the last panel of Table 3.3 we also control for fractionalization at the district level. We notice that even when we do not explicitly control for linguistic distance, the effect of linguistic fractionalization on child mortality is still negative and significant.

Table 3.3: Child mortality: Fractionalization in different radii

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5
fractionalization 25	0.00726	-0.0380	-0.0133	-0.0285**	-0.0293**
	(0.0484)	(0.0418)	(0.0164)	(0.0144)	(0.0148)
fractionalization 50	0.00762	-0.0453	-0.00738	-0.0232	-0.0266
	(0.0610)	(0.0534)	(0.0189)	(0.0164)	(0.0174)
fractionalization 75	-0.00697	-0.0571	-0.0119	-0.0299	-0.0363*
	(0.0729)	(0.0636)	(0.0216)	(0.0188)	(0.0208)
fractionalization 100	-0.0187	-0.0687	-0.0142	-0.0351**	-0.0441**
	(0.0858)	(0.0732)	(0.0205)	(0.0165)	(0.0188)
fractionalization 125	-0.0386	-0.0821	-0.0111	-0.0333*	-0.0440**
	(0.0999)	(0.0845)	(0.0238)	(0.0173)	(0.0193)
fractionalization 150	-0.0506	-0.0923	0.00733	-0.0179	-0.0294
	(0.117)	(0.0979)	(0.0268)	(0.0211)	(0.0230)
fractionalization 175	-0.0664	-0.104	0.0241	-0.00494	-0.0156
	(0.137)	(0.114)	(0.0283)	(0.0257)	(0.0271)
fractionalization 200	-0.0804	-0.119	0.0472	0.0143	0.00405
	(0.159)	(0.131)	(0.0320)	(0.0318)	(0.0326)
fractionalization 250	-0.121	-0.181	0.0188	-0.0352	-0.0439
	(0.203)	(0.163)	(0.0529)	(0.0603)	(0.0583)
frac_district	-0.00138	-0.0465	-0.0250	-0.0443**	-0.0448**
	(0.0445)	(0.0408)	(0.0234)	(0.0204)	(0.0204)
N	658755	658505	649182	648468	648468
pseudo R^2	0.000	0.062	0.084	0.086	0.086

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

[27]See Alesina et al. [2003], Alesina et al. [1999], Easterly and Levine [1997] for example. See Alesina and La Ferrara [2000] for a review of the literature.

Some recent papers like Montalvo and Reynal-Querol [2005] stress the importance of polarization rather than fractionalization particularly in the context of intergroup conflict. Thus, we try to understand whether the effects of polarization on child mortality are different from those on fractionalization. In Appendix Table C.5 we control for polarization in the district along with the ethnic distance variable calculated at the circle level for circles of different radii. We notice that the effects of polarization are quite similar to those of fractionalization. In other words, regardless of whether we measure diversity by the index of fractionalization or of polarization, ethnic diversity if anything seems to reduce child mortality.

So far in all the above tables the two measures of diversity have entered separately in the different specifications and we have found that both fractionalization and polarization have a negative and significant effect on child mortality. In Appendix Table C.7 we control for ELF and polarization together with the linguistic distance variable for different radii and in Appendix Table C.6 we do the same without the linguistic distance variable. When both the measures of diversity enter the specifications together, polarization becomes insignificant, while fractionalization continues to have a negative and significant effect on child mortality.

3.5.3 Heterogenous effects

Up to this point we have assumed that the linguistic distance variable has a homogenous effect on the children of all mothers. However, there are several reasons why this might not be the case. For example, linguistic distance might have a different effect in places which are more fractionalized than places which are not. Again wealthier and more educated mothers might be better able to insulate their children from the negative effects of linguistic distance. Again linguistic distance might have different implications for male and female children. In this section we thus try to identify the heterogeneity in impacts of linguistic distance by focussing on different variables like education, wealth, gender, place of residence (urban or rural), and distance from the capital.

In Table 3.4, we first try to identify if linguistic distance variable has a heterogeneous impact on the probability of child death by the wealth level of the mother. The wealth index variable takes five different values with a higher value indicating a higher level of wealth. In Table 3.4 we use our baseline specification of column 6 from Table 3.1 but add to it the interaction term of linguistic distance and wealth level. We notice that linguistic distance continues to be highly significant and positive in sign. Also higher wealth reduces mortality. As expected we see the negative effects of linguistic distance are partially reduced by higher wealth levels. This is particularly true for the wealth

Table 3.4: Heterogeneous effects of linguistic distance by Wealth level

	ling dist 25	ling dist 50	ling dist 75	ling dist 100	ling dist 125
linguistic distance	0.1211***	0.1439***	0.1564***	0.1547***	0.1516***
	(0.0362)	(0.0407)	(0.0469)	(0.0487)	(0.0566)
wealth index					
2	0.0110	0.0112	0.0102	0.0098	0.0095
	(0.0151)	(0.0152)	(0.0151)	(0.0151)	(0.0151)
3	0.0002	0.0008	0.0002	-0.0002	-0.0007
	(0.0225)	(0.0226)	(0.0224)	(0.0223)	(0.0222)
4	-0.0527**	-0.0518**	-0.0514**	-0.0513**	-0.0518**
	(0.0256)	(0.0258)	(0.0259)	(0.0259)	(0.0258)
5	-0.1401***	-0.1402***	-0.1410***	-0.1416***	-0.1418***
	(0.0258)	(0.0259)	(0.0258)	(0.0258)	(0.0258)
Interaction					
2	-0.0282	-0.0294	-0.0218	-0.0186	-0.0153
	(0.0248)	(0.0242)	(0.0233)	(0.0228)	(0.0227)
3	-0.0876**	-0.0887**	-0.0822**	-0.0777**	-0.0722**
	(0.0435)	(0.0419)	(0.0406)	(0.0397)	(0.0388)
4	-0.0508	-0.0558	-0.0557	-0.0544	-0.0410
	(0.0455)	(0.0462)	(0.0466)	(0.0464)	(0.0457)
5	-0.0964**	-0.0918*	-0.0829**	-0.0755	-0.0715
	(0.0473)	(0.0493)	(0.0496)	(0.0498)	(0.0507)
N	648468	648468	648468	648468	648468
pseudo R^2	0.0861	0.0861	0.0861	0.0861	0.0861

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

levels 3 and 5 compared to the omitted wealth category 1. This means the mothers reach the level of wealth 3 they are able to reduce the negative effects of linguistic distance. One possible explanation could be that if linguistic distance hinders access to public information on healthcare practices or public health facilities, wealthier mothers could perhaps still have access to privately provided health care.

In Table 3.5, we analyse the heterogeneity of the effects on linguistic distance by fractionalization in the circles. We see that while the linguistic distance variable continues to be significant, the fractionalization variable becomes insignificant. The interaction term between the two variables mostly has a negative sign but is almost always insignificant. In the results provided the interaction term is significant only for circles of 100 km.[28] The negative sign of the interaction term implies that if anything in the fractionalized places, the positive effects of fractionalization reduces the negative effects of linguistic distance.

We also have various specifications in which we try to identify the heterogeneity in the effects of linguistic distance by education, gender, place of residence (urban or rural),

[28]The interaction term is also negatively significant for circles of 150, 175 and 200 km at the 10% level. Results not provided and are available upon request.

TABLE 3.5: Heterogenous effects of linguitic distance by fractionalization

	ling dist 25	ling dist 50	ling dist 75	ling dist 100	ling dist 125
linguistic distance	0.0606*	0.1000**	0.1511***	0.1771***	0.1871**
	(0.0357)	(0.04563)	(0.0558)	(0.0687)	(0.0818)
fractionalization	-0.03457*	-0.0280	-0.0328	-0.0376	-0.0363
	(0.0201)	(0.0215)	(0.0253)	(0.0243)	(0.0257)
Interaction	0.0126	-0.0326	-0.0993	-0.1406*	-0.1563
	(0.0710)	(0.0813)	(0.0774)	(0.0833)	(0.0981)
N	648468	648468	648468	648468	648468
pseudo R^2	0.0861	0.0861	0.0861	0.0861	0.0861

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

and distance from the capital. We do not find any heterogeneity in terms of these variables.[29]

3.5.4 Migration

A possible concern in estimating the effects of ethnic distance on child mortality is the possibility of individuals to migrate. One could argue that if individuals know that they incur a cost by living with people who are ethnically distant from them then they might choose to move to neighbourhoods which have people who are ethnically more similar to them. Thus there might be some sort of Tiebout sorting of individuals. However, quite realistically if we introduce some sort of transportation costs in the individual's decision to move then perfect sorting would not happen. In fact perfect sorting is also not observed in reality since there are various barriers to movement. However, if individuals actually are able to move to places where they are less distant to others then if anything we are underestimating the effects of ethnic distance on child mortality and the effects of distance would be stronger.

In order to ensure that migration is not driving our results, in Table 3.6 we provide specifications restricting our sample to mothers who have always lived in their current place of residence. Thus we are looking at the variation in linguistic distance within the sample of those individuals who have never migrated.[30] However, this information is not available for all countries and we lose a lot of our observations when we use the variable that indicates for how long the individuals have been residing in the current place of residence. Our results remain qualitatively similar. Linguistic distance of the mother

[29] Results not provided and are available upon request.
[30] Of course this sample itself might not be random since if there is the possibility of migration, this sample represents individuals who have not been able to make use of the that possibility.

TABLE 3.6: Child mortality: No migration sample

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6
linguistic distance 25	0.0173	0.00895	0.0431	0.0373*	0.0945***	0.0944***
	(0.0475)	(0.0527)	(0.0414)	(0.0221)	(0.0366)	(0.0353)
fractionalization 25		0.0421	-0.0103	0.00194	-0.0117	-0.0101
		(0.0618)	(0.0534)	(0.0205)	(0.0199)	(0.0197)
linguistic distance 50	0.0221	0.0158	0.0470	0.0397*	0.126***	0.127***
	(0.0485)	(0.0526)	(0.0415)	(0.0236)	(0.0452)	(0.0438)
fractionalization 50		0.0426	-0.0122	0.0170	-0.00758	-0.00852
		(0.0727)	(0.0632)	(0.0236)	(0.0225)	(0.0228)
linguistic distance 75	0.0265	0.0233	0.0510	0.0457*	0.164***	0.167***
	(0.0483)	(0.0508)	(0.0403)	(0.0248)	(0.0612)	(0.0600)
fractionalization 75		0.0323	-0.0192	0.0292	-0.00901	-0.0139
		(0.0854)	(0.0745)	(0.0307)	(0.0259)	(0.0268)
linguistic distance 100	0.0273	0.0258	0.0497	0.0440*	0.158***	0.163***
	(0.0483)	(0.0499)	(0.0398)	(0.0251)	(0.0611)	(0.0606)
fractionalization 100		0.0225	-0.0283	0.0415	0.000472	-0.00883
		(0.0979)	(0.0843)	(0.0354)	(0.0275)	(0.0288)
linguistic distance 125	0.0293	0.0290	0.0498	0.0463*	0.182**	0.188**
	(0.0481)	(0.0492)	(0.0397)	(0.0255)	(0.0769)	(0.0763)
fractionalization 125		0.00729	-0.0382	0.0351	-0.00534	-0.0197
		(0.112)	(0.0956)	(0.0468)	(0.0356)	(0.0365)
linguistic distance 175	0.0293	0.0294	0.0481	0.0419	0.194*	0.201*
	(0.0482)	(0.0488)	(0.0397)	(0.0255)	(0.116)	(0.116)
fractionalization 175		-0.00415	-0.0501	0.0464	0.00711	-0.0120
		(0.150)	(0.127)	(0.0617)	(0.0460)	(0.0478)
linguistic distance 200	0.0306	0.0307	0.0484	0.0388	0.189	0.193
	(0.0484)	(0.0488)	(0.0400)	(0.0251)	(0.139)	(0.139)
fractionalization 200		-0.00187	-0.0540	0.0686	0.0265	0.00895
		(0.174)	(0.146)	(0.0743)	(0.0575)	(0.0591)
linguistic distance 250	0.0320	0.0319	0.0465	0.0376	0.160	0.159
	(0.0500)	(0.0498)	(0.0415)	(0.0259)	(0.150)	(0.150)
fractionalization 250		-0.00582	-0.0839	-0.00314	-0.0620	-0.0737
		(0.220)	(0.181)	(0.118)	(0.0992)	(0.0953)
N	239852	239852	239735	235642	235311	235311
pseudo R^2	0.000	0.000	0.061	0.084	0.086	0.086

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

continues to significantly increase the probability of child mortality. Fractionalization on the other hand becomes insignificant.

3.5.5 Selection on Unobservables

Like most empirical studies, the single biggest challenge in the literature on ethnic diversity has been to establish a causal relationship of ethnic diversity with the different socio-economic variables of interest. Like in our case, ethnic diversity is almost always a time invariant measure and observed at a particular point of time. Moreover, there are issues of measurement and omitted variable bias leading to non-causal estimates. Thus most of the literature has to content with finding a correlation.[31]

[31]See Ahlerup [2009] for a discussion on the endogeneity of ethnic diversity.

The biggest challenge in our paper is the possibility of endogeneity arising out of omitted variable bias due to unobserved maternal and family characteristics. We cannot incorporate mother fixed effects since our main variable of interest "linguistic distance" remains unchanged for the mother over time. We are however, able to control for a wide range of individual specific variables that allow us to control for a number of factors that could lead to endogeneity. We also follow the methodology developed by Altonji et al. [2005] and Bellows and Miguel [2009] who present new estimation strategies that can be used when strong prior information regarding the exogeneity of the variable of interest is unavailable. They use the selection on the observables to gauge how strong the selection is on unobservables.

We do not have any clearly exogenous instruments at our disposal which would allow us to cleanly identify the effects of linguistic distance on child mortality. Thus like several recent papers e.g. Nunn and Wantchekon [2009], we exploit the measure and intuition provided by Altonji et al. [2005] and use the strength of selection on observables to assess the potential bias arising from selection on unobservables. In fact, following the methodology developed by Altonji et al. [2005], we can arrive at ratio that tells us how much stronger the selection on unobservables must be, relative to selection on observables, to explain away the full estimated effect of linguistic distance on child mortality. Table 3.7 gives us the corresponding ratios for circles of radii 25 km to 125 km for both the full sample and the sample consisting of births and deaths in the last ten years prior to the survey.

To calculate each of the ratios in Table 3.7 we need two regressions. We need one regression with a restricted set of controls (or no controls) and another with a full set of controls. Let β^R be the coefficient of linguistic distance from the restricted regression and β^F be the coefficient from the regression with the full set of controls. Then the ratio $\beta^F/(\beta^R - \beta^F)$ is our ratio of interest. The smaller the denominator, the less the selection on observables and thus the selection on unobservables need to be stronger in order to explain away the entire effect of linguistic distance. Again, the selection on unobservables has to explain away more the higher the numerator is in magnitude [Nunn and Wantchekon, 2009].

Table 3.7: Selection on Unobservables/Observables

	Full Sample	Last 10 years
25	-4.24	4.70
50	-2.64	12.99
75	-2.22	-10.42
100	-2.27	-50.89
125	-2.32	4.44

In the restricted regression we consider no controls while regressing child mortality on linguistic distance. In the unrestricted regression on the other hand we consider the full set of controls from our baseline regression. All the ratios in Table 3.7 are quite big. In the full sample we notice that the selection on unobservables have to be at least two times larger than that on observables while in some cases it would need to be at least more than four times larger. Again, in the sample from the last ten years we see that the selection on unobservables has to be at least more than four and a half times larger than that on observables while in some cases it would need to be at least more than fifty times. Thus it is highly unlikely that selection on unobservables will explain away the entire effect that we are currently attributing to our linguistic distance variables. Thus, we can argue that linguistic distance indeed increases the probability of child death and this relation is not driven by omitted variable bias.

3.5.6 Other Robustness checks

In this section we subject our results to several additional robustness checks. Our primary variable of interest is the probability of child death which is a 0-1 binary variable. Thus, so far we have used probit regressions in order to estimate the effects of linguistic distance of the mother on child mortality. First, we verify if our results hold true if we use Linear Probability Model (LPM) instead. Hence we rerun our baseline regressions from Table 3.1 using OLS instead of probit regressions. The results from LPM are presented in Table C.10. We notice that our results remain qualitatively unchanged. Linguistic distance continues to significantly increase the probability of child death whereas fractionalization continues to reduce it.

Next, we try to identify if either the linguistic distance variable or the fractionalization variable have a non-linear effect on child mortality. In Table C.8 we present the identical specification used in the complete specifications of the baseline table, but add to it non-linear terms for both fractionalization and linguistic distance in different combinations. We notice that on including nonlinear terms, while the results become less significant in some of the specifications, but overall the results remain similar to what we found in the previous sections. Linguistic distance continues to increase the probability of child death and there is some weak evidence that this effect might be nonlinear and after a certain level it might actually reduce mortality. We do not observe any consistent results for fractionalization.[32]

So far our focus has been on child mortality which is the event that the child dies before reaching the age of five. However, a lot of the literature has focused on Infant

[32] In Table C.9 we consider only the births from the last 10 years. The results remain qualitatively similar.

mortality which is defined as the child dying before reaching the age of 1. In Table C.11 we try out our baseline specifications of child mortality on infant mortality to verify whether linguistic distance has similar negative effects on infant mortality as it does on child mortality. Qualitatively the impact of the linguistic distance of the mother on the probability of mortality of the child before reaching the age of one is similar to that on the probability of mortality of the child before reaching the age of five. However, the results are a lot less significant. When we consider circles of 75 km radius around the mother the linguistic distance variable is significant at level of significance 10%. In Table C.11 apart from this we also see that the fractionalization variable in the complete specification of column 6 continues to reduce infant mortality as it did child mortality

There might be two interpretations for why the linguistic distance variable is a lot less significant for infant mortality. First, "Infant mortality is a rare event and estimating it requires a large number of observations" [Kudamatsu, 2009]. Second, infant mortality has a lot to do with conditions during birth and need not have anything to do with ethnic distance. However, as the child grows the access to certain services might become more crucial and thus child mortality has more to do with linguistic distance than infant mortality.

3.6 Conclusion

There exist huge disparities in child mortality rates across ethnic groups in Africa. The literature so far has not convincingly addressed the reasons behind such ethnic inequality in child mortality rates in Africa. Using high quality DHS data on individual level health outcomes and combining it with a novel dataset on the spatial distribution of ethnic groups at the 1 km^2 level, we estimate the effects individual level ethnic distance and diversity on individual level child health outcomes. The exact ethnic distance variable is calculated using the linguistic distance between individuals. We find that children of mothers who are ethnically distant from their neighbours have a higher probability of dying before reaching the age of five, whereas children of mothers living in more ethnically fractionalized places on the other hand have a lower probability of dying before reaching the age of five. Since we control for ethnicity specific fixed effects our results are not driven by heterogeneity in unobservable characteristics across ethnic groups including heterogeneity in health outcomes across ethnic groups or cultural differences between ethnic groups. To alleviate endogeneity concerns we control for a host of variables that reduce the possibilities of omitted variable bias. Then using recently developed methods by Altonji et al. [2005] we use the selection on the observables to gauge how strong the selection is on unobservables. We do not find evidence of a strong selection on unobservables.

While there are several possible explanations for our results, we follow Ashraf and Galor [2013] and argue that on the one hand, diversity implies a higher stock of knowledge and information, and thus leads to better health outcomes. On the other hand, such knowledge does not flow smoothly to groups which are linguistically distant and thus such groups lose out. One of the biggest issues in child rearing and child deaths is the lack of information. The probability of the survival of the child depends on the number of mistakes the mother makes in taking care of the child. Thus if an individual is very distant linguistically she has less access to the information on how to take care of her child thus commits more mistakes. However, there are several other ways to interpret our results. Our data does not allow us to pin down the exact mechanisms via which linguistic distances affect child mortality. Identifying these channels precisely would be a good avenue of future research.

Appendix A

Appendix for Chapter 1

A.1 Data appendix

Naxalite/Maoist incidents:
The data on the Maoist incidents comes from four different sources:

- Global Terrorism Database (GTD) I: 1970-1997 & II: 1998-2004: In fact now there is one consolidated GTD which has data till 2007.
- Rand-MIPT Terrorism Incident database (1998- present)
- Worldwide Incidents tracking system (WITS), National Counter Terrorism Centre (2004-2007)
- South Asia Terrorism Portal (SATP) maintains a comprehensive web portal tracking all terrorism incidents in South Asia. It has most detailed accounts of the major Naxalite incidents from 2005 onwards.

All the above mentioned sources track data on terrorist/violent incidents from different sources like newspapers, official reports etc.[1] The data is across countries often with the village/city/district name of the place where each particular incident took place specified. We have filtered out all the data pertaining to India and then kept only the incidents that were clearly identifiable as Maoist/Naxalite in nature. To do so we used the name of perpetrator unless it was clearly mentioned as being a Maoist/Naxalite incident. Then except for the cases where the district is clearly mentioned the town or village mentioned was placed in its corresponding district. More often than not we were successful in doing so. However, we had to leave out the cases where it was not possible

[1] The first three sources have been used by Iyer [2009]

to do identify the districts. Finally, we have consolidated all the above data sources to construct a comprehensive consolidated district level Maoist incidents database. Thus, there is information on presence, incidents and the number of deaths and injuries at the district level from 1979-2009.

Consumption: District-specific indicators of income or expenditure are not available for India. Instead, per capita consumption expenditure is calculated from the National Sample Survey (hereafter NSS) data. The NSS conducts every 5 years the Consumption expenditure survey (among other surveys) whence district specific per capita consumption expenditures are calculated. The last four NSS thick rounds [2] were undertaken in 1987-88 (43rd), 1993-94 (50th), 1999-00 (55th), 2004-05 (61st). Given that district specific data is used, the 50th round is unusable since there are no district identifiers. For the purposes of this paper we use the other three rounds. The per capita consumption expenditure is calculated for all the districts using the Marginal per capita consumption (MPC hereafter), using the 30 day recall period since the 30 day recall period is the only one that is common across rounds. The MPC is used as a proxy for per capita income.

Demography: Apart from the Sample surveys by the NSS, India conducts a country wide census every 10 years. The data on demography and public goods are stored in the so called village directory data and is available for each census year since 1961. The 2001 census is used, whence we have data on total population, Schedule tribe and Caste population, population by religious groups. All the above information is available at the village level which has been used to construct the corresponding district level numbers.

Geography: We control for geographical terrain by the fraction of the districts uncultivated area that is barren rocky, sandy and steep sloping. We also control for remoteness by the log of the distance from the state capital. Such data is made available in the Wasteland atlas of India, Department of Land Resources (Ministry of Rural development) in collaboration with the National Remote Sensing agency, (Department of Space). These variables by definition do not vary over time. We also have data on forest cover in the different districts. The data comes from the various "State of forest cover" reports of the Forest Survey of India (FSI) and gives the percentage of forest cover in each district. The 2005 data is used.

Land distribution: The data on operational holdings of agricultural land comes from the Agricultural census of India which collects such data every 5 years. The data gives the number of operated landholdings in the different size classes (viz. Small, semi medium, medium and, large). Both the 1991 and the 2001 censuses are available. The gini coefficient for land inequality is calculated using this data. The computed gini coefficient varies from 0.12 to 0.78 and 0.14 to 0.79 in 1991 and 2001 respectively. The average inequality has gone up from 0.47 in 1991 to 0.5 in the year 2001 (data on Bihar not available in 2001). The land inequalities across the 2 years are highly correlated

[2] only the thick rounds can be used since the thin rounds have too few observations

with Correlation coefficient of 0.97. We use the 1991 data since that way we have data prior to the high conflict years and also no major state is missing.

Colonial Land Institutions: This data comes from Banerjee and Iyer [2005]. This data is available only for 233 districts, i.e. for districts which were directly under British control.

