

HCEO

hceconomics.org

**Human Capital and Economic Opportunity
Global Working Group**

Working Paper Series

Working Paper No. 2014-011

**Colonial Legacy, Linguistic Disenfranchisement and the Civil Conflict
in Sri Lanka**

Paul Castañeda Dower
Victor Ginsburgh
Shlomo Weber

April, 2014

Human Capital and Economic Opportunity Global Working Group
Economics Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
www.hceconomics.org

Colonial Legacy, Linguistic Disenfranchisement and the Civil Conflict in Sri Lanka*

Paul Castañeda Dower

New Economic School

Center for the Study of Diversity and Social Interactions, New Economic School

Victor Ginsburgh

ECARES, Université Libre de Bruxelles

CORE, Université Catholique de Louvain

Center for the Study of Diversity and Social Interactions, New Economic School

Shlomo Weber

Southern Methodist University

Center for the Study of Diversity and Social Interactions, New Economic School

March 4, 2014

Abstract

Polarization measures, that are used in examining the empirical relationship between ethnic divisions and violent conflict, heavily rely on mechanisms of group identification and often use somewhat arbitrary divisions of a society into ethnic groups. In this paper we construct two new measures of polarization, one that accounts for differences in linguistic policies across localities during the colonial era and one that accounts for the differences over time and across localities in the experience of violence throughout the conflict episode. By examining the protracted war in Sri Lanka and applying these indices (and their combination) to a data set describing victims of the civil conflict by district and year, we are able to better identify the effect of ethnolinguistic polarization on the civil conflict in the country. We find that, for each of our polarization indices, there is a positive effect on the conflict. The historical underpinnings of our indices allow us to demonstrate in a quantitative and concrete way the relevance of historical processes for understanding episodes of civil conflict.

*Corresponding Author: Paul Castañeda Dower, Tel: +7 (495) 956-9508 ext. 157, Fax: 8 (499) 129-3722, Email: pdower@nes.ru. The authors are grateful to Dmitry Plotnikov for sharing data and giving us permission to reproduce several of his maps and figures. The authors wish to acknowledge the support of the Ministry of Education and Science of the Russian Federation, grant No. 14.U04.31.0002.

1. Introduction

In order to understand the empirical relationship between ethnic divisions and violent conflict, Esteban et al. (2012) pose two challenges: making a decision on the relevant notion of divisions and, second, estimating the sensitivities of these divisions to the underlying conflict. This paper takes up this claim in the context of the recent Sri Lankan Civil War, a conflict that lasted nearly 30 years taking the lives of an estimated 80,000-100,000 citizens (almost 0.5% of the total population). Through the lens of the colonial legacy and its impact on ethnic divisions, we develop a measure of the relevant ethnic divisions that is different from standard polarization indices. We then allow for these divisions to be sensitive to the underlying conflict as the conflict progresses and spreads throughout the island, creating a dynamic measure of polarization. Using both measures, we find that ethnic divisions centered on linguistic disenfranchisement have exacerbated the conflict.

The two challenges above pose considerable econometric difficulties, especially for the usual venue of empirical analysis in this literature, the much maligned cross-country study. It is well-known that cross-country studies fail to account for two important sources of heterogeneity. First, there is heterogeneity among countries with respect to the choice of an appropriate notion of ethnic division. The standard practice of using a country-neutral measure will tend to attenuate the effect since a uniformly applied selection of ethnic divisions will often result in measurement errors uncorrelated with the unobserved, idiosyncratic determinants of conflict. However, turning to country-specific measures will invite concerns about comparability and data-mining. Second, there may be time-invariant unobservables that explain the

level of conflict. Standard panel data methods to deal with this heterogeneity are rarely appropriate since most polarization indices usually do not change much over time. One remedy is to allow the time-invariant measure to have different causal impacts over time. However, this approach is undesirable for both theoretical, in the sense of the two preconditions above, and econometric reasons, since the total effect may turn out to be poorly estimated and, ultimately this approach may not indicate which effect is the relevant one from a policy point of view.

How does the within-country analysis of the conflict in Sri Lanka that we study deal with these issues? We tailor our measure to the Sri Lankan context by emphasizing the link between colonial heritage and the ethnolinguistic conflict. The backlash against colonial policies is often a source of post-independence conflict