A.2 Figures

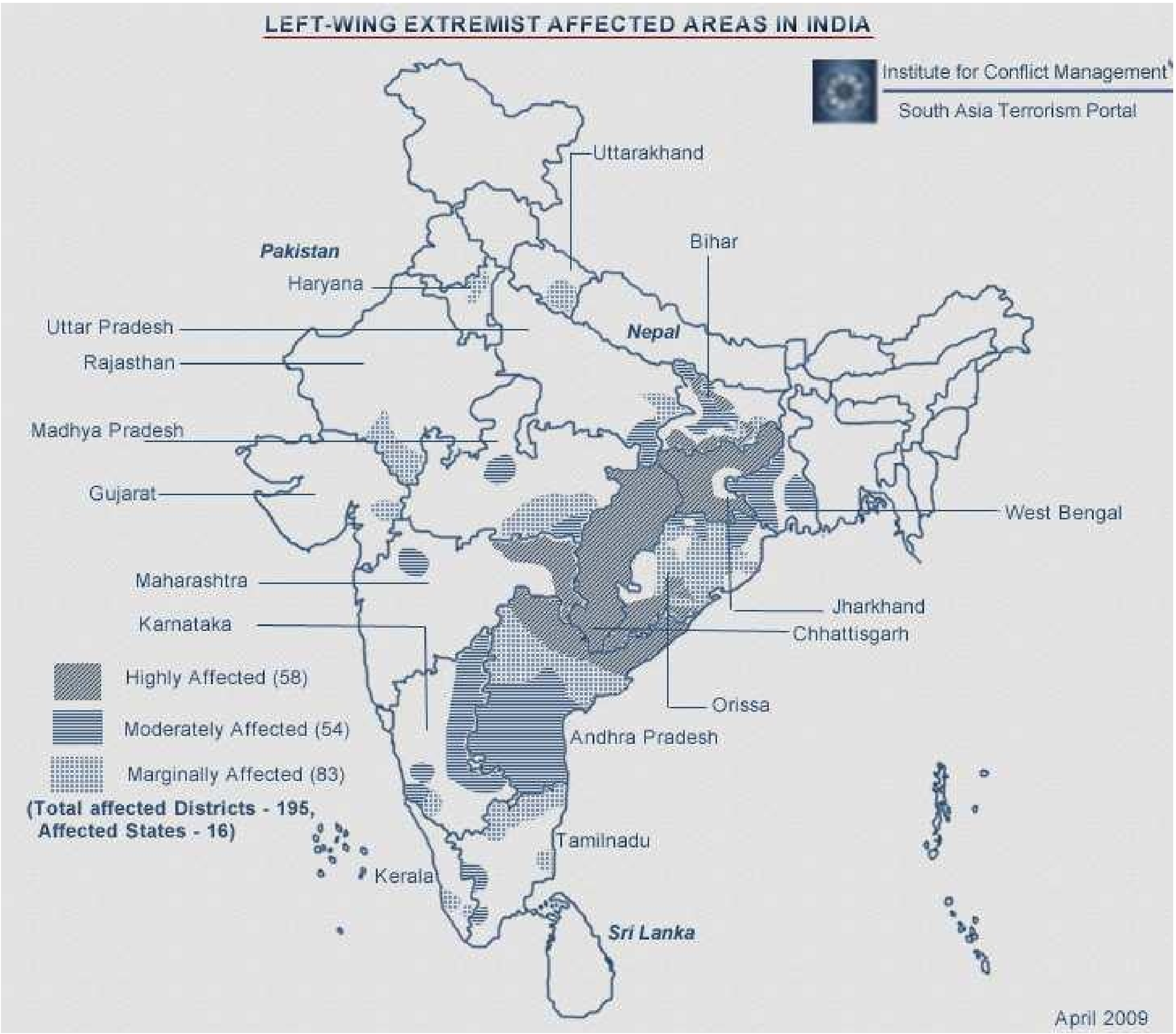

Figure A.1: Naxalite affected areas in India (Source: SATP)

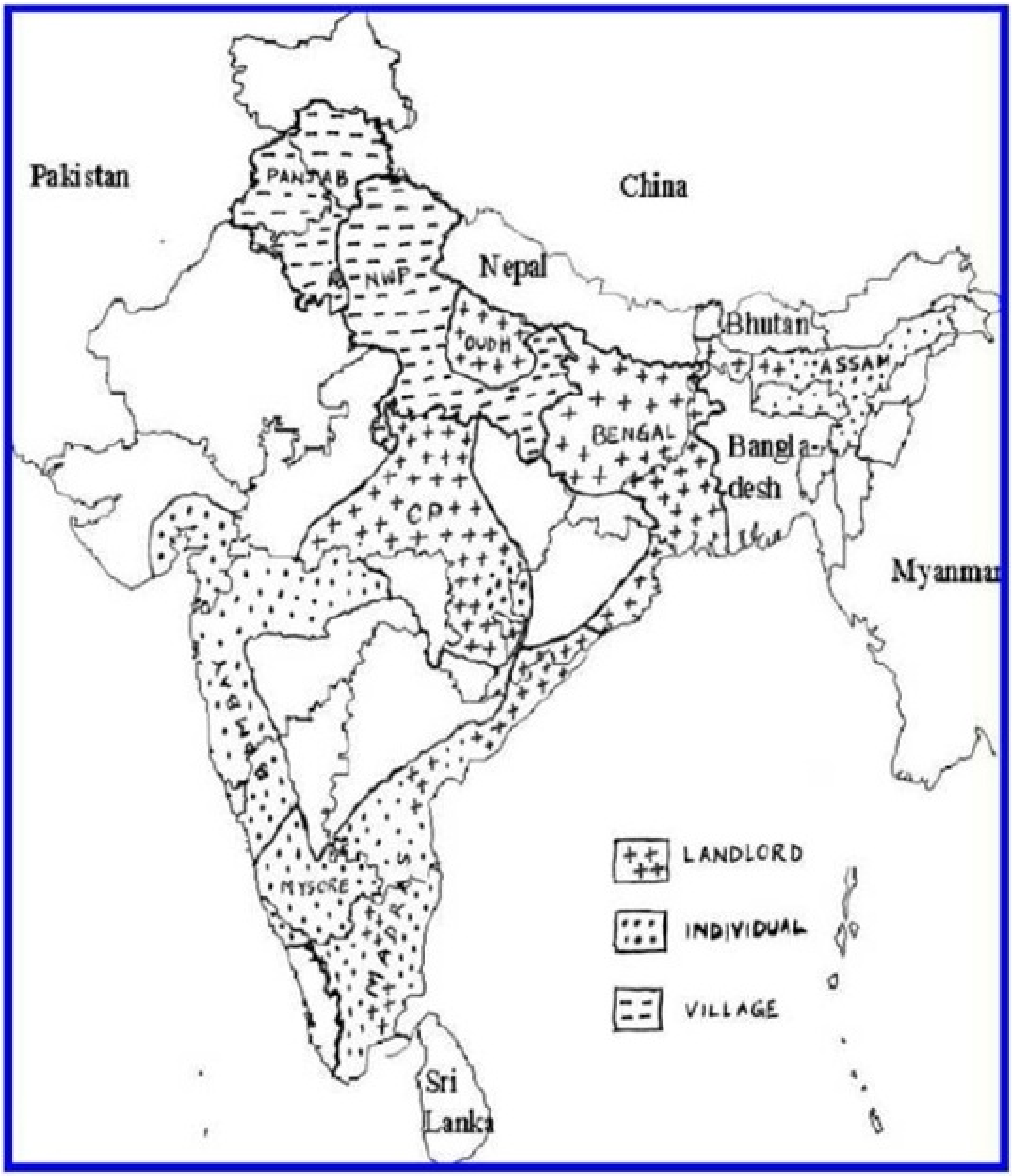

Figure A.2: Landlord & non Landlord districts in India (Source: Banerjee & Iyer, AER 2005)

A.3 Tables

Table A.1: Summary: All 3 rounds

Variable	Obs	Mean	Std. Dev.	Min	Max
maoist	1085	0.170507	0.376251	0	1
deadwounded	1085	6.322581	59.3128	0	1837
nincidents	1085	2.41106	15.79039	0	428
maoist_1	1085	0.087558	0.282781	0	1
gini	1046	0.266298	0.060809	0.103525	0.525915
mpc_87	1030	5.052897	0.244983	4.410083	5.777959
General_MPC_87	1030	5.132978	0.249794	4.498148	6.014436
SCST_MPC_87	1028	4.870675	0.250028	4.180412	5.941787
land inequality in 91	1055	0.47359	0.178855	0.120449	0.788384
proportion forest	1082	0.167094	0.18037	0.000699	0.832942
population density	1082	0.006383	0.022805	0.000692	0.416596
log_mpc	1046	5.849775	0.628196	4.410083	7.352093
log_gen_mpc	1043	6.002583	0.696857	4.498148	8.501844
log_scst_mpc	1045	5.674889	0.628981	4.180412	7.374076
log area	1082	8.721257	0.696394	6.475433	10.7288
proportion sandy	1040	0.006336	0.040049	0	0.688315
proportion barrenrocky	1040	0.00799	0.020191	0	0.265584
proportion steepaloping	1040	0.002556	0.010052	0	0.129534
log state capital distance	1031	5.485296	0.811263	0	6.899219
SCST_percent	1082	0.292082	0.15533	0.016113	0.946497
log_mpc_lag	686	5.608404	0.620172	4.410083	6.848523
log_gen_mpc_lag	685	5.727645	0.666179	4.498148	7.336274
log_scst_mpc_lag	685	5.440889	0.634257	4.180412	7.374076
p_nland	698	0.522766	0.429809	0	1

Table A.2: Correlation Matrix

	maoist	maoist_1	log mpc	proportion sandy	log state capital distance	proportion barren rocky	proportion steep sloping	proportion forest	SC percent	ST percent
maoist	1									
maoist_1	0.5449	1								
log mpc	-0.2713	-0.0596	1							
prop sandy	0.1275	0.0804	0.1252	1						
log state capital distance	0.1327	0.1216	-0.0443	-0.1017	1					
proportion barren rocky	0.0462	0.0597	0.1443	-0.0774	0.0162	1				
proportion steep sloping	-0.1567	-0.0983	0.3246	-0.0626	-0.0396	0.1753	1			
proportion forest	0.114	0.0574	0.0694	-0.1653	0.2297	-0.0088	0.2107	1		
SC percent	-0.0293	-0.0003	0.1513	0.1606	-0.4011	-0.1188	0.0439	-0.2466	1	
ST percent	0.0197	-0.0132	-0.2652	-0.1673	0.1996	0.2404	0.0084	0.3179	-0.507	1
population density	0.0463	0.0192	-0.0216	0.2251	-0.3502	-0.4447	-0.2079	-0.4327	0.1952	-0.5248
log area	0.2169	0.1613	-0.2474	-0.2124	0.2529	0.1663	-0.1171	0.1413	-0.1718	0.3598
gini	0.0218	0.1937	0.4101	-0.0043	0.119	-0.0208	0.0338	0.1363	-0.0309	0.0506
land inequality	0.1982	0.1453	0.0285	0.1137	-0.1999	-0.0849	0.0174	-0.0965	0.169	-0.437
MPC 87	-0.1784	-0.0812	0.6891	0.0768	-0.1229	0.1648	0.3058	0.0201	0.2413	-0.198
prop no landlord	-0.35	-0.2043	0.4941	-0.1214	0.1642	0.1929	0.3112	0.1289	-0.0381	0.0082
log general MPC	-0.2192	-0.0029	0.7489	0.0151	-0.0204	0.1275	0.1908	0.154	0.108	-0.0567
log SC MPC	-0.247	-0.0699	0.8392	0.1908	-0.0318	0.1126	0.3283	0.1087	0.0935	-0.0788
log ST MPC	-0.2613	-0.027	0.6032	0.0468	-0.1384	0.1341	0.2576	-0.0831	0.1407	-0.2957
General MPC growth	-0.1843	-0.1657	-0.0077	-0.0175	0.0385	0.0135	-0.0405	0.1162	0.0213	-0.0096
SC MPC growth	-0.1742	-0.1985	-0.0731	0.0554	0.0077	0.0356	-0.0382	-0.1022	-0.0173	-0.0148
ST MPC growth	-0.1482	-0.0598	-0.0384	-0.0101	-0.1007	-0.0254	-0.0358	-0.1512	0.073	-0.1668
MPC growth	-0.1861	-0.2043	-0.0069	0.05	0.0083	0.005	-0.0052	-0.0522	0.0139	-0.098

	log area	gini	land inequality	MPC 87	prop non landlord	log general MPC	log SC MPC	log ST MPC	General MPC growth	SC MPC growth
log area	1									
gini	0.1235	1								
land inequality	-0.4394	-0.077	1							
MPC 87	-0.2575	0.101	-0.0619	1						
prop no landlord	0.1053	0.1455	-0.253	0.4601	1					
log general MPC	-0.1714	0.4456	-0.0005	0.5304	0.3819	1				
log SC MPC	-0.1468	0.3167	-0.0384	0.5509	0.3794	0.5422	1			
log ST MPC	-0.3048	0.1431	0.1097	0.4364	0.2944	0.4418	0.4632	1		
General MPC growth	-0.0727	-0.3791	0.0565	-0.0156	-0.0025	0.2722	-0.0521	-0.0138	1	
SC MPC growth	0.0245	-0.525	-0.0146	-0.0818	-0.0073	-0.2372	0.1186	-0.0525	0.5474	1
ST MPC growth	-0.1332	-0.4111	0.1093	-0.0482	-0.0132	-0.1229	-0.0685	0.4045	0.4145	0.5549
Mpc growth	-0.0554	-0.5205	0.0658	-0.0829	-0.0028	-0.1254	0.0142	-0.0078	0.7043	0.8869

Table A.3: Same sample as Table 1.2

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
consumption per capita	-1.282***	-1.371***	-1.143	-2.956**	-2.144**	-3.625
	(0.434)	(0.500)	(0.822)	(1.234)	(1.034)	(2.812)
land inequality	2.743***	2.537***	4.552***	9.854***	9.773***	14.32***
	(0.631)	(0.875)	(0.863)	(1.971)	(2.086)	(2.550)
proportion sandy	19.20***	25.04***	30.47**	38.95	-1.637	43.93
	(7.300)	(6.797)	(13.77)	(70.41)	(59.24)	(59.49)
log state capital distance	0.127	0.155	0.139	0.199	0.220	-0.0377
	(0.197)	(0.213)	(0.210)	(0.483)	(0.352)	(0.338)
proportion barrenrocky	17.07***	24.39***	20.34***	46.46***	51.86***	37.59***
	(3.299)	(4.083)	(5.448)	(16.97)	(13.40)	(9.129)
proportion steepsloping	-70.86**	-113.6***	-155.9***	-201.2***	-331.1***	-389.2***
	(33.45)	(42.47)	(44.47)	(60.66)	(116.6)	(95.12)
proportion forest cover	1.211*	1.817***	2.645***	5.579***	6.019***	7.412***
	(0.651)	(0.686)	(0.563)	(2.041)	(1.960)	(1.737)
%Scheduled Castes	-0.862	-0.386	4.110***	-2.788	-3.855	11.19***
	(1.565)	(1.704)	(1.416)	(4.282)	(3.971)	(3.395)
%Scheduled Tribes	4.071	5.503**	1.813	1.400	3.974	0.939
	(2.475)	(2.588)	(1.288)	(6.599)	(8.201)	(2.815)
%Scheduled Tribes square	-6.590	-9.555**	-1.620	-5.864	-14.50	1.763
	(4.192)	(3.910)	(2.259)	(9.987)	(13.57)	(3.586)
population density	0.536**	0.789**	0.401	1.327**	0.836	0.291
	(0.208)	(0.329)	(0.309)	(0.551)	(0.526)	(0.626)
log area	0.768***	0.870***	0.894***	3.006***	2.669***	2.698***
	(0.198)	(0.289)	(0.144)	(0.544)	(0.478)	(0.338)
income inequality	-3.148*	-5.625***	-0.423	-15.45**	-18.13***	-2.350
	(1.780)	(1.848)	(1.794)	(6.189)	(6.455)	(6.539)
initial consumption per capita		-0.116			-3.275***	
		(0.797)			(0.945)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	655	431	655	655	431	655
ll	-186.0	-138.6	-142.8	-651.3	-540.6	-592.8

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In this table we have the same sample as in Table 2 but without the institutions variable. The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.4: The role of income differences across ethnic groups

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
General Castes consumption pc	-0.818**	-0.990***	-1.273***	-1.871**	-1.951***	-2.109***
	(0.319)	(0.363)	(0.202)	(0.874)	(0.756)	(0.680)
Scheduled Castes consumption pc	-0.444	-0.743***	-0.675***	-1.324	-1.338	-1.874*
	(0.348)	(0.245)	(0.186)	(0.832)	(0.831)	(0.986)
Scheduled Tribes consumption pc	-0.218	-0.135	-0.117	-0.670	-0.814	-0.690
	(0.484)	(0.578)	(0.695)	(1.407)	(1.223)	(1.051)
proportion sandy	-3.882	-1.202	3.041	-83.26***	-84.37***	-34.19***
	(5.447)	(5.517)	(3.221)	(30.86)	(29.02)	(9.921)
log state capital distance	-0.141	-0.00183	-0.0944	-0.916***	-0.695**	-0.458**
	(0.163)	(0.118)	(0.250)	(0.247)	(0.323)	(0.194)
proportion barrenrocky	7.414*	16.08***	15.35***	29.98	28.05	17.40***
	(4.261)	(5.217)	(4.608)	(21.45)	(23.02)	(4.523)
proportion steepsloping	-115.0***	-116.2**	-156.6**	-347.1***	-296.4***	-338.8***
	(39.00)	(55.10)	(60.87)	(126.2)	(101.8)	(131.5)
proportion forest cover	0.846	2.221**	3.682***	1.011	0.141	6.027***
	(0.891)	(1.036)	(0.728)	(1.287)	(1.691)	(1.437)
%Scheduled Castes	-0.0988	0.504	2.450	-3.464	-6.589*	4.253
	(1.420)	(1.185)	(1.839)	(4.206)	(3.377)	(5.513)
%Scheduled Tribes	2.080	1.098	1.576	4.489	2.564	0.210
	(1.369)	(2.007)	(1.472)	(3.617)	(4.295)	(2.729)
%Scheduled Tribes square	-1.991	-0.00427	0.158	-3.798	-1.908	2.840
	(2.278)	(2.286)	(2.294)	(3.626)	(4.174)	(3.320)
population density	-0.166	0.344	0.351	-1.684	-1.549	-1.117
	(0.379)	(0.363)	(0.368)	(1.040)	(1.092)	(0.737)
log area	0.806***	0.949***	0.828***	2.784***	2.553***	2.536***
	(0.198)	(0.282)	(0.208)	(0.379)	(0.383)	(0.315)
income inequality	1.504	-0.945	2.555*	-5.349	-7.974*	-5.269
	(1.635)	(2.433)	(1.552)	(3.958)	(4.254)	(4.035)
land inequality	5.412***	5.575***	4.555***	20.47***	18.98***	16.20***
	(0.902)	(1.062)	(0.841)	(5.315)	(5.471)	(4.213)
initial consumption per capita		-0.788*			0.163	
		(0.423)			(1.013)	
State dummies			Yes			Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	567	356	552	567	356	567
ll	-163.0	-99.64	-118.5	-761.9	-616.9	-698.6

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.5: The role of income differences across ethnic groups

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
General Castes Consumption pc	-0.474*	-0.570**	-0.702***	-0.234	-0.575	-1.021
	(0.245)	(0.253)	(0.211)	(0.978)	(0.973)	(0.965)
Scheduled Castes Consumption pc	-0.463	-0.943**	-0.370	-1.123	-1.605*	-1.912**
	(0.371)	(0.417)	(0.389)	(0.908)	(0.874)	(0.945)
Scheduled Tribes Consumption pc	-0.747**	-0.539**	-0.664*	-1.319*	-0.867*	-1.726*
	(0.311)	(0.267)	(0.388)	(0.739)	(0.525)	(1.038)
proportion sandy	16.08**	25.89***	27.67**	-40.88	-62.09***	2.766
	(6.513)	(4.803)	(11.69)	(28.14)	(23.65)	(25.29)
log state capital distance	0.262	0.266	0.194	0.152	0.308	-0.0104
	(0.233)	(0.259)	(0.259)	(0.570)	(0.408)	(0.465)
proportion barrenrocky	22.12***	26.58***	26.58***	57.78***	45.52***	43.53***
	(4.907)	(5.385)	(8.323)	(18.87)	(11.36)	(12.02)
proportion steepsloping	-110.4***	-81.10**	-199.9***	-364.8***	-313.7***	-519.4***
	(33.05)	(40.31)	(66.74)	(129.3)	(92.62)	(168.1)
proportion forest cover	1.657***	2.248***	2.808***	5.460***	4.763***	7.177***
	(0.559)	(0.719)	(0.757)	(1.282)	(1.303)	(1.739)
%Scheduled Castes	-0.0800	0.531	4.503***	-2.296	-3.157	8.422***
	(1.607)	(1.757)	(1.599)	(3.912)	(3.723)	(3.202)
%Scheduled Tribes	0.305	2.021	0.352	-5.308	-5.231	-2.864
	(1.999)	(2.027)	(1.304)	(3.827)	(4.501)	(3.940)
%Scheduled Tribes square	-2.050	-5.541	0.640	0.270	-1.772	4.277
	(3.880)	(3.674)	(2.411)	(7.383)	(9.236)	(5.165)
population density	0.282	0.616*	0.260	0.0281	-0.0317	-0.392
	(0.248)	(0.364)	(0.359)	(0.720)	(0.769)	(0.612)
log area	0.836***	1.060***	0.834***	3.482***	3.329***	2.772***
	(0.226)	(0.363)	(0.162)	(0.535)	(0.452)	(0.434)
income inequality	-0.247	-2.970*	1.169	-15.78***	-18.36***	-5.041
	(1.626)	(1.693)	(1.572)	(5.502)	(5.540)	(6.186)
land inequality	3.062***	2.605***	4.118***	12.64***	10.91***	14.10***
	(0.459)	(0.742)	(0.901)	(2.648)	(2.363)	(2.595)
proportion Non landlord	-0.973***	-1.041***	-1.216**	-2.377***	-1.982***	-1.023
	(0.254)	(0.263)	(0.533)	(0.644)	(0.503)	(0.828)
initial consumption per capita		0.298			-0.741	
		(0.601)			(0.614)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	507	319	507	507	319	507
ll	-139.9	-101.8	-113.7	-535.3	-440.7	-498.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table is same as Table 4 in the paper but also controls for historical institutions. The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.6: The role of income differences across ethnic groups: SC&ST vs. General Castes

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
General Caste consumption pc	-0.492*	-0.566**	-0.722***	-2.169***	-1.418**	-1.808***
	(0.279)	(0.229)	(0.181)	(0.575)	(0.721)	(0.553)
SC & ST consumption pc	-1.037***	-1.039**	-0.625	-2.222**	-2.099*	-2.449*
	(0.389)	(0.502)	(0.547)	(0.927)	(1.165)	(1.472)
land inequality	2.837***	2.902**	4.486***	14.51***	12.97***	14.77***
	(1.084)	(1.218)	(1.004)	(3.438)	(3.905)	(3.438)
proportion sandy	-1.091	-0.771	-0.0382	-46.54	-49.09**	-15.83
	(2.636)	(2.417)	(1.379)	(28.97)	(24.44)	(11.55)
log state capital distance	-0.0494	0.0640	-0.0234	-0.610***	-0.309*	-0.346
	(0.124)	(0.118)	(0.200)	(0.166)	(0.184)	(0.249)
proportion barrenrocky	10.27***	14.06***	11.88***	30.11**	38.16**	23.97***
	(2.192)	(3.016)	(4.389)	(14.30)	(15.75)	(5.205)
proportion steepsloping	-65.18***	-90.58***	-125.6***	-238.9***	-206.4***	-324.7***
	(19.23)	(35.12)	(34.34)	(40.98)	(52.41)	(75.77)
proportion forest cover	1.388*	1.823**	2.300***	2.452	2.511	5.491***
	(0.718)	(0.783)	(0.548)	(1.875)	(1.697)	(1.526)
%Scheduled Castes & Tribes	0.982	1.704**	1.976***	3.192*	3.604**	4.314***
	(0.608)	(0.734)	(0.330)	(1.849)	(1.614)	(1.331)
population density	0.340	0.625**	0.329	-0.472	-0.235	-0.140
	(0.222)	(0.299)	(0.286)	(0.622)	(0.544)	(0.491)
log area	0.696***	0.784***	0.703***	2.357***	2.335***	2.504***
	(0.216)	(0.304)	(0.144)	(0.387)	(0.322)	(0.291)
income inequality	-0.581	-4.248**	1.396	-4.110	-10.37**	-3.628
	(1.937)	(1.874)	(1.281)	(3.588)	(4.289)	(3.982)
Initial consumption per capita	No	Yes	No	No	Yes	No
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	993	651	993	993	651	993
ll	-272.4	-180.9	-193.8	-1013.9	-825.3	-912.2

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In this table we combine the incomes of the Scheduled Castes and tribes together and compare its impact on conflict vis a vis the income of general castes. The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.7: The Effect of Growth on Conflict

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
Consumption growth	-0.0213	-0.0335*	-0.00481	0.0365	-0.00927	-0.0203
	(0.0181)	(0.0183)	(0.0181)	(0.0371)	(0.0389)	(0.0408)
proportion sandy	-0.991	-1.386	0.488	-31.55*	-42.08**	-16.08
	(2.787)	(2.807)	(1.521)	(17.17)	(20.78)	(13.19)
log state capital distance	0.0572	0.0510	0.167	-0.527**	-0.492**	0.102
	(0.161)	(0.150)	(0.195)	(0.207)	(0.191)	(0.272)
proportion barrenrocky	12.03**	15.55***	13.29**	47.66***	52.54***	27.48***
	(5.028)	(3.658)	(6.089)	(15.12)	(14.98)	(9.349)
proportion steepsloping	-69.79***	-92.50***	-97.17***	-194.3***	-247.5***	-248.7***
	(22.75)	(31.15)	(37.30)	(30.34)	(68.15)	(70.50)
proportion forest cover	1.094	1.526**	2.430***	2.730	3.006	5.793***
	(0.812)	(0.732)	(0.782)	(2.373)	(2.187)	(1.982)
%Scheduled Castes	0.296	0.219	3.189**	-3.939	-4.777*	4.423
	(1.201)	(1.399)	(1.479)	(3.478)	(2.841)	(3.665)
%Scheduled Tribes	5.777***	4.969***	3.853***	10.47***	10.78***	9.167***
	(1.550)	(1.786)	(0.835)	(3.743)	(3.135)	(3.442)
%Scheduled Tribes square	-5.987***	-5.526**	-2.844*	-10.77**	-13.80***	-8.308
	(1.971)	(2.170)	(1.726)	(4.668)	(4.550)	(5.261)
population density	0.727***	0.619**	0.649*	0.108	-0.0894	0.858
	(0.258)	(0.304)	(0.373)	(0.689)	(0.643)	(0.792)
log area	0.871***	0.784***	0.913***	2.367***	2.432***	2.687***
	(0.278)	(0.270)	(0.199)	(0.323)	(0.329)	(0.575)
income inequality	-6.188***	-4.961***	-2.238**	-21.51***	-17.64***	-10.17***
	(1.951)	(1.883)	(1.058)	(4.989)	(5.166)	(3.142)
land inequality	3.561***	3.686***	4.187***	13.81***	14.21***	9.756***
	(0.970)	(1.164)	(1.028)	(2.946)	(3.278)	(3.178)
initial consumption per capita		-1.401**			-4.263***	
		(0.608)			(0.932)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	640	638	640	640	638	640
ll	-193.5	-185.3	-143.5	-844.2	-828.3	-770.9

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The growth rates correspond to the 2 latter rounds. The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.8: The Effect of Growth on Conflict

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
Consumption growth	-0.00951	-0.0102	-0.00426	0.0213	0.00485	-0.0304
	(0.0178)	(0.0163)	(0.0192)	(0.0416)	(0.0446)	(0.0601)
proportion sandy	21.75***	19.16***	49.08**	-6.925	-21.67	5.918
	(6.644)	(6.258)	(24.75)	(45.93)	(32.90)	(26.62)
log state capital distance	0.211	0.247	0.243	-0.204	-0.0299	0.258
	(0.213)	(0.216)	(0.249)	(0.356)	(0.335)	(0.552)
proportion barrenrocky	22.84***	25.71***	26.42***	41.98**	50.00***	36.16***
	(5.202)	(4.944)	(7.329)	(18.18)	(16.85)	(12.15)
proportion steepsloping	-47.14*	-100.8***	-105.4**	-137.2***	-291.2**	-203.7***
	(28.57)	(36.85)	(50.28)	(53.07)	(120.0)	(62.36)
proportion forest cover	2.138***	2.158***	3.158***	3.942*	5.177**	7.002***
	(0.720)	(0.673)	(0.979)	(2.211)	(2.169)	(1.942)
%Scheduled Castes	-0.614	-0.609	4.886***	-6.801*	-7.115	6.799**
	(1.319)	(1.453)	(1.837)	(3.815)	(4.564)	(3.162)
%Scheduled Tribes	4.870**	4.791*	4.472**	-0.980	-0.794	0.641
	(2.386)	(2.721)	(2.019)	(6.152)	(6.837)	(5.397)
%Scheduled Tribes square	-8.931**	-8.838**	-6.013**	-6.719	-9.050	-0.467
	(3.735)	(3.945)	(2.740)	(11.11)	(12.37)	(6.949)
population density	0.846***	0.806**	0.734*	-0.367	-0.181	1.175
	(0.297)	(0.334)	(0.444)	(1.021)	(0.776)	(0.978)
log area	1.059***	1.059***	1.217***	3.041***	3.107***	3.312***
	(0.269)	(0.322)	(0.213)	(0.579)	(0.483)	(0.558)
income inequality	-7.272***	-7.164***	-4.630***	-26.04***	-23.95***	-12.25**
	(1.467)	(1.574)	(1.503)	(5.552)	(6.529)	(5.949)
land inequality	2.013**	2.216**	3.041**	9.880***	10.24***	6.677***
	(0.833)	(0.893)	(1.285)	(3.227)	(2.894)	(1.629)
proportion Non landlord	-1.162***	-1.114***	-1.552***	-2.702***	-2.298***	-2.191***
	(0.290)	(0.244)	(0.427)	(0.572)	(0.543)	(0.650)
initial consumption per capita		-0.151			-2.409***	
		(0.731)			(0.616)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	422	420	422	422	420	422
ll	-135.4	-132.9	-105.6	-544.4	-535.4	-500.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table is same as Table 5 in the paper but also controls for historical institutions. The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

TABLE A.9: Change in causes of conflict over time by subgroups

	Round 1 Probability	Round 2 Probability	Round 3 Probability	Round 1 Intensity	Round 2 Intensity	Round 3 Intensity
General Caste consumption pc	0.863	-1.620***	-0.464	-0.246	-2.258	-0.607
	(1.418)	(0.534)	(0.501)	(4.885)	(1.486)	(0.858)
Scheduled Caste consumption pc	0.322	-1.302*	-0.459	3.526	-0.370	-1.219
	(0.870)	(0.685)	(0.882)	(4.910)	(0.887)	(0.996)
Scheduled Tribe consumption pc	-0.730	-0.487	-1.028	-2.752*	-1.881	-1.694
	(0.534)	(0.706)	(1.216)	(1.539)	(1.671)	(1.567)
land inequality	5.158***	8.944***	5.965***	21.03***	26.51***	15.70***
	(1.296)	(1.701)	(1.419)	(3.429)	(6.472)	(3.588)
%Scheduled Tribes	1.691	-3.502	4.303	3.225	1.630	6.614
	(3.532)	(2.235)	(4.225)	(10.65)	(5.616)	(4.066)
%Scheduled Tribes square	-3.811	6.782***	-3.609	-7.120	1.148	-11.35*
	(6.525)	(2.241)	(4.572)	(13.58)	(5.354)	(5.930)
%Scheduled Castes	-2.530	-0.871	5.477***	-6.067	-8.409	-5.112
	(2.727)	(1.377)	(1.856)	(8.469)	(6.048)	(4.129)
proportion sandy	-21.10	-3.417	-1.458	-130.7**	-327.9***	-57.31***
	(25.46)	(6.103)	(5.286)	(64.37)	(81.03)	(16.78)
proportion barrenrocky	1.241	20.94**	16.39**	52.56	22.67	36.79*
	(4.836)	(8.550)	(8.266)	(44.10)	(25.40)	(20.62)
proportion steepsloping	-119.2***	-334.7**	-64.11*	-709.7***	-407.0**	-280.5***
	(42.79)	(132.8)	(34.31)	(240.5)	(194.0)	(82.66)
log state capital distance	-0.291	0.139	0.199	-0.934***	-0.911***	-0.108
	(0.180)	(0.339)	(0.256)	(0.297)	(0.282)	(0.282)
proportion forest cover	-0.174	2.228**	1.979	2.856	-3.910**	1.147
	(1.346)	(1.053)	(1.873)	(2.781)	(1.606)	(1.937)
log area	1.043***	0.899**	1.376***	6.405***	3.707***	4.014***
	(0.375)	(0.444)	(0.493)	(1.855)	(0.826)	(0.451)
population density	-0.518	-0.358	1.037**	0.859	-2.691	0.389
	(0.569)	(0.529)	(0.466)	(1.596)	(1.688)	(0.508)
income inequality	4.769	-1.449	-3.852	-0.849	-8.148	-14.10***
	(3.376)	(4.559)	(4.060)	(13.98)	(11.81)	(5.003)
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	289	136	145	289	152	145
ll	-58.81	-39.90	-33.24	-144.8	-215.9	-329.0

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The table reports coefficients and not marginal effects. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.10: The determinants of the Maoist Conflict in India - Marginal Effects

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
consumption per capita	-0.197***	-0.221***	-0.167**	-880.9	-1,677	-150.6
	(0.0744)	(0.0854)	(0.0809)	(3,312)	(5,862)	(238.8)
land inequality	0.537***	0.548**	0.474***	4,387	14,816	521.5
	(0.166)	(0.216)	(0.0924)	(26,932)	(82,892)	(960.1)
proportion sandy	-0.227	-0.0928	0.0354	-9,231	-46,591**	-364.1
	(0.393)	(0.414)	(0.149)	(6,140)	(23,624)	(314.0)
log state capital distance	-0.0122	0.00801	-0.00215	-232.3	-516.8	-10.78
	(0.0225)	(0.0225)	(0.0190)	(912.5)	(1,822)	(11.69)
proportion barrenrocky	1.560***	2.451***	1.402***	11,975**	52,262***	956.7***
	(0.393)	(0.564)	(0.478)	(4,889)	(18,305)	(165.4)
proportion steepsloping	-10.01***	-13.30***	-13.52***	-79,012***	-265,409***	-12,303***
	(3.134)	(5.046)	(3.518)	(10,693)	(62,016)	(2,579)
proportion forest cover	0.132	0.241**	0.256***	780.0	3,250	215.3
	(0.0913)	(0.110)	(0.0587)	(4,868)	(18,743)	(402.7)
%Scheduled Castes	-0.0934	0.00725	0.312**	-1,311	-5,522*	249.3
	(0.203)	(0.233)	(0.130)	(1,205)	(3,028)	(213.5)
%Scheduled Tribes	0.651***	0.650**	0.308***	2,934	10,971	221.4
	(0.250)	(0.273)	(0.0782)	(17,846)	(61,060)	(375.5)
%Scheduled Tribes square	-0.836*	-0.742**	-0.220	-3,551*	-14,782***	-136.1
	(0.443)	(0.327)	(0.151)	(1,910)	(5,624)	(143.4)
population density	0.0436	0.0950*	0.0352	-43.97	-3.608	-1.358
	(0.0315)	(0.0486)	(0.0314)	(194.6)	(562.5)	(19.10)
log area	0.0997***	0.115***	0.0728***	706.3	2,608	84.86
	(0.0376)	(0.0394)	(0.0142)	(2,147)	(7,103)	(101.6)
income inequality	-0.0195	-0.545*	0.177	-2,147	-15,470**	-68.80
	(0.304)	(0.309)	(0.161)	(1,469)	(6,083)	(109.4)
initial consumption per capita		-0.0755			-4,109	
		(0.101)			(15,599)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	996	656	996	996	656	996
ll	-269.1	-181.9	-192.9	-1013.1	-829.0	-913.4

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This replicates Table 1.1 but reports marginal effects instead of coefficients. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.11: The role of historical institutions - Marginal Effects

	(1) Probability	(2) Probability	(3) Probability	(4) Intensity	(5) Intensity	(6) Intensity
consumption per capita	-0.130	-0.163	-0.110	-32.90	-52.38	-43.04
	(0.0810)	(0.0994)	(0.105)	(36.32)	(52.51)	(54.94)
land inequality	0.367***	0.370**	0.458***	213.7	374.9	184.6
	(0.0972)	(0.163)	(0.104)	(175.4)	(484.5)	(116.2)
proportion Non landlord	-0.141***	-0.178***	-0.139***	-47.61	-79.37	-19.38*
	(0.0438)	(0.0411)	(0.0527)	(40.24)	(64.89)	(11.53)
proportion sandy	2.660**	3.553***	3.756**	853.9	-651.6	584.2
	(1.176)	(1.170)	(1.779)	(779.1)	(1,455)	(765.5)
log state capital distance	0.0287	0.0415	0.0190	-0.413	5.091	-1.758
	(0.0313)	(0.0387)	(0.0246)	(6.548)	(15.56)	(5.088)
proportion barrenrocky	2.864***	4.482***	2.682***	1,048***	1,966***	550.0***
	(0.586)	(0.870)	(0.721)	(312.9)	(565.7)	(130.6)
proportion steepsloping	-8.065**	-16.29***	-15.46***	-3,955***	-12,601***	-5,069***
	(4.109)	(6.010)	(5.091)	(1,503)	(4,408)	(1,409)
proportion forest cover	0.232**	0.360***	0.319***	101.5	207.4	101.3*
	(0.0924)	(0.111)	(0.0728)	(84.53)	(301.1)	(59.62)
%Scheduled Castes	-0.146	-0.101	0.497***	-91.61	-250.6	146.7
	(0.208)	(0.257)	(0.150)	(149.2)	(303.8)	(134.4)
%Scheduled Tribes	0.414	0.729*	0.168	-64.81	-62.32	-1.700
	(0.346)	(0.372)	(0.133)	(97.03)	(237.3)	(44.98)
%Scheduled Tribes square	-0.838	-1.482**	-0.140	-49.12	-339.6	30.47
	(0.612)	(0.586)	(0.233)	(138.2)	(662.5)	(56.36)
population density	0.0728**	0.130**	0.0482	3.793	1.168	1.851
	(0.0317)	(0.0537)	(0.0376)	(12.51)	(25.27)	(8.893)
log area	0.127***	0.172***	0.105***	63.27	115.1	38.45*
	(0.0353)	(0.0507)	(0.0152)	(58.15)	(97.92)	(21.72)
income inequality	-0.423	-0.997***	-0.130	-295.1***	-780.4***	-49.52
	(0.264)	(0.318)	(0.228)	(94.35)	(251.3)	(88.28)
initial consumption per capita		0.0413			-87.98	
		(0.135)			(79.11)	
State dummies	No	No	Yes	No	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	655	431	655	655	431	655
ll	-177.9	-131.6	-138.2	-642.2	-535.5	-590.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table replicates Table 1.2 but reports marginal effects instead of coefficients. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1,2 and 3 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 4, 5 and 6 explain the intensity of conflict (the no. of dead & wounded in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Table A.12: The growth in income of different ethnic groups- Marginal Effects

	(1) Probability	(2) Probability	(3) Intensity	(4) Intensity	(5) Intensity	(6) Intensity
General Castes growth	-0.00374***	-0.00360***	0.644	-0.144	0.424	-0.0560
	(0.00144)	(0.00129)	(1.817)	(1.022)	(0.897)	(0.143)
Scheduled Castes growth	-0.00202	0.000597	4.429	2.327	0.668	0.320
	(0.00168)	(0.00118)	(6.008)	(2.247)	(1.319)	(0.339)
Scheduled Tribes growth	-0.00222	-0.00223	-10.91	-4.194	-2.801	-0.576
	(0.00268)	(0.00306)	(14.31)	(5.183)	(5.914)	(0.618)
proportion sandy	-0.674	-0.00138	-9,948***	-2,376**	-1,094**	-171.9***
	(0.834)	(0.317)	(3,766)	(1,080)	(503.8)	(59.30)
log state capital distance	0.0168	0.0143	-40.77	1.388	-14.24	0.0401
	(0.0137)	(0.0229)	(71.60)	(20.08)	(27.44)	(3.481)
proportion barrenrocky	2.272***	2.387***	3,583	1,632***	1,001	277.0**
	(0.799)	(0.924)	(3,061)	(583.8)	(697.2)	(119.4)
proportion steepsloping	-26.35***	-27.89***	-51,950***	-21,257***	-11,366***	-2,607***
	(7.623)	(10.64)	(15,232)	(7,885)	(3,416)	(982.7)
proportion forest cover	0.350***	0.541***	-136.4	271.6	64.22	71.17
	(0.106)	(0.0843)	(181.0)	(383.7)	(260.1)	(57.07)
%Scheduled Castes	0.135	0.395	-886.6*	100.5	-85.24	41.03
	(0.190)	(0.294)	(537.6)	(316.1)	(145.1)	(79.93)
%Scheduled Tribes	0.220	0.0707	1,056	420.3	214.9	37.44
	(0.334)	(0.202)	(2,533)	(417.8)	(819.1)	(33.02)
%Scheduled Tribes square	-0.143	0.0256	-1,510	-501.2	-280.7	-23.95
	(0.389)	(0.228)	(1,181)	(562.7)	(287.2)	(37.40)
population density	0.0274	0.0292	-82.29	5.585	-4.091	8.718
	(0.0627)	(0.0529)	(123.9)	(55.85)	(29.14)	(9.350)
log area	0.177***	0.110***	434.7	190.1	125.0	31.17*
	(0.0436)	(0.0279)	(510.6)	(176.2)	(220.9)	(17.38)
income inequality	-0.785**	-0.677***	-1,845***	-866.5***	-556.8***	-129.8**
	(0.366)	(0.219)	(544.7)	(324.9)	(146.2)	(53.56)
land inequality	1.092***	0.844***	2,240	758.5	568.7	97.92
	(0.188)	(0.186)	(4,744)	(879.8)	(1,968)	(81.92)
initial consumption per capita	No	Yes	No	Yes	No	Yes
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	No	Yes	No	Yes	No	Yes
Conflict_1	Yes	Yes	Yes	Yes	Yes	Yes
Observations	290	284	290	290	290	290
ll	-74.40	-57.41	-544.7	-511.9	-493.8	-453.0

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table replicates Table 1.4 but reports marginal effects instead of coefficients. The data is from a database built by the author combining myriad databases. All regressions at the district level with robust standard errors, clustered at the state level (in parentheses). Columns 1 and 2 explain the probability of conflict (the presence of conflict in the district) using Probit regressions. Columns 3, 4, 5 and 6 explain the intensity of conflict (columns 3 and 4 - the no. of dead & wounded in the district; columns 5 and 6 - the number of incidents in the district) using Negative Binomial regressions. There are 3 time periods used corresponding to 3 NSS rounds, 1987-88 (43rd), 1999-00 (55th), 2004-05 (61st). The growth rates correspond to the 2 latter rounds. The conflict data is for the years 1979-2009, which are clubbed to corresponding NSS rounds.

Appendix B

Appendix for Chapter2

B.1 Figures

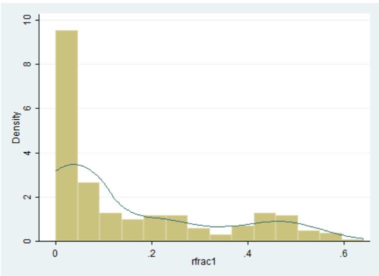

Figure B.1: Histogram: Religious Fractionalization (Aggrgegation Level 1)

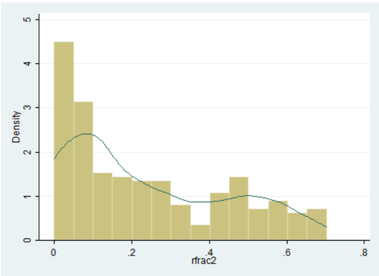

Figure B.2: Histogram: Religious Fractionalization (Aggrgegation Level 2)

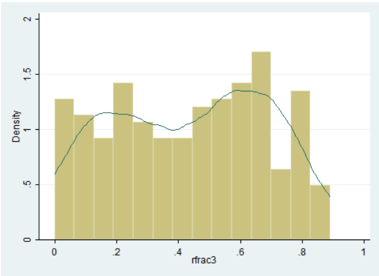

Figure B.3: Histogram: Religious Fractionalization (Aggrgegation Level 3)

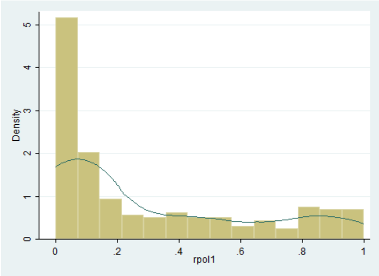

Figure B.4: Histogram: Religious Polarization (Aggrgegation Level 1)

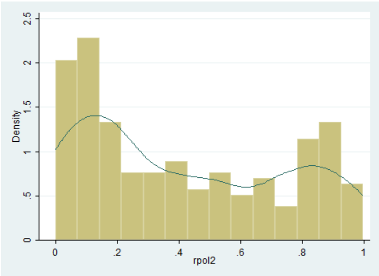

Figure B.5: Histogram: Religious Polarization (Aggrgegation Level 2)

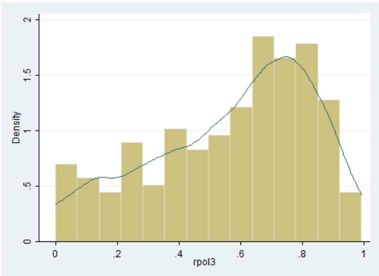

Figure B.6: Histogram: Religious Polarization (Aggrgegation Level 3)

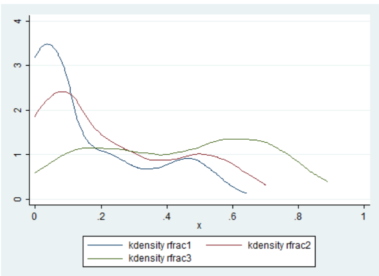

Figure B.7: Density: Religious Fractionalization

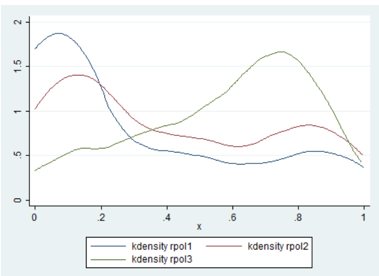

Figure B.8: Density: Religious Polarization

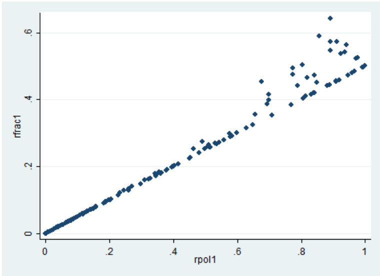

Figure B.9: Scatter Plot: Religious Diversity (Aggrgegation Level 1)

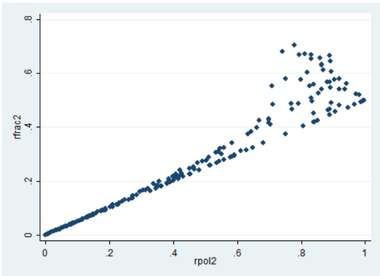

Figure B.10: Scatter Plot: Religious Diversity (Aggrgegation Level 2)

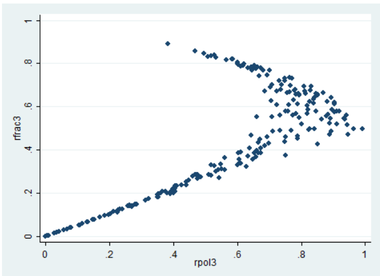

Figure B.11: Scatter Plot: Religious Diversity (Aggrgegation Level 3)

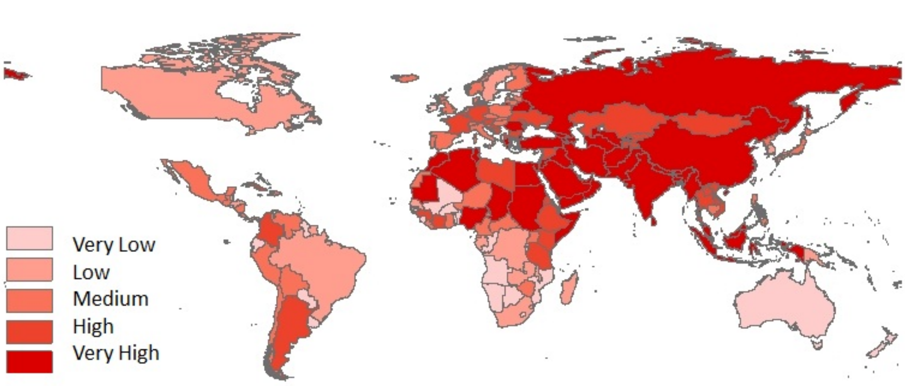

Figure B.12: Religious Intolerance (no data for the U.S.)

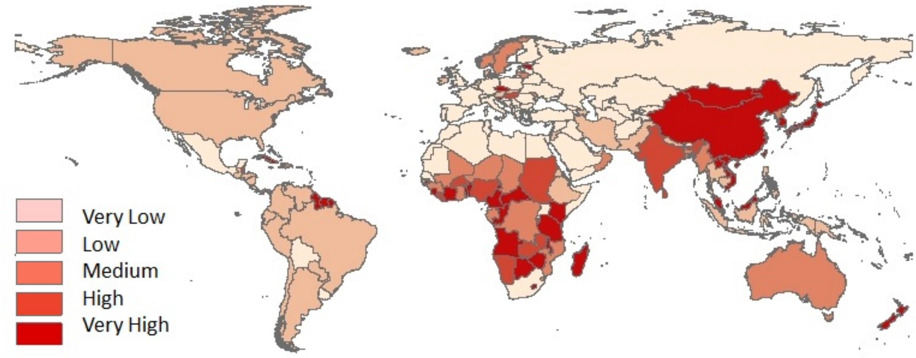

FIGURE B.13: Religious Fractionalization at Level 1

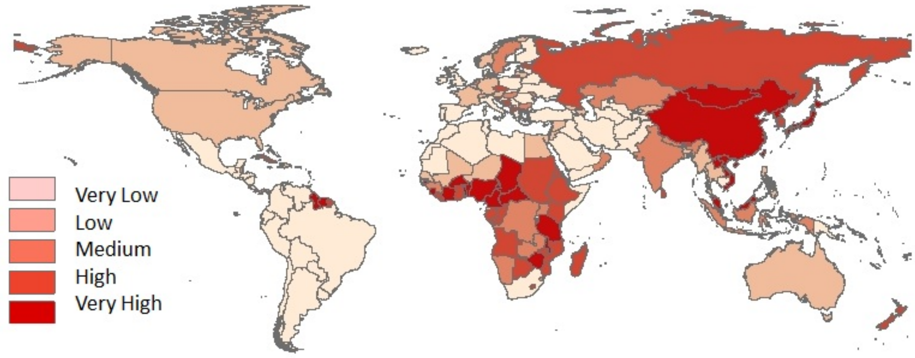

FIGURE B.14: Religious Fractionalization at Level 2

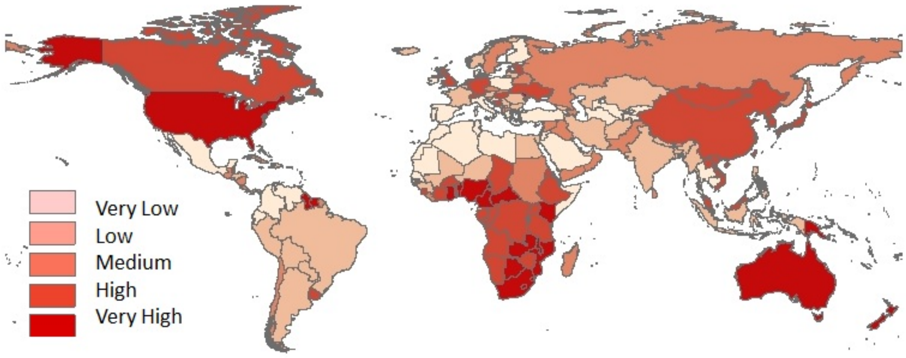

FIGURE B.15: Religious Fractionalization at Level 3

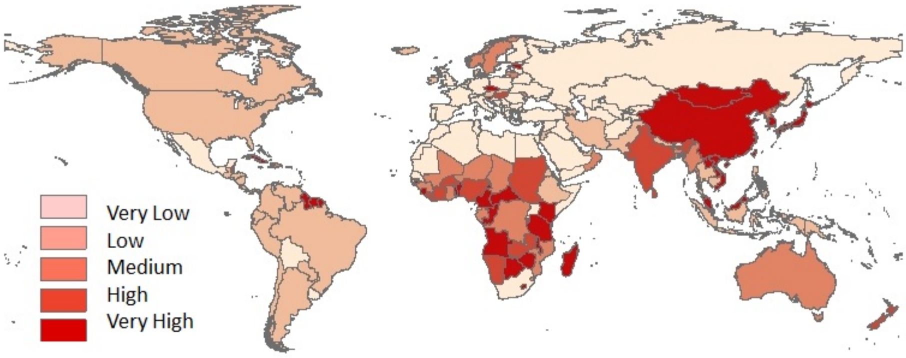

FIGURE B.16: Religious Polarization at Level 1

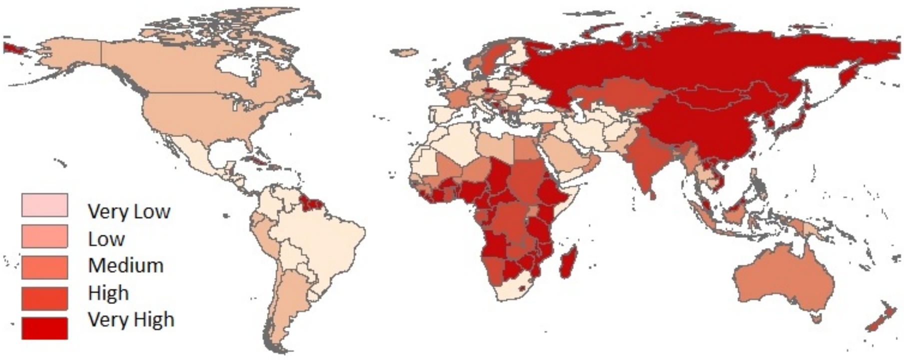

FIGURE B.17: Religious Polarization at Level 2

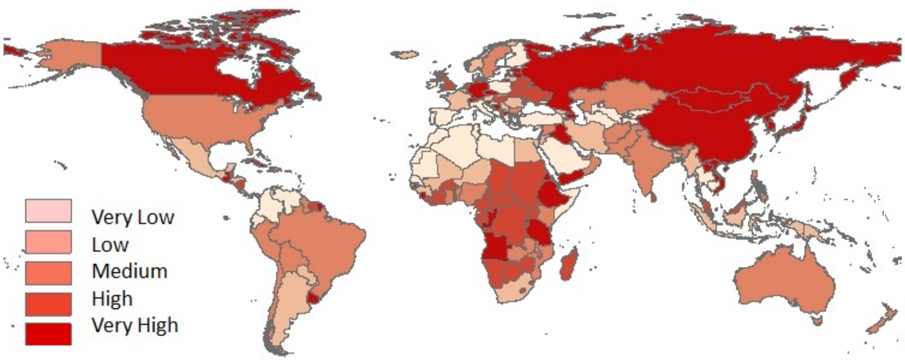

FIGURE B.18: Religious Polarization at Level 3

B.2 Tables

Table B.1: Cross-correlation table Religious diversity and Intolerance (197 obs)

Variables	rfrac1	rpol1	rfrac2	rpol2	rfrac3	rpol3
rpol1	0.991					
rfrac2	0.823	0.813				
rpol2	0.782	0.791	0.967			
rfrac3	0.547	0.550	0.585	0.572		
rpol3	0.502	0.515	0.586	0.628	0.782	
Religious Intolerance	-0.266	-0.286	-0.107	-0.118	-0.395	-0.244

Table B.2: Correlation with Montalvo and Reynal-Querol [2005] measures (137 obs)

	rfrac1	rpol1	rfrac2	rpol2	rfrac3	rpol3
ethpol	0.0256	0.0175	0.1026	0.0648	0.1542	0.1013
ethfrac	0.2659	0.2728	0.4165	0.381	0.2798	0.2539
relpol	0.5898	0.5957	0.6666	0.6564	0.487	0.5801
relfrac	0.6684	0.6597	0.7466	0.7133	0.499	0.5746

Table B.3: Correlation of religious diversity with Desmet et al. [2012] measures (208 obs)

	rfrac1	rpol1	rfrac2	rpol2	rfrac3	rpol3
elf1	0.1445	0.1421	0.1782	0.1682	0.1363	0.1627
elf2	0.0706	0.069	0.1855	0.1797	0.1052	0.1741
elf3	0.1238	0.1088	0.2362	0.21	0.1015	0.1814
elf4	0.1439	0.1276	0.2745	0.2324	0.1352	0.1891
elf5	0.1334	0.1186	0.2805	0.2405	0.1113	0.1703
elf6	0.1036	0.0921	0.2664	0.2292	0.074	0.1621
elf7	0.1233	0.112	0.2829	0.2389	0.0803	0.1549
elf8	0.1115	0.1007	0.2733	0.2312	0.0754	0.1456
elf9	0.1822	0.178	0.3463	0.3034	0.1492	0.1865
elf10	0.2109	0.2099	0.3712	0.3297	0.2059	0.2091
elf11	0.2103	0.2111	0.3702	0.3288	0.2214	0.2102
elf12	0.2094	0.2103	0.3676	0.3265	0.2228	0.2129
elf13	0.2093	0.2102	0.3674	0.3264	0.2229	0.2129
elf14	0.2098	0.211	0.3665	0.326	0.2244	0.2167
elf15	0.2098	0.211	0.3664	0.3259	0.2244	0.2168
pol1	0.1324	0.1317	0.1742	0.1692	0.1245	0.1585
pol2	0.043	0.0431	0.1546	0.1532	0.0715	0.1568
pol3	0.0642	0.0548	0.1728	0.161	0.0684	0.1618
pol4	0.0898	0.0791	0.2064	0.1788	0.1063	0.172
pol5	0.0708	0.0587	0.1855	0.1655	0.0579	0.1363
pol6	0.0271	0.0168	0.1419	0.1267	0.0101	0.1114
pol7	-0.0025	-0.0138	0.1033	0.0936	-0.0232	0.0753
pol8	-0.0358	-0.047	0.0469	0.0371	-0.0781	0.0168
pol9	0.0208	0.0127	0.1123	0.1042	-0.0204	0.0537
pol10	0.0131	0.0034	0.0981	0.0939	-0.0456	0.0461
pol11	-0.0169	-0.0281	0.0529	0.0479	-0.1202	-0.0015
pol12	-0.0175	-0.029	0.0549	0.0495	-0.1269	-0.0085
pol13	-0.0173	-0.0288	0.0551	0.0498	-0.1274	-0.0087
pol14	-0.0169	-0.0284	0.0547	0.0496	-0.1266	-0.007
pol15	-0.017	-0.0285	0.0548	0.0496	-0.1268	-0.0075

Table B.4: Cross-correlation table between groups and intolerance (197 obs)

Variables	%Muslims	%Christians	%None/Atheists
%Christians	-0.739		
%None/Atheists	-0.254	-0.118	
Religious intolerance	0.560	-0.519	-0.057

Table B.5: Religious diversity and incidence of Conflict without intolerance

	(1) rfrac1	(2) rfrac2	(3) rfrac3	(4) rpol1	(5) rpol2	(6) rpol3
Religious Diversity	-1.014	-0.347	-0.110	-0.483	-0.104	0.490
	(0.650)	(0.620)	(0.511)	(0.362)	(0.405)	(0.464)
Lagged civil war	6.261***	6.270***	6.275***	6.263***	6.272***	6.259***
	(0.214)	(0.214)	(0.213)	(0.213)	(0.214)	(0.211)
Log lagged GDP/cap	-0.328**	-0.323**	-0.312**	-0.330**	-0.321**	-0.319**
	(0.144)	(0.144)	(0.142)	(0.143)	(0.143)	(0.141)
Log lagged population	0.317***	0.319***	0.319***	0.314***	0.319***	0.327***
	(0.0640)	(0.0642)	(0.0646)	(0.0644)	(0.0644)	(0.0658)
% mountainous	0.00964**	0.00996***	0.0101***	0.00961**	0.0100***	0.00959**
	(0.00385)	(0.00387)	(0.00382)	(0.00387)	(0.00384)	(0.00390)
Noncontiguos state dummy	0.472	0.497	0.507	0.486	0.506	0.484
	(0.335)	(0.330)	(0.333)	(0.334)	(0.332)	(0.336)
Oil exporter dummy	0.273	0.266	0.262	0.276	0.266	0.243
	(0.272)	(0.268)	(0.269)	(0.270)	(0.269)	(0.259)
New state dummy	1.825***	1.820***	1.821***	1.825***	1.820***	1.820***
	(0.381)	(0.379)	(0.379)	(0.381)	(0.379)	(0.381)
Instability dummy	-0.00683	-0.00135	-0.00429	-0.00515	-0.00240	0.0000379
	(0.279)	(0.278)	(0.277)	(0.278)	(0.277)	(0.277)
Democracy lagged (Polity 2)	0.0132	0.0122	0.0123	0.0132	0.0122	0.0131
	(0.0202)	(0.0200)	(0.0199)	(0.0202)	(0.0200)	(0.0203)
French legal origin dummy	2.304***	2.530***	2.597***	2.332***	2.566***	2.829***
	(0.546)	(0.539)	(0.526)	(0.553)	(0.558)	(0.546)
UK legal origin dummy	2.233***	2.396***	2.455***	2.253***	2.417***	2.611***
	(0.551)	(0.534)	(0.529)	(0.556)	(0.547)	(0.544)
Socialist legal origin dummy	1.850***	1.984***	2.013***	1.845***	1.995***	2.116***
	(0.576)	(0.560)	(0.565)	(0.582)	(0.562)	(0.570)
Latin America and Carribean Dummy	0.169	0.0961	0.125	0.162	0.107	0.0381
	(0.339)	(0.326)	(0.347)	(0.338)	(0.323)	(0.354)
Sub-Saharan Africa dummy	0.610**	0.511	0.465	0.591*	0.466	0.318
	(0.306)	(0.329)	(0.344)	(0.308)	(0.325)	(0.300)
East and Southeast Asia Dummy	0.533	0.399	0.372	0.486	0.371	0.337
	(0.336)	(0.317)	(0.313)	(0.327)	(0.312)	(0.306)
Constant	-7.224***	-7.495***	-7.674***	-7.213***	-7.569***	-8.123***
	(1.345)	(1.366)	(1.377)	(1.354)	(1.386)	(1.426)
Observations	5733	5733	5733	5733	5733	5733
Pseudo R^2	0.743	0.743	0.743	0.743	0.743	0.743
ll	-594.8	-595.8	-596.0	-595.1	-595.9	-595.1

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust Standard errors clustered at the Country level in parentheses.

The dependent variable is the Incidence of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

TABLE B.6: Religious diversity and incidence of Conflict without intolerance

	(1) r1	(2) r2	(3) r3	(4) rr1	(5) rr2	(6) rr3
Religious Diversity	-0.964	-0.441	0.161	-0.443	-0.137	0.569
	(0.671)	(0.609)	(0.525)	(0.371)	(0.397)	(0.455)
Religious intolerance	0.164*	0.175*	0.175*	0.163	0.170*	0.178*
	(0.0995)	(0.0990)	(0.105)	(0.0995)	(0.1000)	(0.101)
Lagged civil war	6.205***	6.211***	6.212***	6.208***	6.215***	6.197***
	(0.209)	(0.210)	(0.207)	(0.209)	(0.209)	(0.205)
Log lagged GDP/cap	-0.234	-0.225	-0.228	-0.235*	-0.226	-0.218
	(0.143)	(0.144)	(0.141)	(0.142)	(0.143)	(0.141)
Log lagged population	0.287***	0.285***	0.289***	0.284***	0.286***	0.292***
	(0.0735)	(0.0743)	(0.0753)	(0.0739)	(0.0746)	(0.0761)
% mountainous	0.00912**	0.00939**	0.00919**	0.00910**	0.00945**	0.00884**
	(0.00394)	(0.00394)	(0.00386)	(0.00396)	(0.00390)	(0.00394)
Noncontiguos state dummy	0.606*	0.630*	0.632*	0.619*	0.639*	0.607*
	(0.336)	(0.332)	(0.331)	(0.334)	(0.333)	(0.335)
Oil exporter dummy	0.106	0.0878	0.0790	0.109	0.0918	0.0520
	(0.298)	(0.295)	(0.292)	(0.295)	(0.294)	(0.284)
New state dummy	1.776***	1.771***	1.772***	1.777***	1.772***	1.773***
	(0.393)	(0.391)	(0.391)	(0.393)	(0.390)	(0.392)
Instability dummy	-0.00590	-0.000832	-0.00272	-0.00418	-0.00177	0.00314
	(0.280)	(0.280)	(0.278)	(0.279)	(0.279)	(0.279)
Democracy lagged (Polity 2)	0.0184	0.0178	0.0179	0.0183	0.0177	0.0193
	(0.0207)	(0.0205)	(0.0206)	(0.0207)	(0.0205)	(0.0209)
French legal origin dummy	2.139***	2.308***	2.456***	2.177***	2.354***	2.657***
	(0.595)	(0.589)	(0.575)	(0.602)	(0.605)	(0.609)
UK legal origin dummy	2.034***	2.145***	2.232***	2.061***	2.173***	2.381***
	(0.616)	(0.608)	(0.604)	(0.621)	(0.617)	(0.625)
Socialist legal origin dummy	1.751***	1.856***	1.912***	1.754***	1.871***	2.009***
	(0.614)	(0.604)	(0.612)	(0.621)	(0.606)	(0.624)
Latin America and Carribean Dummy	0.440	0.383	0.391	0.430	0.388	0.304
	(0.371)	(0.367)	(0.369)	(0.370)	(0.365)	(0.385)
Sub-Saharan Africa dummy	0.990***	0.948**	0.789**	0.964**	0.881**	0.719*
	(0.375)	(0.395)	(0.387)	(0.376)	(0.392)	(0.385)
East and Southeast Asia Dummy	0.645*	0.538	0.460	0.593*	0.495	0.455
	(0.362)	(0.345)	(0.344)	(0.353)	(0.342)	(0.344)
Constant	-7.682***	-7.886***	-8.134***	-7.682***	-7.952***	-8.606***
	(1.423)	(1.453)	(1.456)	(1.431)	(1.470)	(1.516)
Observations	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.744	0.743	0.743	0.744	0.743	0.744
ll	-591.0	-591.7	-592.0	-591.3	-592.0	-590.9

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust Standard errors clustered at the Country level in parentheses.

The dependent variable is the Incidence of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

Table B.7: Religious diversity and Onset of Conflict without intolerance

	(1) rfrac1	(2) rfrac2	(3) rfrac3	(4) rpol1	(5) rpol2	(6) rpol3
Religious Diversity	-1.022	-0.548	0.0277	-0.485	-0.206	0.652
	(0.714)	(0.837)	(0.728)	(0.416)	(0.554)	(0.674)
Lagged civil war	-0.850***	-0.854***	-0.851***	-0.847***	-0.847***	-0.866***
	(0.256)	(0.261)	(0.259)	(0.256)	(0.259)	(0.258)
Log lagged GDP/cap	-0.616***	-0.621***	-0.617***	-0.618***	-0.621***	-0.605***
	(0.149)	(0.148)	(0.148)	(0.149)	(0.148)	(0.146)
Log lagged population	0.297***	0.299***	0.295***	0.292***	0.294***	0.299***
	(0.0727)	(0.0726)	(0.0704)	(0.0717)	(0.0708)	(0.0708)
% mountainous	0.00853*	0.00895*	0.00883*	0.00849*	0.00903*	0.00785
	(0.00494)	(0.00479)	(0.00491)	(0.00495)	(0.00482)	(0.00507)
Noncontiguos state dummy	0.487	0.520	0.514	0.507	0.531	0.456
	(0.354)	(0.358)	(0.361)	(0.353)	(0.364)	(0.358)
Oil exporter dummy	0.724***	0.751***	0.728***	0.730***	0.746***	0.721***
	(0.239)	(0.239)	(0.239)	(0.239)	(0.241)	(0.234)
New state dummy	1.777***	1.769***	1.775***	1.777***	1.770***	1.785***
	(0.371)	(0.368)	(0.370)	(0.371)	(0.367)	(0.374)
Instability dummy	0.625***	0.630***	0.646***	0.631***	0.637***	0.657***
	(0.219)	(0.216)	(0.218)	(0.219)	(0.217)	(0.217)
Democracy lagged (Polity 2)	0.0207	0.0195	0.0195	0.0206	0.0195	0.0203
	(0.0210)	(0.0210)	(0.0207)	(0.0210)	(0.0209)	(0.0211)
French legal origin dummy	1.160*	1.324*	1.478**	1.194*	1.363*	1.757**
	(0.701)	(0.697)	(0.676)	(0.711)	(0.723)	(0.723)
UK legal origin dummy	0.958	1.079	1.172*	0.981	1.091	1.380**
	(0.698)	(0.684)	(0.660)	(0.705)	(0.701)	(0.693)
Socialist legal origin dummy	1.096	1.212*	1.245*	1.095	1.210*	1.322*
	(0.719)	(0.710)	(0.703)	(0.726)	(0.713)	(0.705)
Latin America and Carribean Dummy	0.183	0.111	0.0969	0.171	0.105	-0.0132
	(0.403)	(0.395)	(0.421)	(0.402)	(0.390)	(0.422)
Sub-Saharan Africa dummy	0.394	0.335	0.143	0.369	0.257	-0.0494
	(0.384)	(0.475)	(0.482)	(0.390)	(0.476)	(0.413)
East and Southeast Asia Dummy	0.461	0.353	0.267	0.414	0.309	0.232
	(0.347)	(0.351)	(0.335)	(0.340)	(0.342)	(0.318)
Constant	-3.904**	-4.014**	-4.203***	-3.873**	-4.024**	-4.800***
	(1.629)	(1.622)	(1.610)	(1.641)	(1.650)	(1.710)
Observations	5733	5733	5733	5733	5733	5733
Pseudo R^2	0.098	0.097	0.096	0.098	0.097	0.098
ll	-454.5	-455.1	-455.5	-454.8	-455.4	-454.6

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The dependent variable is the onset of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

TABLE B.8: Religious diversity and Onset of Conflict with intolerance

	(1) rfrac1	(2) rfrac2	(3) rfrac3	(4) rpol1	(5) rpol2	(6) rpol3
Religious Diversity	-0.878	-0.525	0.377	-0.385	-0.147	0.758
	(0.760)	(0.812)	(0.732)	(0.441)	(0.539)	(0.633)
Rel intolerance	0.209**	0.221**	0.240**	0.210**	0.221**	0.241**
	(0.104)	(0.103)	(0.111)	(0.104)	(0.105)	(0.106)
Lagged civil war	-0.900***	-0.902***	-0.899***	-0.897***	-0.896***	-0.913***
	(0.246)	(0.250)	(0.244)	(0.246)	(0.247)	(0.243)
Log lagged GDP/cap	-0.537***	-0.538***	-0.537***	-0.538***	-0.539***	-0.517***
	(0.159)	(0.158)	(0.156)	(0.159)	(0.158)	(0.157)
Log lagged population	0.238***	0.237***	0.227***	0.233***	0.231***	0.230***
	(0.0853)	(0.0861)	(0.0824)	(0.0842)	(0.0841)	(0.0827)
%mountainous	0.00755	0.00795	0.00735	0.00754	0.00795	0.00645
	(0.00562)	(0.00547)	(0.00567)	(0.00563)	(0.00550)	(0.00573)
Noncontiguos state dummy	0.609*	0.648*	0.643*	0.628*	0.656*	0.600*
	(0.353)	(0.354)	(0.348)	(0.351)	(0.357)	(0.349)
Oil exporter dummy	0.543**	0.566**	0.530**	0.550**	0.558**	0.522**
	(0.256)	(0.250)	(0.248)	(0.254)	(0.251)	(0.241)
New state dummy	1.728***	1.720***	1.726***	1.728***	1.723***	1.736***
	(0.381)	(0.377)	(0.380)	(0.381)	(0.377)	(0.383)
Instability dummy	0.624***	0.626***	0.645***	0.629***	0.634***	0.647***
	(0.220)	(0.217)	(0.219)	(0.220)	(0.217)	(0.219)
Democracy lagged (Polity 2)	0.0272	0.0265	0.0267	0.0271	0.0265	0.0284
	(0.0207)	(0.0207)	(0.0206)	(0.0207)	(0.0206)	(0.0209)
French legal origin dummy	0.996	1.102	1.297*	1.038	1.155	1.532*
	(0.782)	(0.776)	(0.755)	(0.789)	(0.791)	(0.820)
UK legal origin dummy	0.696	0.762	0.826	0.721	0.780	1.029
	(0.789)	(0.780)	(0.758)	(0.793)	(0.789)	(0.804)
Socialist legal origin dummy	1.011	1.091	1.102	1.015	1.090	1.186
	(0.785)	(0.780)	(0.768)	(0.789)	(0.779)	(0.786)
Latin America and Carribean Dummy	0.549	0.512	0.491	0.538	0.506	0.411
	(0.409)	(0.400)	(0.403)	(0.408)	(0.395)	(0.423)
Sub-Saharan Africa dummy	0.825*	0.829*	0.555	0.794*	0.729	0.479
	(0.424)	(0.492)	(0.466)	(0.428)	(0.488)	(0.443)
East and Southeast Asia Dummy	0.595	0.516	0.389	0.542	0.456	0.379
	(0.385)	(0.379)	(0.360)	(0.376)	(0.372)	(0.352)
Constant	-4.029**	-4.105**	-4.337**	-4.012**	-4.120**	-4.916***
	(1.721)	(1.723)	(1.717)	(1.732)	(1.735)	(1.826)
Observations	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.102	0.101	0.101	0.101	0.101	0.103
ll	-451.7	-452.1	-452.2	-452.0	-452.3	-451.2

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The dependent variable is the onset of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

Table B.9: Percentage of different groups (Aggregation Level 2) with Intolerance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Religious fractionalization	-3.805*	-3.834*	-3.782*	-3.794*	-3.844*	-3.879*	-3.908*
	(2.187)	(2.190)	(2.150)	(2.176)	(2.219)	(2.172)	(2.198)
Religious polarization	2.363*	2.231*	2.215	2.333*	2.353*	2.114	2.218
	(1.358)	(1.343)	(1.353)	(1.360)	(1.361)	(1.374)	(1.379)
Rel intolerance	0.236**	0.268**	0.240**	0.219*	0.253**	0.273**	0.257**
	(0.108)	(0.120)	(0.112)	(0.117)	(0.123)	(0.121)	(0.124)
Percentage Nonreligious/Atheists	-0.00753			-0.00825	-0.00751		-0.00945
	(0.00945)			(0.00949)	(0.00947)		(0.00954)
Percentage Muslims		-0.00118			-0.00115	-0.00364	-0.00454
		(0.00417)			(0.00413)	(0.00635)	(0.00612)
Percentage Christians			-0.000990	-0.00166		-0.00339	-0.00472
			(0.00419)	(0.00422)		(0.00639)	(0.00630)
Observations	5678	5678	5678	5678	5678	5678	5678
Pseudo R^2	0.104	0.104	0.104	0.104	0.104	0.104	0.105
ll	-450.7	-450.9	-450.9	-450.6	-450.6	-450.8	-450.3

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust standard errors, clustered at the level of countries, in parentheses.

The dependent variable is the onset of civil conflict. This Table uses Religious Diversity calculated at aggregation level 1. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section. The other controls are: a constant term, lagged civil war, the log of per capita GDP (lagged), the percentage of the country that is mountainous, non-contiguous state dummy, oil exporter dummy, new state dummy, Instability dummy, democracy lagged (polity2), continent dummy variables for Sub-Saharan Africa, East and Southeast Asia, Latin America and the Caribbean, and legal origin dummies from La Porta et al. [1999].

Table B.10: Percentage of different groups (Aggregation Level 2) without Intolerance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Religious fractionalization	-2.891	-2.932	-2.979	-3.048	-2.966	-2.982	-3.048
	(2.130)	(2.095)	(2.076)	(2.100)	(2.124)	(2.077)	(2.101)
Religious polarization	1.783	1.760	1.584	1.802	1.919	1.651	1.793
	(1.341)	(1.298)	(1.284)	(1.304)	(1.326)	(1.337)	(1.343)
Percentage Nonreligious/Atheists	-0.00995			-0.0112	-0.00922		-0.0112
	(0.00957)			(0.00918)	(0.00958)		(0.00956)
Percentage Muslims		0.00359			0.00325	0.00127	-0.000181
		(0.00356)			(0.00359)	(0.00579)	(0.00561)
Percentage Christians			-0.00433	-0.00478		-0.00335	-0.00492
			(0.00388)	(0.00386)		(0.00628)	(0.00608)
Observations	5733	5733	5733	5733	5733	5733	5733
Pseudo R^2	0.099	0.099	0.099	0.101	0.100	0.099	0.101
ll	-454.0	-454.1	-453.9	-453.2	-453.6	-453.9	-453.2

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The dependent variable is the onset of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section. The other controls are: a constant term, lagged civil war, the log of per capita GDP (lagged), the percentage of the country that is mountainous, non-contiguous state dummy, oil exporter dummy, new state dummy, Instability dummy, democracy lagged (polity2), continent dummy variables for Sub-Saharan Africa, East and Southeast Asia, Latin America and the Caribbean, and legal origin dummies from La Porta et al. [1999].

Table B.11: Components of intolerance in Montalvo and Reynal-Querol [2005] specification

	(1) Level1	(2) Level2	(3) Level3	(4) Level1	(5) Level2	(6) Level3
religious fractionalization	-33.14***	-11.39***	-3.563*	-47.53**	-14.44***	-2.859
	(12.55)	(3.765)	(1.923)	(20.52)	(4.464)	(2.223)
Religious polarization	15.68**	6.924***	3.809**	23.14**	8.855***	3.566*
	(6.803)	(2.296)	(1.806)	(10.86)	(2.866)	(2.002)
ETHPOL	2.175**	2.276**	2.360**	2.051*	1.935*	2.134*
	(1.088)	(1.109)	(1.132)	(1.134)	(1.148)	(1.192)
ETHFRAC	0.257	0.526	0.270	0.996	1.440	0.602
	(0.920)	(0.968)	(1.004)	(0.986)	(0.988)	(1.044)
Government Regulation Index				**0.179***	**0.201****	**0.151***
				(0.0947)	(0.0937)	(0.0915)
N	846	846	846	838	838	838
pseudo R^2	0.178	0.154	0.157	0.210	0.191	0.176
ll	-294.1	-302.7	-301.5	-281.8	-288.4	-293.8
religious fractionalization	-33.14***	-11.39***	-3.563*	-50.09**	-12.80***	-2.030
	(12.55)	(3.765)	(1.923)	(24.49)	(4.140)	(1.966)
Religious polarization	15.68**	6.924***	3.809**	24.60*	7.932***	3.207*
	(6.803)	(2.296)	(1.806)	(12.83)	(2.731)	(1.709)
ETHPOL	2.175**	2.276**	2.360**	1.927*	1.908*	1.850*
	(1.088)	(1.109)	(1.132)	(1.086)	(1.075)	(1.115)
ETHFRAC	0.257	0.526	0.270	1.020	1.248	0.480
	(0.920)	(0.968)	(1.004)	(1.019)	(0.984)	(1.038)
Social Regulation Index				**0.222****	**0.248*****	**0.246*****
				(0.0986)	(0.0916)	(0.0910)
N	846	846	846	838	838	838
pseudo R^2	0.178	0.154	0.157	0.226	0.217	0.211
ll	-294.1	-302.7	-301.5	-275.7	-279.2	-281.3
religious fractionalization	-33.14***	-11.39***	-3.563*	-39.72***	-12.80***	-2.720
	(12.55)	(3.765)	(1.923)	(14.68)	(4.038)	(2.005)
Religious polarization	15.68**	6.924***	3.809**	19.30**	8.068***	3.505*
	(6.803)	(2.296)	(1.806)	(7.938)	(2.633)	(1.875)
ETHPOL	2.175**	2.276**	2.360**	2.139*	2.135*	2.190*
	(1.088)	(1.109)	(1.132)	(1.162)	(1.161)	(1.162)
ETHFRAC	0.257	0.526	0.270	0.553	0.981	0.461
	(0.920)	(0.968)	(1.004)	(0.961)	(0.997)	(1.023)
Government Favoritism Index				0.0802	0.114	0.0990
				(0.0962)	(0.0962)	(0.0999)
N	846	846	846	838	838	838
pseudo R^2	0.178	0.154	0.157	0.191	0.171	0.164
ll	-294.1	-302.7	-301.5	-288.5	-295.6	-297.9

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The dependent variable is the incidence of civil war (intermediate and high-intensity civil wars of PRIO). Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Montalvo and Reynal-Querol [2005] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section. The sample is divided into 5 year periods. The other controls are: a constant term, the log of per capita GDP, the log of population (both at the beginning of the period), the percentage of the country that is mountainous, non-contiguous state dummy, level of democracy (Polity IV dataset).

Table B.12: Controlling ETHFRAC and ETHPOL in Desmet et al. [2012] (Incidence)

	(1) r1	(2) r2	(3) r3	(4) r4	(5) r5	(6) r6
Religious fractionalization 2	-4.150*	-4.130*	-3.998*	-3.879	-3.360	-3.642
	(2.261)	(2.247)	(2.264)	(2.582)	(2.474)	(2.729)
Religious polarization 2	2.426*	2.402*	2.302*	2.029	1.733	1.884
	(1.388)	(1.380)	(1.396)	(1.700)	(1.642)	(1.765)
elf1	0.972*		-2.818			
	(0.573)		(3.200)			
Religious intolerance	0.221**	0.219**	0.222**	0.196	0.172	0.179
	(0.105)	(0.105)	(0.106)	(0.120)	(0.120)	(0.127)
pol1		0.607*	2.208			
		(0.330)	(1.748)			
Ethnic fractionalization MRQ				0.508		0.333
				(0.546)		(0.761)
Ethnic polarization MRQ					0.531	0.345
					(0.549)	(0.769)
Observations	5678	5678	5678	4898	4898	4898
Pseudo R^2	0.106	0.106	0.107	0.104	0.104	0.104
ll	-449.7	-449.5	-449.3	-391.6	-391.6	-391.5
Religious fractionalization 3	-1.113	-1.099	-1.044	-2.304*	-2.272*	-2.266*
	(1.287)	(1.292)	(1.301)	(1.248)	(1.212)	(1.210)
Religious polarization 3	1.402	1.377	1.322	2.488**	2.435**	2.435**
	(0.966)	(0.970)	(0.975)	(1.055)	(1.030)	(1.027)
elf1	0.768		-2.744			
	(0.593)		(3.398)			
Religious intolerance	0.184	0.182	0.189*	0.141	0.119	0.118
	(0.114)	(0.114)	(0.114)	(0.113)	(0.117)	(0.118)
pol1		0.487	2.045			
		(0.340)	(1.873)			
Ethnic fractionalization MRQ				0.199		-0.0508
				(0.522)		(0.720)
Ethnic polarization MRQ					0.472	0.501
					(0.598)	(0.802)
Observations	5678	5678	5678	4898	4898	4898
Pseudo R^2	0.105	0.106	0.106	0.107	0.108	0.108
ll	-450.0	-449.8	-449.6	-390.0	-389.7	-389.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust Standard errors clustered at the Country level in parentheses.

The dependent variable is the Incidence of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

TABLE B.13: Controlling ETHFRAC and ETHPOL in Desmet et al. [2012] (Onset)

	(1)	(2)	(3)	(4)	(5)	(6)
	r1	r2	r3	r4	r5	r6
Religious fractionalization 2	-4.150*	-4.130*	-3.998*	-3.879	-3.360	-3.642
	(2.261)	(2.247)	(2.264)	(2.582)	(2.474)	(2.729)
Religious polarization 2	2.426*	2.402*	2.302*	2.029	1.733	1.884
	(1.388)	(1.380)	(1.396)	(1.700)	(1.642)	(1.765)
elf1	0.972*		-2.818			
	(0.573)		(3.200)			
Religious intolerance	0.221**	0.219**	0.222**	0.196	0.172	0.179
	(0.105)	(0.105)	(0.106)	(0.120)	(0.120)	(0.127)
pol1		0.607*	2.208			
		(0.330)	(1.748)			
Ethnic fractionalization MRQ				0.508		0.333
				(0.546)		(0.761)
Ethnic polarization MRQ					0.531	0.345
					(0.549)	(0.769)
Observations	5678	5678	5678	4898	4898	4898
Pseudo R^2	0.106	0.106	0.107	0.104	0.104	0.104
ll	-449.7	-449.5	-449.3	-391.6	-391.6	-391.5
Religious fractionalization 3	-1.113	-1.099	-1.044	-2.304*	-2.272*	-2.266*
	(1.287)	(1.292)	(1.301)	(1.248)	(1.212)	(1.210)
Religious polarization 3	1.402	1.377	1.322	2.488**	2.435**	2.435**
	(0.966)	(0.970)	(0.975)	(1.055)	(1.030)	(1.027)
elf1	0.768		-2.744			
	(0.593)		(3.398)			
Religious intolerance	0.184	0.182	0.189*	0.141	0.119	0.118
	(0.114)	(0.114)	(0.114)	(0.113)	(0.117)	(0.118)
pol1		0.487	2.045			
		(0.340)	(1.873)			
Ethnic fractionalization MRQ				0.199		-0.0508
				(0.522)		(0.720)
Ethnic polarization MRQ					0.472	0.501
					(0.598)	(0.802)
Observations	5678	5678	5678	4898	4898	4898
Pseudo R^2	0.105	0.106	0.106	0.107	0.108	0.108
ll	-450.0	-449.8	-449.6	-390.0	-389.7	-389.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust Standard errors clustered at the Country level in parentheses.

The dependent variable is the onset of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

TABLE B.14: Controlling for percentage of different groups in Montalvo and Reynal-Querol [2005] (Aggregation Level 2)

	(1) c1	(2) c2	(3) c3	(4) c4	(5) c5	(6) c6	(7) c7
Religious polarization	8.151***	8.948***	8.970***	8.237***	8.978***	8.155***	8.165***
	(2.808)	(2.914)	(2.871)	(2.782)	(2.877)	(2.776)	(2.717)
religious fractionalization	-12.58***	-14.56***	-14.19***	-12.57***	-14.45***	-12.90***	-12.88***
	(4.117)	(4.462)	(4.370)	(4.109)	(4.566)	(4.108)	(4.200)
LGDPC	-0.124	-0.286	-0.283	-0.159	-0.294	-0.168	-0.169
	(0.242)	(0.235)	(0.258)	(0.263)	(0.248)	(0.237)	(0.250)
LPOP	0.438***	0.306*	0.335**	0.445***	0.316*	0.419**	0.420**
	(0.166)	(0.171)	(0.168)	(0.167)	(0.187)	(0.166)	(0.178)
PRIMEXP	-1.559	-1.036	-1.117	-1.516	-1.046	-1.408	-1.409
	(1.731)	(1.622)	(1.709)	(1.761)	(1.633)	(1.658)	(1.658)
MOUNTAINS	-0.00765	-0.00886	-0.00846	-0.00935	-0.00920	-0.0107	-0.0108
	(0.0105)	(0.0115)	(0.0108)	(0.0109)	(0.0114)	(0.0117)	(0.0118)
NONCONT	0.0218	0.0257	0.00640	-0.0121	0.0105	-0.00224	-0.00513
	(0.573)	(0.545)	(0.547)	(0.575)	(0.545)	(0.568)	(0.572)
DEMOCRACY	0.242	0.229	0.285	0.214	0.239	0.119	0.121
	(0.382)	(0.396)	(0.410)	(0.393)	(0.383)	(0.380)	(0.368)
ETHPOL	2.264*	1.783	1.853	2.301**	1.819	2.265*	2.270*
	(1.191)	(1.142)	(1.137)	(1.173)	(1.163)	(1.172)	(1.195)
ETHFRAC	1.097	1.485	1.358	0.979	1.417	1.082	1.069
	(1.040)	(1.000)	(1.015)	(1.049)	(1.071)	(1.023)	(1.116)
Rel intolerance	0.366*	0.525**	0.519***	0.433**	0.540***	0.464**	0.467**
	(0.187)	(0.229)	(0.188)	(0.182)	(0.209)	(0.221)	(0.201)
Percentage Nonreligious/Atheists	-0.0336*			-0.0337		-0.0362*	-0.0361*
	(0.0202)			(0.0205)		(0.0209)	(0.0212)
Percentage Muslims		-0.00489			-0.00345	-0.00656	-0.00629
		(0.00697)			(0.0105)	(0.00712)	(0.0103)
Percentage Christians			0.00454	0.00421	0.00243		0.000458
			(0.00659)	(0.00629)	(0.0101)		(0.00960)
_cons	-9.723***	-6.240*	-7.181**	-9.748***	-6.537*	-8.733***	-8.781**
	(3.159)	(3.197)	(2.925)	(3.153)	(3.509)	(3.317)	(3.517)
N	838	838	838	838	838	838	838
pseudo R^2	0.213	0.204	0.204	0.215	0.204	0.217	0.217
ll	-280.5	-283.8	-283.9	-279.9	-283.7	-279.2	-279.2

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

The dependent variable is the onset of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Montalvo and Reynal-Querol [2005] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

Table B.15: Correlates of Incidence of Civil wars - Marginal Effects

	(1) Level1	(2) Level2	(3) Level3	(4) Level1	(5) Level2	(6) Level3
Religious fractionalization	-0.185*	-0.0749*	-0.0368*	-0.234**	-0.100**	-0.0265
	(0.105)	(0.0446)	(0.0210)	(0.106)	(0.0447)	(0.0218)
Religious polarization	0.0894	0.0460	0.0369**	0.117**	0.0614**	0.0313*
	(0.0582)	(0.0282)	(0.0164)	(0.0586)	(0.0284)	(0.0167)
Lagged civil war	0.826***	0.831***	0.827***	0.816***	0.822***	0.821***
	(0.0224)	(0.0211)	(0.0210)	(0.0233)	(0.0222)	(0.0214)
Log lagged GDP/cap	-0.00718**	-0.00707**	-0.00538*	-0.00456	-0.00428	-0.00410
	(0.00335)	(0.00336)	(0.00323)	(0.00335)	(0.00336)	(0.00327)
Log lagged population	0.00810***	0.00791***	0.00743***	0.00749***	0.00707***	0.00689***
	(0.00188)	(0.00164)	(0.00166)	(0.00200)	(0.00178)	(0.00188)
% mountainous	0.000242***	0.000217**	0.000246***	0.000234**	0.000199**	0.000227**
	(9.08e-05)	(9.41e-05)	(9.06e-05)	(9.31e-05)	(9.80e-05)	(9.12e-05)
Noncontiguos state dummy	0.00967	0.00981	0.0124	0.0130	0.0129	0.0150
	(0.00882)	(0.00864)	(0.00884)	(0.00918)	(0.00890)	(0.00925)
Oil exporter dummy	0.00568	0.00536	0.00473	0.000684	3.34e-05	0.00129
	(0.00684)	(0.00646)	(0.00625)	(0.00716)	(0.00660)	(0.00672)
New state dummy	0.0678***	0.0688***	0.0685***	0.0645***	0.0657***	0.0661***
	(0.0236)	(0.0241)	(0.0240)	(0.0236)	(0.0240)	(0.0240)
Instability dummy	-0.000385	-0.000293	0.000298	-0.000471	-0.000337	0.000210
	(0.00653)	(0.00647)	(0.00658)	(0.00662)	(0.00657)	(0.00666)
Democracy lagged (Polity 2)	0.000292	0.000286	0.000378	0.000435	0.000453	0.000485
	(0.000479)	(0.000475)	(0.000479)	(0.000495)	(0.000490)	(0.000496)
French legal origin dummy	0.0895***	0.118***	0.125***	0.0786**	0.105**	0.111**
	(0.0339)	(0.0410)	(0.0455)	(0.0334)	(0.0414)	(0.0452)
UK legal origin dummy	0.0959**	0.127**	0.141**	0.0805**	0.111**	0.118**
	(0.0407)	(0.0496)	(0.0591)	(0.0402)	(0.0509)	(0.0590)
Socialist legal origin dummy	0.0852**	0.0939**	0.0975*	0.0798*	0.0863*	0.0881*
	(0.0425)	(0.0435)	(0.0507)	(0.0428)	(0.0441)	(0.0498)
Latin America and Carribean Dummy	0.00386	0.00246	-0.000749	0.0120	0.0110	0.00514
	(0.00829)	(0.00785)	(0.00841)	(0.0101)	(0.00969)	(0.00968)
Sub-Saharan Africa dummy	0.0137*	0.0116	0.0147	0.0265**	0.0263**	0.0223*
	(0.00819)	(0.00854)	(0.00964)	(0.0122)	(0.0128)	(0.0119)
East and Southeast Asia Dummy	0.0189*	0.0122	0.0123	0.0254*	0.0185*	0.0140
	(0.0115)	(0.00890)	(0.00904)	(0.0140)	(0.0107)	(0.0101)
Religious intolerance				0.00454*	0.00488**	0.00352
				(0.00237)	(0.00227)	(0.00254)
Observations	5733	5733	5733	5678	5678	5678
Pseudo R^2	0.744	0.743	0.744	0.745	0.744	0.744
ll	-593.8	-594.7	-592.6	-589.4	-590.0	-589.7

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors, clustered at the level of countries, in parentheses.

This table replicates the Table 2.4, but reports the marginal effects instead of the coefficients. The dependent variable is the incidence of civil conflict. Column 1 (4) , 2 (5) and 3 (6) correspond to religious diversity measured at the 1^{st}, 2^{nd} and 3^{rd} levels of aggregation respectively. All the data are from Desmet et al. [2012] except for the measures of religious diversity and religious intolerance which are based on the author's own calculations using data from myriad sources explained in detail in the data section.

Table B.16: List of Religions

Religion Level 1	**Religion Level 2**	**Religion Level 3**
Abrahamic	Baha'i	Baha'i
Abrahamic	Christian	African Christian
Abrahamic	Christian	African Methodist Episcopal
Abrahamic	Christian	African Protestant
Abrahamic	Christian	Aglipayan
Abrahamic	Christian	Albanian Orthodox
Abrahamic	Christian	Anglican
Abrahamic	Christian	Apostolic Faith
Abrahamic	Christian	Armenian Apostolic (Orthodox)
Abrahamic	Christian	Armenian Gregorian
Abrahamic	Christian	Assemblies of God
Abrahamic	Christian	Baptist
Abrahamic	Christian	Belarusian Orthodox
Abrahamic	Christian	Black Independent Churches
Abrahamic	Christian	Bulgarian Orthodox
Abrahamic	Christian	Roman Catholic
Abrahamic	Christian	Christian
Abrahamic	Christian	Christian unaffiliated
Abrahamic	Christian	Church of Christ
Abrahamic	Christian	Congregational
Abrahamic	Christian	Coptic Orthodox
Abrahamic	Christian	Czechoslovak Hussite
Abrahamic	Christian	Dutch Reformed Church
Abrahamic	Christian	Eastern Orthodox
Abrahamic	Christian	Eritrean Orthodox
Abrahamic	Christian	Estonian Orthodox
Abrahamic	Christian	Ethiopian Orthodox
Abrahamic	Christian	Evangelical Church of Czech Brethren
Abrahamic	Christian	Evangelical Lutheran
Abrahamic	Christian	Evangelical Protestant
Abrahamic	Christian	Free Wesleyan
Abrahamic	Christian	Full Gospel
Abrahamic	Christian	Georgian Orthodox
Abrahamic	Christian	Greek Catholic (Melchite)
Abrahamic	Christian	Greek Orthodox
Abrahamic	Christian	Independent
Abrahamic	Christian	Kimbanguist
Abrahamic	Christian	Lutheran
Abrahamic	Christian	Methodist
Abrahamic	Christian	Mormon
Abrahamic	Christian	New Apostolic
Abrahamic	Christian	Orthodox
Abrahamic	Christian	Other Apostolic
Abrahamic	Christian	Other Black Independent
Abrahamic	Christian	Other Christian
Abrahamic	Christian	other Protestant
Abrahamic	Christian	Pentecostal
Abrahamic	Christian	Polish Orthodox
Abrahamic	Christian	Presbyterian
Abrahamic	Christian	Protestant
Abrahamic	Christian	Reformed Churches
Abrahamic	Christian	Romanian Orthodox
Abrahamic	Christian	Russian Orthodox
Abrahamic	Christian	Salvation Army
Abrahamic	Christian	Serbian Orthodox
Abrahamic	Christian	Seventh Day Adventist
Abrahamic	Christian	Silesian Evangelical
Abrahamic	Christian	Slovak Evangelical
Abrahamic	Christian	Swiss Christian

Table B.17: List of Religions

Religion Level 1	**Religion Level 2**	**Religion Level 3**
Abrahamic	Christian	Ukrainian Catholic
Abrahamic	Christian	Ukrainian Orthodox (Autocephalous)
Abrahamic	Christian	Ukrainian Orthodox (Kiev)
Abrahamic	Christian	Ukrainian Orthodox (Russian)
Abrahamic	Christian	United Congregational
Abrahamic	Christian	Uniting Church
Abrahamic	Druze	Druze
Abrahamic	Jewish	Jewish
Abrahamic	Muslim	Ibadiyah Muslim
Abrahamic	Muslim	Muslim
Abrahamic	Muslim	other Muslim
Abrahamic	Muslim	Shii Muslim
Abrahamic	Muslim	Sunni Muslim
Indian	Buddhist	Buddhist
Indian	Buddhist	Hoa Hao
Indian	Buddhist	Lamaistic Buddhist
Indian	Buddhist	Tantric Buddhist
Indian	Hindu	Hindu
Indian	Jain	Jain
Indian	Sikh	Sikh
Indigenous	Animist	Animist
Indigenous	Buddhist and Taoism	Buddhist and Taoism
Indigenous	Burkinan Traditional	Burkinan Traditional
Indigenous	Chinese Folk	Chinese Folk
Indigenous	Chondogyo	Chondogyo
Indigenous	Confucian	Confucian
Indigenous	Ethnic Religionist	Ethnic Religionist
Indigenous	Indigenous	Indigenous
Indigenous	GB Traditional	GB Traditional
Indigenous	Hsuan Yuan Chiao	Hsuan Yuan Chiao
Indigenous	Indigenous	Cao Dai
Indigenous	Indigenous	I Kuan Tao
Indigenous	Indigenous	Tien Te Chiao
Indigenous	Ivoirian Traditional	Ivoirian Traditional
Indigenous	Laos Traditional	Laos Traditional
Indigenous	Madagascar Traditional	Madagascar Traditional
Indigenous	Malawi Traditional	Malawi Traditional
Indigenous	Modekngei (Indigenous)	Modekngei (Indigenous)
Indigenous	Mozambique Traditional	Mozambique Traditional
Indigenous	Myanmar Traditional	Myanmar Traditional
Indigenous	Niger Traditional	Niger Traditional
Indigenous	Nigeria Traditional	Nigeria Traditional
Indigenous	NK Traditional	NK Traditional
Indigenous	Ratana	Ratana
Indigenous	Senegal Traditional	Senegal Traditional
Indigenous	Shintoist	Shintoist
Indigenous	SL Traditional	SL Traditional
Indigenous	Swaziland Traditional	Swaziland Traditional
Indigenous	Tanzania Traditional	Tanzania Traditional
Indigenous	Taoist	Taoist
Indigenous	Togo Traditional	Togo Traditional
Indigenous	Traditional	Traditional
Indigenous	Voodoo	Voodoo
Indigenous	Wonbulgyo	Wonbulgyo
Indigenous	Zambia Traditional	Zambia Traditional
Indigenous	Zimbabwe Traditional	Zimbabwe Traditional
Iranian	Zoroastrian	Zoroastrian
Other	New Religionist	New Religionist
Other	Other	Other

Table B.18: Ranking of countries by religious fractionalization (High to Low)

Ranks	rfrac1	rfrac2	rfrac3
1	Macau	Benin	Jamaica
2	China	Singapore	Antigua and Barbuda
3	Mongolia	Taiwan	Papua New Guinea
4	Singapore	Malawi	New Zealand
5	Taiwan	Cote d'Ivoire	Trinidad and Tobago
6	Laos	Tanzania	South Africa
7	Japan	Nigeria	Guyana
8	Vietnam	Suriname	Malawi
9	Mauritius	Macau	Ghana
10	Korea, South	Mauritius	Vanuatu
11	Trinidad and Tobago	Zimbabwe	Benin
12	Togo	Cameroon	Solomon Islands
13	Angola	Central African Republic	United States
14	Estonia	Togo	Samoa
15	French Guiana	China	American Samoa
16	Benin	Guinea-Bissau	Lebanon
17	Tonga	Malaysia	Bahamas, The
18	Guinea-Bissau	Mongolia	Bermuda
19	Malawi	Trinidad and Tobago	Zambia
20	Isle of Man	Lebanon	Saint Kitts and Nevis
21	Guyana	Chad	Barbados
22	Suriname	Korea, North	Kenya
23	Madagascar	Guyana	Fiji
24	Central African Republic	Burkina Faso	Suriname
25	Tanzania	Brunei	Australia
26	Botswana	Laos	Mozambique
27	Brunei	Japan	Moldova
28	Kenya	Sierra Leone	Nigeria
29	Cuba	Vietnam	Cameroon
30	Congo, Republic of the	Kenya	Botswana
31	Malaysia	Korea, South	Central African Republic
32	Sierra Leone	Eritrea	Lesotho
33	Lesotho	Madagascar	Swaziland
34	Slovenia	Angola	Ukraine
35	Hong Kong	Estonia	Grenadine
36	Fiji	French Guiana	Cote d'Ivoire
37	Czech Republic	Fiji	Tanzania
38	American Samoa	Bosnia and Herzegovina	Nauru
39	Cote d'Ivoire	Ethiopia	Singapore
40	New Zealand	Liberia	Congo, Democratic Republic of the
41	Zimbabwe	Burundi	Taiwan
42	Cameroon	Sri Lanka	Virgin Islands
43	Nigeria	Tonga	Mauritius
44	Burundi	Russia	Uganda
45	Northern Mariana Islands	Isle of Man	Netherlands
46	Namibia	Mozambique	Belize
47	Liberia	Macedonia	Zimbabwe
48	Sudan	Congo, Republic of the	Congo, Republic of the
49	Jamaica	Botswana	Namibia
50	CZECHOSLOVAKIA	Slovenia	Gabon
51	Liechtenstein	Cuba	Ethiopia
52	Vanuatu	Sudan	Togo
53	Faroe Islands	Ghana	Palau
54	Burkina Faso	East Timor	Macau
55	Belize	Kuwait	Guernsey
56	India	Albania	Bosnia and Herzegovina
57	Sri Lanka	Lesotho	Jersey
58	Bermuda	Hong Kong	Germany
59	Nauru	New Zealand	Latvia
60	Palau	Czech Republic	Vietnam

Ranking of countries by religious fractionalization (High to Low)

Ranks	rfrac1	rfrac2	rfrac3
61	Zambia	American Samoa	Chad
62	Qatar	Israel	Grenada
63	Saint Kitts and Nevis	Gabon	Korea, South
64	Kuwait	Bhutan	Canada
65	Hungary	Qatar	Tonga
66	Guam	Yugoslavia	Estonia
67	Sweden	Congo, Democratic Republic of the	Angola
68	Dominican Republic	Moldova	Marshall Islands
69	Bangladesh	Northern Mariana Islands	Kuwait
70	Niger	India	New Caledonia
71	Moldova	Namibia	China
72	Swaziland	Montenegro	Eritrea
73	French Polynesia	West Bank	Rwanda
74	Gabon	Bahrain	Guinea-Bissau
75	Burma (Myanmar)	Zambia	Uruguay
76	Guernsey	Kazakhstan	Switzerland
77	Chad	Cyprus	Hungary
78	Congo, Democratic Republic of the	Jamaica	Malaysia
79	Ghana	CZECHOSLOVAKIA	Mongolia
80	San Marino	Nepal	Burkina Faso
81	Mali	Liechtenstein	Northern Mariana Islands
82	Mozambique	Sweden	Kiribati
83	Oman	Vanuatu	Korea, North
84	Norway	Faroe Islands	Bahrain
85	Panama	Belize	United Kingdom
86	Bahrain	Guinea	French Polynesia
87	Guinea	Bermuda	Dominica
88	Virgin Islands	Nauru	Brunei
89	Korea, North	Palau	Burundi
90	Grenadine	Netherlands	Laos
91	Australia	Uganda	Micronesia
92	Lithuania	Panama	Japan
93	Barbados	Saint Kitts and Nevis	Sierra Leone
94	Thailand	Indonesia	Cuba
95	Aruba	Hungary	Russia
96	Nepal	Oman	Madagascar
97	Costa Rica	Guam	Guatemala
98	Chile	Palestine	Yemen
99	New Caledonia	Bulgaria	Slovenia
100	Equatorial Guinea	Swaziland	Czech Republic
101	Ecuador	Bangladesh	French Guiana
102	Austria	Dominican Republic	Azerbaijan
103	United Arab Emirates	Syria	Belarus
104	Iceland	Niger	Liberia
105	Micronesia	Burma (Myanmar)	Sri Lanka
106	Sao Tome	Reunion	Iraq
107	Israel	Georgia	Isle of Man
108	Netherlands	French Polynesia	Macedonia
109	Seychelles	Grenadine	Saint Lucia
110	Slovakia	Gambia, The	El Salvador
111	Cambodia	Mali	CZECHOSLOVAKIA
112	Peru	Norway	Albania
113	Canada	Guernsey	Nicaragua
114	Ethiopia	Egypt	Panama
115	Papua New Guinea	San Marino	Yugoslavia
116	Indonesia	France	Oman
117	Pakistan	Austria	Sudan
118	Lebanon	Australia	East Timor
119	Haiti	Kosovo	Hong Kong
120	Guatemala	Equatorial Guinea	Guam

Ranking of countries by religious fractionalization (High to Low)

Ranks	rfrac1	rfrac2	rfrac3
121	Martinique	Jordan	Puerto Rico
122	Honduras	New Caledonia	Israel
123	El Salvador	Belgium	Sweden
124	Argentina	Kyrgyzstan	Chile
125	Solomon Islands	Virgin Islands	Liechtenstein
126	Brazil	Philippines	Syria
127	United States	Azerbaijan	Pakistan
128	Ireland	Switzerland	Bhutan
129	Guadeloupe	Lithuania	Qatar
130	Antigua and Barbuda	Senegal	Montenegro
131	Puerto Rico	Barbados	Slovakia
132	Nicaragua	United States	Cyprus
133	Gambia, The	Djibouti	Costa Rica
134	Venezuela	Germany	Brazil
135	Denmark	Canada	Haiti
136	Jordan	Aruba	Bolivia
137	Iran	Thailand	United Arab Emirates
138	Eritrea	Pakistan	Philippines
139	Paraguay	Denmark	India
140	Samoa	Costa Rica	Afghanistan
141	Philippines	Chile	Dominican Republic
142	Colombia	Seychelles	West Bank
143	Azerbaijan	Argentina	Kazakhstan
144	Kiribati	Rwanda	Bulgaria
145	Uganda	Grenada	Seychelles
146	Dominica	Serbia and Montenegro	Nepal
147	Macedonia	United Kingdom	Equatorial Guinea
148	Grenada	Ecuador	Netherlands Antilles
149	United Kingdom	Kiribati	Peru
150	France	United Arab Emirates	Aruba
151	Netherlands Antilles	Iceland	Georgia
152	Uruguay	Libya	Faroe Islands
153	Mayotte	Micronesia	Guinea
154	Armenia	Sao Tome	Austria
155	Libya	Slovakia	Norway
156	Switzerland	Papua New Guinea	Martinique
157	Portugal	Mayotte	Romania
158	Saint Lucia	Cambodia	Honduras
159	Andorra	Saint Lucia	Sao Tome
160	Marshall Islands	Peru	Indonesia
161	Tunisia	Serbia	Lithuania
162	Senegal	Saudi Arabia	Argentina
163	Afghanistan	Dominica	Belgium
164	Croatia	Puerto Rico	Palestine
165	Yemen	Haiti	Bangladesh
166	Bahamas, The	Guatemala	France
167	Georgia	Martinique	Serbia and Montenegro
168	Greece	Iraq	Iceland
169	Jersey	Honduras	San Marino
170	South Africa	El Salvador	Niger
171	Montenegro	Turkmenistan	Guadeloupe
172	Mauritania	Luxembourg	Iran
173	Tuvalu	Solomon Islands	Burma (Myanmar)
174	Bolivia	Croatia	Reunion
175	Malta	Brazil	Serbia
176	Germany	Monaco	Gambia, The
177	Mexico	Ireland	Mali
178	Saudi Arabia	Netherlands Antilles	Egypt
179	Turkey	Guadeloupe	Paraguay
180	Morocco	Antigua and Barbuda	Kosovo

Ranking of countries by religious fractionalization (High to Low)

Ranks	rfrac1	rfrac2	rfrac3
181	Somalia	Ukraine	Mexico
182	Finland	Nicaragua	Jordan
183	Rwanda	Tunisia	Colombia
184	Serbia and Montenegro	Greece	Ecuador
185	Gaza Strip	Colombia	Tajikistan
186	Poland	Venezuela	Croatia
187	Tajikistan	Iran	Ireland
188	Serbia	Comoros	Kyrgyzstan
189	Egypt	Paraguay	Saudi Arabia
190	Palestine	Samoa	Portugal
191	Bulgaria	Uruguay	Monaco
192	Syria	Spain	Andorra
193	Cyprus	Andorra	Luxembourg
194	Russia	Tajikistan	Senegal
195	Western Sahara	Tuvalu	Djibouti
196	Maldives	Italy	Thailand
197	Algeria	Armenia	Armenia
198	Uzbekistan	South Africa	Denmark
199	Italy	Belarus	Malta
200	Spain	Portugal	Venezuela
201	Comoros	Yemen	Cape Verde
202	Turkmenistan	Uzbekistan	Libya
203	Cape Verde	Marshall Islands	Mayotte
204	Luxembourg	Malta	Cambodia
205	Djibouti	Afghanistan	Tuvalu
206	Monaco	Mauritania	Turkmenistan
207	Kyrgyzstan	Romania	Finland
208	Kosovo	Bahamas, The	Poland
209	Reunion	Gaza Strip	Tunisia
210	Belgium	Jersey	Greece
211	Romania	Bolivia	Comoros
212	West Bank	Mexico	Spain
213	Kazakhstan	Algeria	Italy
214	Belarus	Turkey	Uzbekistan
215	Ukraine	Morocco	Mauritania
216	Yugoslavia	Somalia	Algeria
217	Bhutan	Finland	Gaza Strip
218	Albania	Poland	Turkey
219	Bosnia and Herzegovina	Western Sahara	Morocco
220	East Timor	Maldives	Somalia
221	Iraq	Cape Verde	Western Sahara
222	Latvia	Latvia	Maldives

Table B.19: Ranking of countries by religious polarization (High to Low)

Ranks	rpol1	rpol2	rpol3
1	Togo	Angola	French Guiana
2	Angola	Estonia	Yemen
3	Estonia	French Guiana	Isle of Man
4	French Guiana	Bosnia and Herzegovina	Guatemala
5	Mauritius	Eritrea	Korea, North
6	Korea, South	Korea, South	Japan
7	Tonga	Tonga	Micronesia
8	Guinea-Bissau	Isle of Man	Guinea-Bissau
9	Isle of Man	Korea, North	Switzerland
10	Taiwan	Japan	Russia
11	Japan	Russia	Tonga
12	Vietnam	Guinea-Bissau	Mongolia
13	Madagascar	Vietnam	Angola
14	Mongolia	Botswana	Eritrea
15	Central African Republic	Mongolia	Latvia
16	Tanzania	Central African Republic	Iraq
17	Botswana	Togo	Laos
18	Macau	Ethiopia	Cuba
19	Laos	Macau	Macedonia
20	Singapore	Laos	Kiribati
21	Cuba	Chad	Korea, South
22	Congo, Republic of the	Tanzania	French Polynesia
23	China	Cuba	Congo, Republic of the
24	Kenya	Macedonia	Sierra Leone
25	Guyana	Congo, Republic of the	Macau
26	Sierra Leone	Lebanon	East Timor
27	Lesotho	Cameroon	Bosnia and Herzegovina
28	Slovenia	Sierra Leone	Ethiopia
29	Suriname	Zimbabwe	Uruguay
30	Fiji	Mauritius	Germany
31	Czech Republic	Nigeria	Slovenia
32	American Samoa	Kenya	Vietnam
33	Trinidad and Tobago	East Timor	Estonia
34	Malaysia	Albania	Tanzania
35	Malawi	Slovenia	Virgin Islands
36	Benin	Guyana	Canada
37	Zimbabwe	Lesotho	Czech Republic
38	Cameroon	Fiji	Azerbaijan
39	Burundi	Cote d'Ivoire	Rwanda
40	Cote d'Ivoire	Madagascar	Northern Mariana Islands
41	Hong Kong	Suriname	China
42	New Zealand	Burkina Faso	Gabon
43	Brunei	China	Uganda
44	Nigeria	Taiwan	Nauru
45	Northern Mariana Islands	Czech Republic	Madagascar
46	Namibia	American Samoa	Jersey
47	Liberia	Trinidad and Tobago	Guernsey
48	Jamaica	Malawi	Burundi
49	CZECHOSLOVAKIA	Burundi	Togo
50	Sudan	Benin	El Salvador
51	Liechtenstein	Mozambique	Marshall Islands
52	Faroe Islands	Liberia	Nicaragua
53	Belize	Malaysia	Burkina Faso
54	Burkina Faso	Bhutan	Palau
55	Bermuda	Singapore	Taiwan
56	Nauru	Sri Lanka	Mauritius
57	India	Brunei	Chad
58	Sri Lanka	New Zealand	Bahrain
59	Zambia	Ghana	Central African Republic
60	Vanuatu	Hong Kong	Liberia

Ranking of countries by religious polarization (High to Low)

Ranks	rpol1	rpol2	rpol3
61	Saint Kitts and Nevis	Sudan	Cote d'Ivoire
62	Qatar	Yugoslavia	Cameroon
63	Palau	Kuwait	Zimbabwe
64	Kuwait	Israel	New Caledonia
65	Hungary	Northern Mariana Islands	Hungary
66	Guam	Gabon	Namibia
67	Sweden	Qatar	Albania
68	Dominican Republic	Namibia	Malaysia
69	Bangladesh	Montenegro	Bhutan
70	Niger	Kazakhstan	Lesotho
71	Moldova	Cyprus	Botswana
72	Swaziland	Congo, Democratic Republic of the	Yugoslavia
73	French Polynesia	Jamaica	Ukraine
74	Gabon	CZECHOSLOVAKIA	Congo, Democratic Republic of the
75	Guernsey	Moldova	Belarus
76	Congo, Democratic Republic of the	Liechtenstein	United Kingdom
77	Burma (Myanmar)	Zambia	Swaziland
78	Ghana	India	Netherlands
79	San Marino	Faroe Islands	Kuwait
80	Chad	West Bank	Sri Lanka
81	Mali	Belize	Brunei
82	Mozambique	Bahrain	Grenadine
83	Oman	Bermuda	Grenada
84	Norway	Nauru	Singapore
85	Panama	Nepal	Suriname
86	Virgin Islands	Sweden	Sudan
87	Guinea	Vanuatu	CZECHOSLOVAKIA
88	Bahrain	Saint Kitts and Nevis	Moldova
89	Korea, North	Guinea	Hong Kong
90	Grenadine	Uganda	Belize
91	Australia	Palau	Mozambique
92	Lithuania	Hungary	Guam
93	Barbados	Netherlands	Saint Kitts and Nevis
94	Thailand	Guam	Pakistan
95	Aruba	Bulgaria	Panama
96	Nepal	Indonesia	Nigeria
97	Costa Rica	Panama	Oman
98	Chile	Swaziland	Liechtenstein
99	New Caledonia	Dominican Republic	Bermuda
100	Equatorial Guinea	Oman	Dominica
101	Ecuador	Bangladesh	Israel
102	Austria	Niger	Kenya
103	United Arab Emirates	Palestine	Montenegro
104	Iceland	Reunion	American Samoa
105	Micronesia	Syria	Chile
106	Sao Tome	French Polynesia	Cyprus
107	Israel	Georgia	Fiji
108	Netherlands	Guernsey	Australia
109	Slovakia	Egypt	Zambia
110	Seychelles	Burma (Myanmar)	Saint Lucia
111	Peru	Mali	Puerto Rico
112	Cambodia	Gambia, The	Qatar
113	Papua New Guinea	Norway	Barbados
114	Canada	San Marino	Afghanistan
115	Ethiopia	Grenadine	United States
116	Indonesia	Kosovo	Lebanon
117	Lebanon	France	Brazil
118	Haiti	Austria	Bahamas, The
119	Pakistan	Australia	Sweden
120	Martinique	Equatorial Guinea	Costa Rica

Ranking of countries by religious polarization (High to Low)

Ranks	rpol1	rpol2	rpol3
121	Guatemala	Jordan	Bolivia
122	Honduras	New Caledonia	Samoa
123	El Salvador	Belgium	Syria
124	Argentina	Kyrgyzstan	Solomon Islands
125	Solomon Islands	Virgin Islands	United Arab Emirates
126	Brazil	Philippines	Benin
127	Ireland	Azerbaijan	Malawi
128	United States	Switzerland	Ghana
129	Guadeloupe	Lithuania	Vanuatu
130	Antigua and Barbuda	Barbados	Slovakia
131	Puerto Rico	Djibouti	Kazakhstan
132	Nicaragua	Germany	Dominican Republic
133	Gambia, The	Senegal	Haiti
134	Venezuela	United States	Faroe Islands
135	Eritrea	Aruba	West Bank
136	Iran	Canada	Guyana
137	Denmark	Thailand	Bulgaria
138	Jordan	Costa Rica	India
139	Paraguay	Chile	New Zealand
140	Samoa	Pakistan	Peru
141	Philippines	Denmark	Trinidad and Tobago
142	Azerbaijan	Rwanda	Philippines
143	Colombia	Seychelles	Nepal
144	Kiribati	Argentina	South Africa
145	Dominica	Serbia and Montenegro	Papua New Guinea
146	Uganda	Grenada	Georgia
147	Macedonia	United Kingdom	Guinea
148	Grenada	Ecuador	Seychelles
149	United Kingdom	United Arab Emirates	Netherlands Antilles
150	France	Iceland	Equatorial Guinea
151	Netherlands Antilles	Kiribati	Antigua and Barbuda
152	Uruguay	Micronesia	Aruba
153	Mayotte	Sao Tome	Austria
154	Armenia	Libya	Norway
155	Libya	Slovakia	Honduras
156	Switzerland	Papua New Guinea	Martinique
157	Saint Lucia	Mayotte	Romania
158	Portugal	Peru	Lithuania
159	Andorra	Serbia	Indonesia
160	Marshall Islands	Saint Lucia	Sao Tome
161	Afghanistan	Cambodia	Belgium
162	Senegal	Saudi Arabia	Bangladesh
163	Tunisia	Dominica	Argentina
164	Croatia	Haiti	Niger
165	Yemen	Puerto Rico	Palestine
166	Bahamas, The	Martinique	Reunion
167	Georgia	Iraq	San Marino
168	Greece	Guatemala	Jamaica
169	Jersey	Honduras	Iran
170	South Africa	El Salvador	Serbia and Montenegro
171	Montenegro	Turkmenistan	France
172	Tuvalu	Luxembourg	Iceland
173	Mauritania	Solomon Islands	Guadeloupe
174	Bolivia	Croatia	Egypt
175	Malta	Brazil	Mali
176	Germany	Monaco	Burma (Myanmar)
177	Mexico	Ireland	Serbia
178	Saudi Arabia	Netherlands Antilles	Gambia, The
179	Turkey	Guadeloupe	Paraguay
180	Morocco	Antigua and Barbuda	Kosovo

Ranking of countries by religious polarization (High to Low)

Ranks	rpol1	rpol2	rpol3
181	Somalia	Ukraine	Jordan
182	Finland	Nicaragua	Mexico
183	Rwanda	Venezuela	Tajikistan
184	Serbia and Montenegro	Iran	Kyrgyzstan
185	Poland	Comoros	Ecuador
186	Gaza Strip	Colombia	Colombia
187	Serbia	Greece	Croatia
188	Tajikistan	Tunisia	Ireland
189	Cyprus	Paraguay	Saudi Arabia
190	Russia	Samoa	Portugal
191	Bulgaria	Spain	Monaco
192	Egypt	Uruguay	Andorra
193	Palestine	Andorra	Djibouti
194	Syria	Tajikistan	Senegal
195	Kazakhstan	Tuvalu	Luxembourg
196	Belarus	Italy	Thailand
197	Iraq	Armenia	Armenia
198	Belgium	South Africa	Denmark
199	Monaco	Belarus	Malta
200	Albania	Portugal	Cape Verde
201	Reunion	Yemen	Venezuela
202	Cape Verde	Uzbekistan	Libya
203	Spain	Marshall Islands	Mayotte
204	Luxembourg	Malta	Cambodia
205	Djibouti	Afghanistan	Tuvalu
206	Italy	Romania	Turkmenistan
207	Algeria	Mauritania	Finland
208	Ukraine	Bahamas, The	Poland
209	Bosnia and Herzegovina	Gaza Strip	Comoros
210	Latvia	Jersey	Greece
211	Yugoslavia	Bolivia	Tunisia
212	East Timor	Mexico	Spain
213	Bhutan	Algeria	Italy
214	West Bank	Turkey	Uzbekistan
215	Romania	Morocco	Algeria
216	Kosovo	Somalia	Mauritania
217	Kyrgyzstan	Finland	Gaza Strip
218	Turkmenistan	Poland	Turkey
219	Comoros	Cape Verde	Morocco
220	Uzbekistan	Latvia	Somalia
221	Western Sahara	Western Sahara	Western Sahara
222	Maldives	Maldives	Maldives

Table B.20: Ranking of countries by religious intolerance (High to Low)

Rank	Country	Religious intolerance	Govt. Regulation	Social regulation	Govt. Favouritism
1	Saudi Arabia	3.2583	9.444	9.556	9.278
2	Iran	3.2449	8.796	10	9.389
3	Pakistan	3.1382	8.796	10	8.811
4	Burma (Myanmar)	2.8537	9.259	8.667	8.289
5	Afghanistan	2.8437	7.685	9.778	8.644
6	Egypt	2.7928	8.333	9.556	7.933
7	Iraq	2.6477	7.315	9.333	8.478
8	Uzbekistan	2.5316	8.982	7.778	7.844
9	Kuwait	2.4743	7.87	8.445	7.956
10	Maldives	2.4462	9.722	6	8.611
11	Armenia	2.4217	7.87	7.556	8.678
12	Algeria	2.3786	6.759	8.222	8.867
13	Jordan	2.3708	8.333	6.889	8.667
14	Sudan	2.3605	8.056	9.111	6.389
15	Indonesia	2.3583	6.667	9.556	7.344
16	Comoros	2.2839	8.796	8.445	5.944
17	Belarus	2.2443	7.963	7.778	7.367
18	Georgia	2.2312	7.037	8.445	7.522
19	Bhutan	2.2106	8.056	6.667	8.344
20	Bahrain	2.1737	7.5	6.667	8.733
21	Malaysia	2.1726	7.593	7.556	7.622
22	Qatar	2.1101	8.796	5.111	8.778
23	India	2.1059	6.296	10	5.867
24	Romania	2.0645	6.296	8.222	7.656
25	Greece	2.0601	6.759	7.111	8.4
26	Israel	2.0222	4.815	9.111	7.989
27	Turkmenistan	2.0127	8.982	4.667	8.556
28	Palestine	1.9677	4.352	9.333	7.933
29	Bangladesh	1.9431	7.13	7.333	7.122
30	Mauritania	1.9167	7.778	5.334	8.556
31	Brunei	1.9149	9.445	5.778	6.278
32	Nigeria	1.9040	6.852	7.111	7.456
33	Turkey	1.8874	5.185	9.111	6.867
34	China	1.8830	8.796	5.556	7.044
35	Morocco	1.8495	6.482	7.334	7.3
36	Azerbaijan	1.8170	8.056	8.444	4.2
37	Tunisia	1.7436	5.926	6.445	8.322
38	Russia	1.7297	6.482	7.556	6.4
39	Nepal	1.7056	6.389	9.333	4.356
40	Sri Lanka	1.6337	5.556	9.111	5.1
41	United Arab Emirates	1.6217	6.389	5.556	8.178
42	Cyprus	1.5932	4.63	7.556	7.622
43	Oman	1.5858	6.759	6	7.089
44	Yemen	1.5792	5.926	7.778	5.922
45	Chad	1.5471	6.574	5.556	7.578
46	Somalia	1.5221	7.222	8	3.989
47	Yugoslavia	1.4264	6.111	5.333	7.667
48	Bulgaria	1.2264	7.5	4	6.622
49	Lebanon	1.1631	5.741	6.667	5.122
50	Bosnia and Herzegovina	1.1402	5.833	5.333	6.411
51	Libya	1.1284	6.667	4.222	6.722
52	Syria	1.0031	5.741	6	5.011
53	Ethiopia	0.9920	4.167	6.889	5.611
54	Colombia	0.9720	4.167	4.889	7.767
55	Laos	0.9150	8.889	3.556	3.967
56	Kazakhstan	0.7505	6.574	5.111	3.767
57	Moldova	0.6914	4.445	3.778	7.211
58	Kosovo	0.6853	3.056	6	6.133
59	Singapore	0.6560	7.87	1.778	5.656
60	Eritrea	0.6557	8.148	4.667	2.089
61	Kyrgyzstan	0.6111	6.019	6.445	2.089
62	Vietnam	0.6083	8.241	4	2.489
63	Djibouti	0.5956	5.833	4.445	4.467
64	Ukraine	0.5778	4.722	4.889	5.044
65	Cote d'Ivoire	0.5504	4.259	3.778	6.644

Ranking of countries by religious religious intolerance (High to Low)

Rank	Country	Religious intolerance	Govt. Regulation	Social regulation	Govt. Favouritism
66	Thailand	0.5261	4.815	2.445	7.433
67	Belgium	0.5068	3.148	4	7.333
68	Cuba	0.4870	7.222	4.667	2.156
69	France	0.4650	4.445	4.667	4.978
70	Macedonia	0.4453	5.278	4.667	3.989
71	Argentina	0.4003	2.037	4	7.933
72	Guinea	0.3986	2.315	4.222	7.378
73	Vanuatu	0.3920	1.667	5.111	7.022
74	Serbia and Montenegro	0.3913	2.222	4.667	6.933
75	Germany	0.3905	3.333	4.222	6.256
76	Kenya	0.3398	3.333	4	6.233
77	Austria	0.3385	1.945	4.889	6.689
78	Tajikistan	0.3230	5	5.333	2.867
79	Mongolia	0.3207	5.556	2.889	5.033
80	Croatia	0.2610	1.667	3.778	7.822
81	Tanzania	0.2126	5.185	3.111	4.589
82	Lithuania	0.1507	3.889	2.222	6.633
83	Western Sahara	0.1146	5.278	2.889	4.211
84	Spain	0.1023	1.019	3.556	7.9
85	Italy	0.0904	1.204	4.445	6.633
86	Latvia	0.0879	3.889	2	6.544
87	Mexico	0.0846	3.333	5.556	3.089
88	Liberia	0.0330	3.056	3.556	5.367
89	Korea, North	0.0215	8.889	2	0.889
90	Zimbabwe	-0.0040	3.056	2.889	5.922
91	Peru	-0.0086	2.778	0.889	8.456
92	Nicaragua	-0.0144	0.741	3.778	7.311
93	Venezuela	-0.0527	1.204	2.445	8.122
94	Norway	-0.1221	1.759	2.889	6.656
95	Hungary	-0.1424	1.389	2.667	7.189
96	Central African Republic	-0.1571	5.278	3.778	1.733
97	Niger	-0.1581	2.5	4.222	4.167
98	Guatemala	-0.1633	1.204	4.222	5.511
99	Nauru	-0.1732	5.185	3.778	1.744
100	Slovakia	-0.1761	1.204	3.111	6.7
101	Iceland	-0.2247	0.926	2	7.989
102	Uganda	-0.2403	3.889	5.111	1.244
103	Cameroon	-0.2864	2.778	4.889	2.422
104	Equatorial Guinea	-0.3115	4.259	2	3.989
105	Philippines	-0.3207	1.759	4.667	3.567
106	Dominican Republic	-0.3207	1.574	1.111	7.789
107	Cambodia	-0.3227	2.315	0.222	8
108	Monaco	-0.3418	5.185	2.667	2.089
109	Switzerland	-0.3469	1.019	2.889	6.222
110	Czech Republic	-0.3689	0.185	2.445	7.489
111	Finland	-0.3749	1.574	2	6.489
112	Denmark	-0.3899	1.759	1.333	6.967
113	Slovenia	-0.4111	0.926	4.445	4.211
114	Chile	-0.4477	2.222	1.556	5.911
115	East Timor	-0.4888	1.667	5.333	2
116	United Kingdom	-0.4920	1.204	3.111	4.989
117	Costa Rica	-0.5578	1.019	0.889	7.344
118	Ghana	-0.5688	2.037	1.778	5.2
119	Poland	-0.5769	0	3.556	5.3
120	Netherlands	-0.5819	0	3.778	5.022
121	Portugal	-0.6084	1.574	0	7.489
122	Mauritius	-0.6351	0.556	3.111	4.9
123	Panama	-0.6529	1.296	1.333	6.033
124	Bolivia	-0.6700	0	0.667	8.067
125	Japan	-0.6704	2.315	2.445	3.6
126	Malta	-0.7644	0	0	8.311
127	Andorra	-0.8143	0.741	0	7.256
128	Haiti	-0.8287	0.278	2.445	4.9
129	Canada	-0.8677	0.278	1.778	5.444
130	Luxembourg	-0.8800	0.185	0	7.489

Ranking of countries by religious religious intolerance (High to Low)

Rank	Country	Religious intolerance	Govt. Regulation	Social regulation	Govt. Favouritism
131	Hong Kong	-0.9017	1.019	0.667	5.733
132	Saint Lucia	-0.9167	1.019	4	1.878
133	Rwanda	-0.9295	4.074	1.111	1.844
134	Malawi	-0.9339	0	2.889	4.122
135	Congo, Democratic Republic	-0.9706	3.704	2.889	0
136	Solomon Islands	-0.9707	0.556	2.222	4.089
137	San Marino	-0.9776	0	0	7.156
138	Fiji	-1.0125	0.741	3.556	2.156
139	Swaziland	-1.0228	2.963	0.222	3.522
140	Madagascar	-1.0508	1.204	0.889	4.478
141	Liechtenstein	-1.0595	0.463	0	6.222
142	Tuvalu	-1.0987	0.741	3.556	1.689
143	Papua New Guinea	-1.1082	0	2.222	3.933
144	Palau	-1.1830	0.648	1.778	3.344
145	Honduras	-1.2031	1.296	0.222	4.311
146	South Africa	-1.2341	0	3.778	1.489
147	Korea, South	-1.2461	0.463	0.667	4.456
148	Cape Verde	-1.2930	0	0.445	4.944
149	Senegal	-1.3079	0	0	5.367
150	Gabon	-1.3127	1.759	0	3.478
151	Brazil	-1.3158	0.833	3.334	0.667
152	Belize	-1.3643	0.278	0	4.767
153	Suriname	-1.3921	0	0.667	4.156
154	Zambia	-1.4113	0.185	0	4.611
155	Sweden	-1.4194	0.278	1.111	3.211
156	Trinidad and Tobago	-1.4277	0.833	1.111	2.578
157	Seychelles	-1.4434	0	0	4.633
158	Jamaica	-1.4979	1.482	2.445	0
159	Gambia, The	-1.5163	0.278	0	3.944
160	Tonga	-1.5313	1.019	0	3.078
161	Samoa	-1.5458	0.37	2.667	0.667
162	Bahamas, The	-1.5595	0.463	0.444	3.011
163	Lesotho	-1.5787	0	0	3.9
164	Estonia	-1.6004	0.278	0.667	2.733
165	El Salvador	-1.6384	1.111	0	2.4
166	Grenadine	-1.6407	0	2	1.3
167	Angola	-1.7055	0.741	1.556	0.667
168	Mozambique	-1.7153	1.111	1.556	0.222
169	Sierra Leone	-1.7386	0.278	1.556	0.978
170	Australia	-1.7492	0.463	2	0.222
171	Paraguay	-1.7596	0	1.333	1.411
172	Albania	-1.7626	0.463	1.111	1.156
173	New Zealand	-1.7703	0	0.222	2.611
174	Dominica	-1.8285	0.741	0.222	1.511
175	Guyana	-1.8294	0.278	0.445	1.744
176	Ireland	-1.8626	0	0.222	2.111
177	Barbados	-1.8729	0	0.889	1.3
178	Macau	-1.8758	1.296	0.222	0.667
179	Mali	-1.8861	0	1.778	0.222
180	Congo, Republic of the	-1.9271	0	1.778	0
181	Saint Kitts and Nevis	-1.9846	0	1.111	0.444
182	Guinea-Bissau	-2.0042	0.556	0.889	0
183	Ecuador	-2.0115	0	0	1.556
184	Togo	-2.0259	0.556	0	0.889
185	Taiwan	-2.0639	0.278	0	0.978
186	Burundi	-2.0745	0.463	0.444	0.222
187	Burkina Faso	-2.1129	0	0.889	0
188	Uruguay	-2.1129	0	0.889	0
189	Botswana	-2.1437	0.556	0.222	0
190	Benin	-2.1551	0	0	0.778
191	Antigua and Barbuda	-2.1629	0.695	0	0
192	Namibia	-2.2578	0	0	0.222
193	Sao Tome	-2.2578	0	0	0.222
194	Grenada	-2.2988	0	0	0
195	Kiribati	-2.2988	0	0	0
196	Marshall Islands	-2.2988	0	0	0
197	Micronesia	-2.2988	0	0	0

Appendix C

Appendix for Chapter3

C.1 Figures

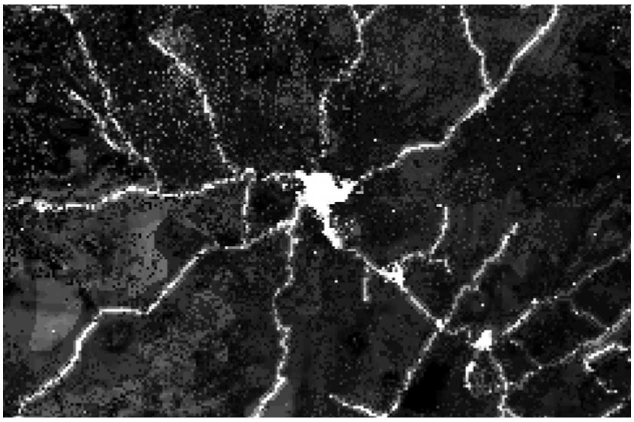

FIGURE C.1: Addis Abbaba Population

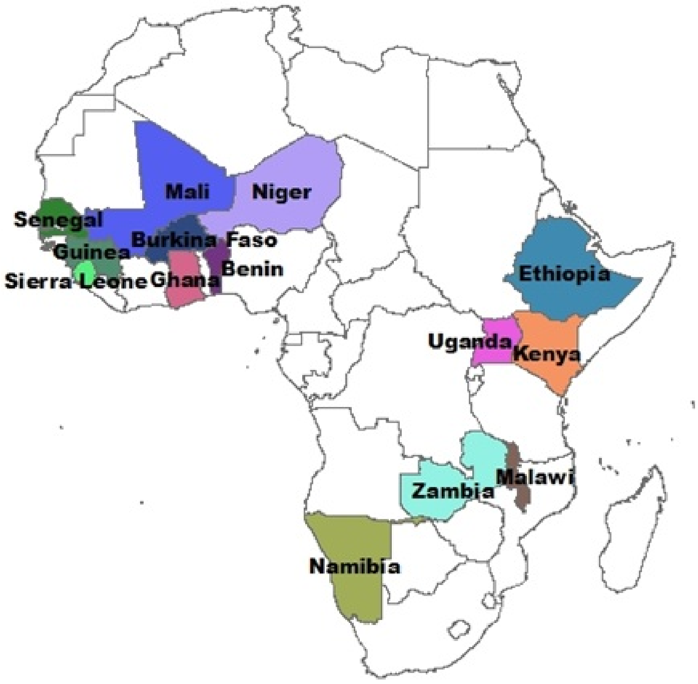

Figure C.2: Countries Used

C.2 Tables

TABLE C.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
child death	0.229	0.42	0	1	658755
infant death	0.12	0.325	0	1	821918
linguistic distance 25	0.121	0.266	0	1	868629
linguistic distance 50	0.125	0.265	0	1	868629
linguistic distance 75	0.128	0.266	0	1	868629
linguistic distance 100	0.131	0.267	0	1	868629
linguistic distance 125	0.133	0.27	0	1	868629
linguistic distance 150	0.135	0.271	0	1	868629
linguistic distance 175	0.138	0.272	0	1	868629
linguistic distance 200	0.14	0.274	0	1	868629
linguistic distance 250	0.145	0.276	0	1	868629
fractionalization district	0.399	0.289	0	0.997	868629
fractionalization 25	0.388	0.298	0	0.905	868761
fractionalization 50	0.467	0.284	0	0.921	868761
fractionalization 75	0.52	0.268	0	0.935	868761
fractionalization 100	0.556	0.252	0	0.929	868761
fractionalization 125	0.584	0.238	0	0.933	868761
fractionalization 150	0.608	0.225	0	0.935	868761
fractionalization 175	0.629	0.21	0	0.933	868761
fractionalization 200	0.648	0.195	0	0.933	868761
fractionalization 250	0.676	0.172	0.093	0.930	868761
polarization district	0.464	0.299	0	1	868629
urban	0.225	0.417	0	1	868629
female	0.49	0.5	0	1	868629
age at birth	24.991	6.422	8	50	868629
age at birth sqaured	665.786	348.351	64	2500	868629
multiple birth	0.032	0.177	0	1	868629
birth order	3.442	2.316	1	18	868629
birth order squared	17.216	22.533	1	324	868629
short birth space prior	0.209	0.407	0	1	868629
short birth space post	0.209	0.406	0	1	868629
education years	2.013	3.415	0	26	868306
wealth index	2.871	1.401	1	5	868629
birth year	1992.261	9.364	1955	2011	868629
log of distance to capital	5.123	1.217	-2.614	7.188	868629
log of geographic distance	5.656	0.39	4.662	7.028	868629

TABLE C.2: Summary statistics Child mortality

Variable	Mean	Std. Dev.	Obs
Benin	0.246	0.43	28,184
Burkina Faso	0.241	0.428	93,972
Ethiopia	0.215	0.411	95,686
Ghana	0.162	0.369	35,612
Guinea	0.265	0.442	37,924
Kenya	0.130	0.337	31,867
Malawi	0.228	0.420	103,901
Mali	0.310	0.462	96,365
Namibia	0.089	0.285	10,646
Niger	0.358	0.479	21,879
Senegal	0.168	0.374	54,651
Sierra Leone	0.226	0.419	13,780
Uganda	0.165	0.371	18,253
Zambia	0.185	0.389	15,038
All	0.229	0.420	657758

TABLE C.3: Child mortality: Baseline-last 10 years

	(1)	(2)	(3)	(4)	(5)	(6)
linguistic distance 75	0.0599	0.0598	0.0992*	0.0649**	0.0770**	0.0779**
	(0.0621)	(0.0653)	(0.0571)	(0.0267)	(0.0375)	(0.0392)
fractionalization 75		0.000408	-0.0608	-0.0101	-0.0243	-0.0233
		(0.0871)	(0.0744)	(0.0207)	(0.0209)	(0.0244)
urban			-0.0642*	-0.0451*	-0.0481**	-0.0474*
			(0.0366)	(0.0242)	(0.0243)	(0.0246)
female			-0.0675***	-0.0697***	-0.0701***	-0.0701***
			(0.00709)	(0.00655)	(0.00660)	(0.00660)
education_years			-0.0292***	-0.0179***	-0.0180***	-0.0180***
			(0.00569)	(0.00282)	(0.00279)	(0.00280)
wealth index 2			0.0587**	0.0321	0.0255	0.0253
			(0.0229)	(0.0204)	(0.0195)	(0.0195)
wealth index 3			0.0505**	0.0218	0.0128	0.0131
			(0.0246)	(0.0301)	(0.0284)	(0.0283)
wealth index 4			0.0229	-0.0142	-0.0219	-0.0218
			(0.0291)	(0.0329)	(0.0314)	(0.0313)
wealth index 5			-0.0459	-0.0941***	-0.106***	-0.105***
			(0.0423)	(0.0354)	(0.0326)	(0.0317)
lndist2cap						0.00699
						(0.0164)
ln_geog_dist						-0.0895*
						(0.0507)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	273822	273822	273706	264046	263730	263730
pseudo R^2	0.000	0.000	0.059	0.089	0.091	0.091

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table C.4: Child mortality: Alternative radii, last 10 years

	(1)	(2)	(3)	(4)	(5)	(6)
linguistic distance 25	0.0505	0.0473	0.0921	0.0573**	0.0483	0.0489
	(0.0610)	(0.0656)	(0.0584)	(0.0254)	(0.0310)	(0.0320)
fractionalization 25		0.0198	-0.0386	-0.0118	-0.0235	-0.0213
		(0.0608)	(0.0503)	(0.0196)	(0.0190)	(0.0201)
linguistic distance 50	0.0566	0.0545	0.0973*	0.0612**	0.0611*	0.0618*
	(0.0627)	(0.0666)	(0.0587)	(0.0255)	(0.0345)	(0.0355)
fractionalization 50		0.0184	-0.0459	0.00163	-0.00916	-0.00743
		(0.0739)	(0.0627)	(0.0193)	(0.0199)	(0.0218)
linguistic distance 75	0.0599	0.0598	0.0992*	0.0649**	0.0770**	0.0779**
	(0.0621)	(0.0653)	(0.0571)	(0.0267)	(0.0375)	(0.0392)
fractionalization 75		0.000408	-0.0608	-0.0101	-0.0243	-0.0233
		(0.0871)	(0.0744)	(0.0207)	(0.0209)	(0.0244)
linguistic distance 100	0.0607	0.0615	0.0970*	0.0624**	0.0719**	0.0729*
	(0.0615)	(0.0641)	(0.0556)	(0.0265)	(0.0362)	(0.0378)
fractionalization 100		-0.0113	-0.0733	-0.0213	-0.0384*	-0.0383
		(0.102)	(0.0859)	(0.0226)	(0.0200)	(0.0233)
linguistic distance 125	0.0606	0.0618	0.0941*	0.0598**	0.0578	0.0581
	(0.0606)	(0.0628)	(0.0541)	(0.0261)	(0.0446)	(0.0463)
fractionalization 125		-0.0252	-0.0821	-0.0117	-0.0272	-0.0285
		(0.118)	(0.0997)	(0.0262)	(0.0243)	(0.0296)
linguistic distance 150	0.0588	0.0599	0.0909*	0.0568**	0.0410	0.0407
	(0.0600)	(0.0620)	(0.0530)	(0.0260)	(0.0604)	(0.0620)
fractionalization 150		-0.0289	-0.0857	0.0216	0.00721	0.00540
		(0.139)	(0.116)	(0.0315)	(0.0312)	(0.0370)
linguistic distance 175	0.0575	0.0589	0.0890*	0.0555**	0.0460	0.0445
	(0.0596)	(0.0613)	(0.0520)	(0.0257)	(0.0779)	(0.0799)
fractionalization 175		-0.0378	-0.0932	0.0462	0.0322	0.0304
		(0.163)	(0.135)	(0.0375)	(0.0408)	(0.0458)
linguistic distance 200	0.0570	0.0584	0.0877*	0.0537**	0.0450	0.0413
	(0.0596)	(0.0611)	(0.0515)	(0.0251)	(0.0940)	(0.0964)
fractionalization 200		-0.0456	-0.105	0.0820*	0.0715	0.0706
		(0.190)	(0.157)	(0.0469)	(0.0492)	(0.0535)
linguistic distance 250	0.0567	0.0573	0.0846*	0.0561**	0.0763	0.0693
	(0.0609)	(0.0606)	(0.0507)	(0.0246)	(0.119)	(0.122)
fractionalization 250		-0.0807	-0.163	0.0551	0.0397	0.0404
		(0.244)	(0.198)	(0.0769)	(0.0772)	(0.0755)
Individual controls	No	No	Yes	Yes	Yes	Yes
region	No	No	No	Yes	Yes	Yes
country*time	No	No	No	Yes	Yes	Yes
Religion	No	No	No	No	Yes	Yes
Ethnicity	No	No	No	No	Yes	Yes
N	273822	273822	273706	264046	263730	263730

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table C.5: Child mortality: Polarization

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5
ling_dist_25	0.0359	0.0756	0.0279	0.0622***	0.0620***
	(0.0559)	(0.0466)	(0.0218)	(0.0213)	(0.0210)
pol_district	0.0625	-0.00121	-0.0253	-0.0360*	-0.0360*
	(0.0511)	(0.0429)	(0.0213)	(0.0188)	(0.0189)
ling_dist_50	0.0399	0.0793*	0.0319	0.0853***	0.0850***
	(0.0576)	(0.0475)	(0.0231)	(0.0257)	(0.0253)
pol_district	0.0623	-0.00140	-0.0254	-0.0364*	-0.0364*
	(0.0511)	(0.0428)	(0.0213)	(0.0188)	(0.0189)
ling_dist_75	0.0428	0.0808*	0.0331	0.0977***	0.0974***
	(0.0580)	(0.0474)	(0.0238)	(0.0318)	(0.0317)
pol_district	0.0621	-0.00126	-0.0252	-0.0361*	-0.0361*
	(0.0510)	(0.0427)	(0.0214)	(0.0189)	(0.0190)
ling_dist_100	0.0449	0.0800*	0.0309	0.0970***	0.0968***
	(0.0582)	(0.0473)	(0.0239)	(0.0336)	(0.0335)
pol_district	0.0621	-0.000898	-0.0249	-0.0356*	-0.0356*
	(0.0509)	(0.0426)	(0.0214)	(0.0190)	(0.0190)
ling_dist_125	0.0458	0.0785*	0.0297	0.0976**	0.0971**
	(0.0577)	(0.0468)	(0.0237)	(0.0422)	(0.0422)
pol_district	0.0621	-0.000610	-0.0248	-0.0353*	-0.0354*
	(0.0508)	(0.0426)	(0.0213)	(0.0190)	(0.0190)
ling_dist_150	0.0444	0.0759	0.0270	0.0842	0.0837
	(0.0579)	(0.0469)	(0.0237)	(0.0533)	(0.0534)
pol_district	0.0623	-0.000275	-0.0247	-0.0350*	-0.0350*
	(0.0508)	(0.0425)	(0.0213)	(0.0190)	(0.0191)
ling_dist_175	0.0431	0.0740	0.0254	0.0806	0.0801
	(0.0581)	(0.0469)	(0.0235)	(0.0654)	(0.0656)
pol_district	0.0625	0.0000177	-0.0246	-0.0347*	-0.0348*
	(0.0507)	(0.0425)	(0.0213)	(0.0190)	(0.0191)
ling_dist_200	0.0427	0.0728	0.0245	0.0778	0.0767
	(0.0586)	(0.0474)	(0.0233)	(0.0764)	(0.0767)
pol_district	0.0626	0.000249	-0.0245	-0.0345*	-0.0345*
	(0.0507)	(0.0424)	(0.0213)	(0.0189)	(0.0190)
ling_dist_250	0.0419	0.0700	0.0240	0.0553	0.0535
	(0.0606)	(0.0494)	(0.0233)	(0.0884)	(0.0891)
pol_district	0.0629	0.000880	-0.0244	-0.0339*	-0.0339*
	(0.0506)	(0.0423)	(0.0213)	(0.0188)	(0.0189)
N	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and polarization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table C.6: Child mortality: Fractionalization & Polarization

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5
frac_district	-0.174*	-0.156**	-0.0136	-0.0414	-0.0426*
	(0.0950)	(0.0784)	(0.0283)	(0.0255)	(0.0253)
pol_district	0.203**	0.130	-0.0137	-0.00328	-0.00258
	(0.102)	(0.0797)	(0.0252)	(0.0229)	(0.0228)
urban		-0.0904***	-0.0691***	-0.0715***	-0.0688***
		(0.0325)	(0.0154)	(0.0148)	(0.0148)
female		-0.0655***	-0.0674***	-0.0679***	-0.0678***
		(0.00433)	(0.00433)	(0.00433)	(0.00433)
education_years		-0.0351***	-0.0221***	-0.0212***	-0.0211***
		(0.00514)	(0.00254)	(0.00248)	(0.00247)
_Iwealth_in_2		0.0295	0.0104	0.00763	0.00763
		(0.0188)	(0.0145)	(0.0141)	(0.0141)
_Iwealth_in_3		0.0125	-0.00602	-0.0102	-0.00990
		(0.0211)	(0.0220)	(0.0212)	(0.0212)
_Iwealth_in_4		-0.0238	-0.0554**	-0.0588**	-0.0584**
		(0.0252)	(0.0261)	(0.0249)	(0.0249)
_Iwealth_in_5		-0.101***	-0.153***	-0.155***	-0.152***
		(0.0376)	(0.0274)	(0.0255)	(0.0251)
lndist2cap					0.0158**
					(0.00758)
ln_geog_dist					-0.0517
					(0.0413)
N	658755	658505	649182	648468	648468
pseudo R^2	0.001	0.062	0.084	0.086	0.086

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level.

Table C.7: Child mortality: Fractionalization & Polarization with distance

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5
ling_dist_25	0.0503	0.0894*	0.0294	0.0655***	0.0654***
	(0.0576)	(0.0477)	(0.0223)	(0.0216)	(0.0213)
frac_district	-0.184*	-0.175**	-0.0191	-0.0449*	-0.0460*
	(0.0973)	(0.0796)	(0.0276)	(0.0254)	(0.0253)
pol_district	0.206**	0.135*	-0.0113	-0.00343	-0.00272
	(0.102)	(0.0790)	(0.0249)	(0.0222)	(0.0222)
ling_dist_50	0.0540	0.0928*	0.0334	0.0893***	0.0889***
	(0.0592)	(0.0485)	(0.0238)	(0.0265)	(0.0262)
frac_district	-0.185*	-0.175**	-0.0194	-0.0455*	-0.0466*
	(0.0970)	(0.0794)	(0.0277)	(0.0254)	(0.0252)
pol_district	0.206**	0.134*	-0.0111	-0.00335	-0.00264
	(0.102)	(0.0789)	(0.0249)	(0.0222)	(0.0221)
ling_dist_75	0.0562	0.0936*	0.0345	0.102***	0.102***
	(0.0593)	(0.0483)	(0.0244)	(0.0328)	(0.0327)
frac_district	-0.185*	-0.174**	-0.0191	-0.0455*	-0.0466*
	(0.0966)	(0.0791)	(0.0276)	(0.0253)	(0.0252)
pol_district	0.206**	0.134*	-0.0111	-0.00302	-0.00231
	(0.102)	(0.0789)	(0.0249)	(0.0223)	(0.0223)
ling_dist_100	0.0578	0.0922*	0.0321	0.102***	0.102***
	(0.0592)	(0.0479)	(0.0245)	(0.0350)	(0.0350)
frac_district	-0.185*	-0.173**	-0.0183	-0.0452*	-0.0463*
	(0.0964)	(0.0789)	(0.0276)	(0.0255)	(0.0253)
pol_district	0.206**	0.134*	-0.0114	-0.00277	-0.00207
	(0.102)	(0.0789)	(0.0249)	(0.0225)	(0.0224)
ling_dist_125	0.0581	0.0900*	0.0308	0.102**	0.102**
	(0.0584)	(0.0471)	(0.0241)	(0.0435)	(0.0434)
frac_district	-0.184*	-0.172**	-0.0178	-0.0446*	-0.0458*
	(0.0961)	(0.0787)	(0.0274)	(0.0255)	(0.0253)
pol_district	0.206**	0.134*	-0.0117	-0.00287	-0.00218
	(0.102)	(0.0789)	(0.0249)	(0.0226)	(0.0225)
ling_dist_150	0.0560	0.0868*	0.0280	0.0889	0.0885
	(0.0584)	(0.0470)	(0.0240)	(0.0544)	(0.0545)
frac_district	-0.184*	-0.171**	-0.0172	-0.0438*	-0.0450*
	(0.0959)	(0.0784)	(0.0274)	(0.0256)	(0.0254)
pol_district	0.205**	0.133*	-0.0121	-0.00312	-0.00243
	(0.102)	(0.0789)	(0.0250)	(0.0227)	(0.0226)
ling_dist_175	0.0542	0.0843*	0.0263	0.0848	0.0845
	(0.0585)	(0.0469)	(0.0239)	(0.0661)	(0.0663)
frac_district	-0.183*	-0.170**	-0.0167	-0.0432*	-0.0443*
	(0.0957)	(0.0783)	(0.0273)	(0.0256)	(0.0254)
pol_district	0.205**	0.133*	-0.0123	-0.00329	-0.00261
	(0.102)	(0.0790)	(0.0250)	(0.0227)	(0.0227)
ling_dist_200	0.0533	0.0827*	0.0253	0.0811	0.0802
	(0.0588)	(0.0472)	(0.0236)	(0.0767)	(0.0770)
frac_district	-0.183*	-0.170**	-0.0164	-0.0426*	-0.0437*
	(0.0956)	(0.0781)	(0.0273)	(0.0255)	(0.0253)
pol_district	0.205**	0.133*	-0.0124	-0.00346	-0.00278
	(0.102)	(0.0789)	(0.0250)	(0.0228)	(0.0227)
ling_dist_250	0.0519	0.0793	0.0247	0.0565	0.0549
	(0.0605)	(0.0490)	(0.0235)	(0.0878)	(0.0884)
frac_district	-0.182*	-0.169**	-0.0161	-0.0417	-0.0428*
	(0.0954)	(0.0779)	(0.0272)	(0.0255)	(0.0253)
pol_district	0.205**	0.133*	-0.0125	-0.00350	-0.00281
	(0.102)	(0.0789)	(0.0250)	(0.0228)	(0.0228)
N	658755	658505	649182	648468	648468

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table C.8: Child mortality: Non-linearities (Full Sample)

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5
linguistic distance 25	0.0860	0.205*		0.0648***	0.217**
	(0.0896)	(0.106)		(0.0219)	(0.104)
linguistic distance 25_2	-0.0326	-0.150			-0.163
	(0.0931)	(0.108)			(0.105)
fractionalization 25		-0.0451***	0.0319	0.0276	0.0219
		(0.0166)	(0.0537)	(0.0530)	(0.0533)
fractionalization 25_2			-0.0797	-0.0798	-0.0886
			(0.0666)	(0.0653)	(0.0626)
linguistic distance 50	0.175*	0.272**		0.0876***	0.291**
	(0.105)	(0.126)		(0.0280)	(0.127)
linguistic distance 50_2	-0.106	-0.199			-0.219*
	(0.110)	(0.128)			(0.128)
fractionalization 50		-0.0454**	0.106	0.102	0.0945
		(0.0199)	(0.0665)	(0.0653)	(0.0644)
fractionalization 50_2			-0.164**	-0.166**	-0.174**
			(0.0762)	(0.0747)	(0.0744)
linguistic distance 75	0.118	0.208		0.106***	0.221*
	(0.114)	(0.133)		(0.0350)	(0.133)
linguistic distance 75_2	-0.0292	-0.113			-0.124
	(0.128)	(0.141)			(0.140)
fractionalization 75		-0.0505**	0.139*	0.134*	0.129*
		(0.0231)	(0.0723)	(0.0709)	(0.0717)
fractionalization 75_2			-0.206**	-0.211**	-0.213***
			(0.0835)	(0.0819)	(0.0817)
linguistic distance 100	0.107	0.202		0.106***	0.207
	(0.111)	(0.129)		(0.0361)	(0.130)
linguistic distance 100_2	-0.0179	-0.104			-0.108
	(0.123)	(0.137)			(0.137)
fractionalization 100		-0.0580***	0.0800	0.0737	0.0689
		(0.0208)	(0.0831)	(0.0825)	(0.0832)
fractionalization 100_2			-0.137	-0.140	-0.141
			(0.0934)	(0.0929)	(0.0925)
linguistic distance 125	-0.00560	0.0817		0.104**	0.0816
	(0.125)	(0.148)		(0.0433)	(0.149)
linguistic distance 125_2	0.103	0.0238			0.0237
	(0.128)	(0.147)			(0.147)
fractionalization 125		-0.0512**	-0.00564	-0.0164	-0.0151
		(0.0217)	(0.0845)	(0.0844)	(0.0856)
fractionalization 125_2			-0.0398	-0.0375	-0.0375
			(0.0905)	(0.0902)	(0.0901)
linguistic distance 150	-0.116	-0.0707		0.0861	-0.0693
	(0.167)	(0.184)		(0.0535)	(0.183)
linguistic distance 150_2	0.204	0.163			0.162
	(0.161)	(0.175)			(0.174)
fractionalization 150		-0.0265	-0.0712	-0.0813	-0.0697
		(0.0243)	(0.0761)	(0.0761)	(0.0765)
fractionalization 150_2			0.0411	0.0440	0.0425
			(0.0830)	(0.0827)	(0.0820)
N	648468	648468	648468	648468	648468
pseudo R^2	0.086	0.086	0.086	0.086	0.086

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level.

Table C.9: Child mortality:Non-linearities (Last 10 years)

	(1)	(2)	(3)	(4)	(5)
	est1	est2	est3	est4	est5
linguistic distance 25	0.126	0.220*		0.0492	0.241*
	(0.114)	(0.131)		(0.0321)	(0.132)
linguistic distance 25_2	-0.0904	-0.183			-0.205*
	(0.111)	(0.123)			(0.124)
fractionalization 25		-0.0353*	0.0854	0.0824	0.0757
		(0.0212)	(0.0783)	(0.0776)	(0.0759)
fractionalization 25_2			-0.135	-0.135	-0.147
			(0.103)	(0.103)	(0.102)
linguistic distance 50	0.215*	0.262**		0.0636*	0.286**
	(0.124)	(0.131)		(0.0365)	(0.133)
linguistic distance 50_2	-0.170	-0.215*			-0.240*
	(0.122)	(0.124)			(0.126)
fractionalization 50		-0.0215	0.163	0.160	0.152
		(0.0214)	(0.106)	(0.106)	(0.105)
fractionalization 50_2			-0.205	-0.207*	-0.216*
			(0.126)	(0.126)	(0.125)
linguistic distance 75	0.127	0.181		0.0811**	0.195
	(0.148)	(0.162)		(0.0399)	(0.162)
linguistic distance 75_2	-0.0611	-0.111			-0.123
	(0.152)	(0.163)			(0.162)
fractionalization 75		-0.0294	0.180	0.177	0.172
		(0.0258)	(0.125)	(0.124)	(0.124)
fractionalization 75_2			-0.232*	-0.236*	-0.238*
			(0.139)	(0.138)	(0.138)
linguistic distance 100	0.0861	0.158		0.0745**	0.164
	(0.141)	(0.150)		(0.0377)	(0.151)
linguistic distance 100_2	-0.0262	-0.0912			-0.0957
	(0.143)	(0.152)			(0.153)
fractionalization 100		-0.0431*	0.134	0.130	0.126
		(0.0247)	(0.137)	(0.136)	(0.137)
fractionalization 100_2			-0.184	-0.186	-0.187
			(0.150)	(0.150)	(0.150)
linguistic distance 125	0.0113	0.0610		0.0577	0.0607
	(0.133)	(0.142)		(0.0461)	(0.143)
linguistic distance 125_2	0.0420	-0.00303			-0.00313
	(0.136)	(0.146)			(0.147)
fractionalization 125		-0.0286	0.0394	0.0336	0.0334
		(0.0312)	(0.128)	(0.129)	(0.130)
fractionalization 125_2			-0.0653	-0.0642	-0.0642
			(0.144)	(0.144)	(0.144)
linguistic distance 150	-0.0765	-0.102		0.0413	-0.100
	(0.121)	(0.122)		(0.0624)	(0.121)
linguistic distance 150_2	0.126	0.150			0.148
	(0.136)	(0.140)			(0.138)
fractionalization 150		0.0147	-0.0510	-0.0558	-0.0454
		(0.0386)	(0.116)	(0.117)	(0.116)
fractionalization 150_2			0.0588	0.0601	0.0589
			(0.127)	(0.127)	(0.126)
N	263730	263730	263730	263730	263730
pseudo R^2	0.091	0.091	0.091	0.091	0.091

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table C.10: Child mortality: Linear Probability Model

	(1) est1	(2) est2	(3) est3	(4) est4	(5) est5	(6) est6	(7) est7
linguistic distance 25	0.0134	0.0133	0.00706	0.0233*	0.00723	0.0145**	0.0146***
	(0.0165)	(0.0177)	(0.00544)	(0.0135)	(0.00569)	(0.00536)	(0.00530)
fractionalization 25		0.000458	-0.00205	-0.0139	-0.00489	-0.00888**	-0.00902**
		(0.0155)	(0.00636)	(0.0125)	(0.00430)	(0.00398)	(0.00404)
linguistic distance 50	0.0145	0.0144	0.00912	0.0238*	0.00803	0.0194***	0.0196***
	(0.0171)	(0.0180)	(0.00584)	(0.0137)	(0.00590)	(0.00660)	(0.00657)
fractionalization 50		0.000790	-0.000581	-0.0155	-0.00308	-0.00777*	-0.00863*
		(0.0190)	(0.00779)	(0.0156)	(0.00479)	(0.00452)	(0.00477)
linguistic distance 75	0.0153	0.0156	0.0104	0.0239*	0.00834	0.0222**	0.0226**
	(0.0173)	(0.0179)	(0.00620)	(0.0136)	(0.00604)	(0.00822)	(0.00829)
fractionalization 75		-0.00345	-0.00241	-0.0186	-0.00436	-0.0103*	-0.0120**
		(0.0224)	(0.00918)	(0.0184)	(0.00564)	(0.00534)	(0.00582)
linguistic distance 100	0.0159	0.0163	0.0110*	0.0234*	0.00778	0.0216**	0.0222**
	(0.0174)	(0.0178)	(0.00640)	(0.0135)	(0.00609)	(0.00863)	(0.00874)
fractionalization 100		-0.00685	-0.00223	-0.0217	-0.00499	-0.0120**	-0.0144**
		(0.0263)	(0.0103)	(0.0212)	(0.00565)	(0.00478)	(0.00532)
linguistic distance 125	0.0161	0.0167	0.0115*	0.0228*	0.00747	0.0207**	0.0213**
	(0.0173)	(0.0176)	(0.00648)	(0.0132)	(0.00605)	(0.0100)	(0.0101)
fractionalization 125		-0.0128	-0.00336	-0.0255	-0.00426	-0.0116**	-0.0145**
		(0.0306)	(0.0114)	(0.0246)	(0.00661)	(0.00496)	(0.00542)
linguistic distance 150	0.0156	0.0162	0.0111	0.0220	0.00669	0.0154	0.0160
	(0.0173)	(0.0175)	(0.00658)	(0.0132)	(0.00607)	(0.0128)	(0.0129)
fractionalization 150		-0.0164	-0.00106	-0.0285	0.000457	-0.00698	-0.0103
		(0.0358)	(0.0124)	(0.0286)	(0.00762)	(0.00607)	(0.00656)
linguistic distance 175	0.0152	0.0159	0.0108	0.0215	0.00617	0.0128	0.0133
	(0.0174)	(0.0175)	(0.00662)	(0.0131)	(0.00604)	(0.0154)	(0.0155)
fractionalization 175		-0.0213	-0.000880	-0.0319	0.00453	-0.00346	-0.00657
		(0.0419)	(0.0125)	(0.0333)	(0.00813)	(0.00718)	(0.00756)
linguistic distance 200	0.0150	0.0158	0.0106	0.0212	0.00576	0.0109	0.0110
	(0.0176)	(0.0176)	(0.00659)	(0.0132)	(0.00597)	(0.0176)	(0.0179)
fractionalization 200		-0.0256	0.000424	-0.0363	0.0107	0.00214	-0.000864
		(0.0489)	(0.0132)	(0.0386)	(0.00928)	(0.00882)	(0.00902)
linguistic distance 250	0.0145	0.0150	0.0117*	0.0199	0.00610	0.00795	0.00738
	(0.0183)	(0.0178)	(0.00668)	(0.0134)	(0.00600)	(0.0204)	(0.0207)
fractionalization 250		-0.0377	-0.0146	-0.0530	0.00208	-0.0115	-0.0140
		(0.0626)	(0.0165)	(0.0483)	(0.0150)	(0.0164)	(0.0157)
N	658755	658755	658755	658505	658505	657795	657795

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
The dependent variable is the probability of child death. OLS regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Table C.11: Infant mortality

	(1)	(2)	(3)	(4)	(5)	(6)
	est1	est2	est3	est4	est5	est6
infant_death						
linguistic distance 75	0.0440	0.0431	0.0699**	0.0273*	0.0509	0.0523*
	(0.0420)	(0.0433)	(0.0353)	(0.0152)	(0.0313)	(0.0317)
fractionalization 75		0.0110	-0.0404	-0.0155	-0.0338	-0.0410*
		(0.0474)	(0.0388)	(0.0243)	(0.0210)	(0.0222)
urban			-0.0765***	-0.0512***	-0.0519***	-0.0488***
			(0.0226)	(0.0109)	(0.0109)	(0.0112)
female			-0.0831***	-0.0846***	-0.0850***	-0.0850***
			(0.00482)	(0.00492)	(0.00493)	(0.00493)
education_years			-0.0254***	-0.0146***	-0.0138***	-0.0138***
			(0.00372)	(0.00224)	(0.00217)	(0.00216)
wealth index 2			0.0237*	0.00827	0.00670	0.00673
			(0.0140)	(0.0102)	(0.0101)	(0.00998)
wealth index 3			0.0150	-0.000726	-0.00250	-0.00219
			(0.0171)	(0.0154)	(0.0157)	(0.0158)
wealth index 4			-0.00292	-0.0341**	-0.0371**	-0.0366**
			(0.0193)	(0.0173)	(0.0173)	(0.0173)
wealth index 5			-0.0525*	-0.107***	-0.110***	-0.106***
			(0.0290)	(0.0212)	(0.0206)	(0.0202)
lndist2cap						0.0173**
						(0.00700)
ln_geog_dist						-0.0421
						(0.0340)
N	821918	821918	821609	820832	819820	819820
pseudo R^2	0.000	0.000	0.084	0.098	0.099	0.099

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is the probability of infant death. Pooled probit regressions have been used. Standard errors in parentheses are clustered at the DHS survey-country level. The numbers after linguistic distance and fractionalization variables indicate the radius of the circle around the mother in which these variables have been calculated.

Bibliography

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review 91 (5)*, 1369–1401.

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics 117 (4)*, 1231–94.

Ahlerup, P. (2009). The causal effects of ethnic diversity:an instrumental variables approach. *WORKING PAPERS IN ECONOMICS No 386, September.*

Alesina, A., R. Bakir, and W. Easterly (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics, MIT Press 114(4) November*, 1243–1284.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth 8, no. 2, June*, 155–194.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics 128*(2), 469–530.

Alesina, A. and E. La Ferrara (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics, MIT Press 115(3) August*, 847–904.

Alesina, A. F., S. Michalopoulos, and E. Papaioannou (2012). Ethnic inequality. *NBER Working Paper No. 18512, November.*

Alesina, A. F. and E. Zhuravskaya (2011). Segregation and the quality of government in a cross section of countries. *American Economic Review, American Economic Association vol. 101(5), August.*

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy Vol. 113, No. 1, February*, 151–184.

Andre, C. and J.-P. Platteau (1996). Land tenure under unendurable stress: Rwanda caught in the malthusian trap. *Cahiers de la Faculte des Sciences Econommiques et Sociales de Namur, Serie Recherche 164 - 1996/7.*

Ashraf, Q. and O. Galor (2013). The 'out of africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review Vol. 103 No. 1, February*, 1–46.

Baird, S., J. Friedman, and N. Schady (2011). Aggregate income shocks and infant mortality in the developing world. *Review of Economics and Statistics 93*(3), 847–856.

Baldwin, K. and J. D. Huber (2010). Ecnomic versus cultural differences: Forms of ethnic diversity and public good provision. *American Political Science Review 104, No. 4, November.*

Banerjee, A. and L. Iyer (2005). History, institutions and economic performance: The legacy of colonial land tenure systmes in india. *American Economic Review*.

Banerjee, A. and R. Somanathan (2007). The political economy of public goods: Some evidence from india. *Journal of Development Economics, Elsevier 82(2)*, 287–314.

Barooah, V. K. (2008). Deprivation, violence and conflict: An analysis of naxalite activity in the districts of india. *International Journal of Conflict and Violence 2 (2)*, 317–333.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76(2)*, 169–217.

Bellows, J. and E. Miguel (2009). War and local collective action in sierra leone. *Journal of Public Economics 93*, 1144–1157.

Besley, T. and R. Burgess (2000). Land reform, poverty reduction, and growth: Evidence from india. *Quaterly Journal of Economics 115 (2)*, 389–430.

Besley, T. and M. Kudamatsu (2006). Health and democracy. *The American Economic Eeview 96*(2), 313–318.

Besley, T. J. and T. Persson (2008). The incidence of civil war: Theory and evidence. *NBER Working Paper Series, Working Paper 14585*.

Bhatia, B. (2005). The naxalite movement in central bihar. *Economic and Political Weekly 40 (15) Apr 9-15.*

Blattman, C. and E. Miguel (2010). Civil war. *The Journal of Economic Literature 48(1)*, 3–57(55).

Bowen, H. (2010). Information at the grassroots: Analyzing the media use and communication habits of kenyans to support effective development. *Africa Development Research Series, available at: http://www.audiencescapes.org/africa-research-survey-quantitative-analysis-ghana-kenya*.

Brahm, E. (2005). Religion and conflict. *Beyond Intractability. Eds. Guy Burgess and Heidi Burgess. Conflict Information Consortium, University of Colorado, Boulder. Posted: November 2005 ¡http://www.beyondintractability.org/bi-essay/religion-and-conflict¿*.

Brockerhoff, M. and P. Hewett (2000). Inequality of child mortality among ethnic groups in sub-saharan africa. *Bulletin of the World Health Organization 78*(1), 30–41.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics:Methods and Applications.*

Caselli, F. and W. J. Coleman (2013). On the theory of ethnic conflict. *Journal of the European Economic Association January.*

Chakravarti, S. (2008). Red sun-travels in naxalite country. *Penguin Global*.

Ciccone, A. (2010). Transitory economic shocks and civil conflict. *unpublished 82(2)*, 287–314.

Collier, P. and A. Hoeffler (1998). On economic causes of civil war. *Oxford Economic Papers 50(3)*, 563–573.

Collier, P. and A. Hoeffler (2004). Greed and grievance in civil war. *Oxford Economic Papers 56(4)*, 563–596.

Desmet, K., J. F. Gomes, I. Ortuno-Ortin, and D. Puga (2013). The geography of linguistic divesrity (working title). *Work in Progress*.

Desmet, K., I. Ortuno-Ortin, and R. Wacziarg (2012). The political economy of linguistic cleavages. *Journal of Development Economics 97*, 322–338.

Desmet, K., I. Ortuno-Ortin, and I. Weber (2009). Linguistic diversity and redistribution. *Journal of European Economic Association 7(6)*, 1291–1318.

Do, Q.-T. and L. Iyer (2009). Geography, poverty and conflict in nepal. *HBS Working Paper 07-065.*

Dube, O. and J. F. Vargas (2008). Commodity price shocks and civil confict:evidence from colombia. *CID Graduate Student and Postdoctoral Fellow Working Paper No. 14*.

Duflo, E. and R. Pande (2007). Dams. *The Quarterly Journal of Economics 122(2)*, 601–646, 05.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics 112, no.4 November*, 1203–1250.

Engerman, S. L. and K. L. Sokoloff (1997). Factor endowments, institutions, and differential paths of growth among new world economies: A view from economic historians of the united states. *in Steven Haber, ed., How Latin America fell behind: Essays on the economic histories of Brazil and Mexico, 18001914. Stanford: Stanford University Press*.

Engerman, S. L. and K. L. Sokoloff (2002). Factor endowments, inequality, and paths of development among new world economies. *Economia: Journal of the Latin American and Caribbean Economic Association 3 (1)*, 41–88.

Esteban, J., L. Mayoral, and D. Ray (2012a). Ethnicity and conflict: An empirical study. *American Economic Review 102, No.4*, 1310–1342.

Esteban, J., L. Mayoral, and D. Ray (2012b). Ethnicity and conflict: Theory and facts. *Science 336, 858*.

Esteban, J. and D. Ray (1999). Conflict and distribution. *Journal of Economic Theory 87(2)*, 379–415.

Fearon, J. and D. Laitin (1996). Explaining interethnic cooperation. *American Political Science Review 90 (4)*, 715–35.

Fearon, J. and D. Laitin (2003). Ethnicity, insurgency and civil war. *American Political Science Review 97 (1)*, 75–90.

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*(2), 195–222.

Fox, J. (1997). The saliance of religious issues in ethnic conflicts: A large-n study. *Nationalism and Ethnic Politics 3, No. 3*, 1–19.

Fox, J. (2004). The rise of religious nationalism and conflict: Ethnic conflict and revolutionary wars. *Journal of Peace Research 41, No. 6*, 715–731.

Franck, R. and I. Rainer (2012). Does the leaders ethnicity matter? ethnic favoritism, education and health in sub-saharan africa. *American Political Science Review 106*(2, May).

Gawande, K., D. Kapur, and S. Satyanath (2012). Natural resource shock and conflict in india's red belt. *unpublished*.

Ghobarah, H. A., P. Huth, and B. Russett (2004). Comparative public health: The political economy of human misery and well-being. *International Studies Quarterly 48*(1), 73–94.

Gomes, J. F. (2012). The political economy of the maoist conflict in india: An empirical analysis. *UC3M Working Paper 12-18, Economic Series, June.*

Gomes, J. F. (2013). Religious diversity, intolerance and civil conflicts. *UC3M Working Paper 13-11, Economic Series, May.*

Guha, R. (2007). Adivasis,naxalites and indian democracy. *Economic and Political Weekly (Special Articles) August 11.*

Gyimah, S. O. (2002). Ethnicity and infant mortality in sub-saharan africa: The case of ghana. *PSC Discussion Papers Series 16*(10), 1.

Habyarimana, J., M. Humphreys, D. N. Posner, and J. M. Weinstein (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review 101*(04), 709–725.

Harris, J. (2010). The naxalite/maoist movement in india: A review of recent literature. *ISAS Working paper No. 109, 8 July.*

Hoelscher, K., J. Mikilian, and K. C. Vadlamannati (2011). Hearts and mines: A district-level analysis of the maoist conflict in india. *available at www.uni-heidelberg.de/md/awi/professuren/intwipol/india.pdf*.

Horowitz, D. L. (1985). *Ethnic groups in Conflict.* University of California Press.

Huntington, S. P. (1993a). The clash of civilizations? *Foreign Affairs 72(3)*, 22–49.

Huntington, S. P. (1993b). If not civilizations, what? samuel huntington responds to his crtics. *Foreign Affairs 72(5)*, 186–94.

Huntington, S. P. (1998). The clash of the civilizations and the remaking of the world order. *Simon & Schuster Ltd. West Garden Place Kendal Street London W2 2AQ 72(5)*, 186–94.

Isphording, I. E. (2013). Disadvantages of linguistic origin:evidence from immigrant literacy scores. *IZA DP No. 7360, April.*

Iyer, L. (2009). The bloody millennium: Internal conflict in south asia. *Harvard Business School, Working Paper 09-086.*

Jha, S. (2008). Complementarities and religious tolerance: evidence from india. *Stanford GSB Working Paper No. 2004*.

Kudamatsu, M. (2009). *Ethnic Favoritism: Micro Evidence from Guinea.* unpublished.

Kudamatsu, M., T. Persson, and D. Strmberg (2012). Weather and infant mortality in africa. *Prepared for the conference on Climate and the Economy in Stockholm, September 5-8*.

Kujur, R. (2008). Naxal movement in india: A profile. *Institute of Peace and Conflict Studies, New Delhi*.

Kumar, S., E. Dansereau, and C. Murray (2012). Does distance matter for institutional delivery in rural india: An instrumental variable approach. *Available at SSRN 2243709*.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny (1998). Law and finance. *Journal of Political Economy 106 (6)*, 113–55.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny (1999). The quality of government. *Journal of Law, Economics, and Organization 15 (1)*, 222–79.

La Porta, R., F. Lopez de Silanes, A. Shleifer, and R. Vishny (2000). Investor protection and corporate governance. *Journal of Financial Economics 58 (1-2)*, 3–27.

Lieberman, E. S. (2007). Ethnic politics, risk, and policy-making a cross-national statistical analysis of government responses to hiv/aids. *Comparative Political Studies 40*(12), 1407–1432.

Macours, K. (2011). Increasing inequality and civil conflict in nepal. *Oxford Economic Papers 63(1)*, 1–26.

Malhotra, N. (2012). Inadequate feeding of infant and young children in india: lack of nutritional information or food affordability? *Public Health Nutrition 1*(1), 1–9.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of Public Economics 89*(11), 2325–2368.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy 112 (41)*, 725–753.

Mitra, A. and D. Ray (2010). Implications of an economic thory of conflict: Hindu-muslim violence in india. *unpublished*.

Montalvo, J. and M. Reynal-Querol (2000). A theory of religious conflict and its effect on growth. *IVIE working papers WP-EC 2000-04*.

Montalvo, J. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict and civil war. *The American Economic Review 95(3) June*, 796–816.

Montez, D. (2011). Identifying health information needs in tanzania evidence from the audiencescapes national survey. *AudienceScapes Development Research Report, available at: http://www.audiencescapes.org/africa-research-reports-identifying-health-information-needs-Tanzania-malaria-maternal-child*.

Nunn, N. and L. Wantchekon (2009). The slave trade and the origins of mistrust in africa. *American Economic Review*.

of India, G. (2008). Development challenges in extremist affected areas. *Report of an Expert Group to the Planning Commission*.

Pande, R. (2003). Can mandated political representation increase policy influence for disadvantaged minorities? theory and evidence from india. *American Economic Review 93, No.4*, 1132–1151.

Platas, M. R. (2010). Africa's health tragedy? ethnic diversity and health outcomes. *Prepared for delivery at the Winter 2010 Working Group on African Political Economy No. 7360, April*, December 17–18.

Pongou, R. (2009). Anonymity and infidelity: Ethnic identity, strategic cross-ethnic sexual network formation, and hiv/aids in africa. *Unpublished paper, Department of Economics, Brown University*.

Shariff, A. F. and M. Rhemtulla (2012, 06). Divergent effects of beliefs in heaven and hell on national crime rates. *PLoS ONE 7*(6), e39048.

Singleton, K. and E. Krause (2009). Understanding cultural and linguistic barriers to health literacy. *OJIN: The online journal of issues in nursing 14*(3).

Spolaore, E. and R. Wacziarg (2009). The diffusion of development. *The Quarterly Journal of Economics 124*(2), 469–529.

Sundar, N. (2008). Subalterns and sovereigns, an anthropological history of bastar (1854-2006). *Oxford University Press*.

UNICEF (2012). Levels trends in child mortality: Report 2012. *UN Inter-agency Group for Child Mortality Estimation, United Nations Childrens Fund, UNICEF*.

Vanden Eynde, O. (2011). Targets of violence: Evidence from india's naxalite conflict. *Job market paper, LSE*.

Verwimp, P. (2003). Micro-level evidence from rwanda. *HiCN Working Papers 08, Households in Conflict Network*.

Vigdor, J. L. (2004). Community composition and collective action: Analyzing initial mail response to the 2000 census. *Review of Economics and Statistics 86*(1), 303–312.

Zhou, Y. (2010). Health information gaps in zambia evidence from the audiencescapes national survey. *Africa Development Research Series, available at: http://www.audiencescapes.org/zambia-health-information-gaps-urban-rural-HIV-AIDS-malaria-family-planning-maternal-child-data-services-media-development*.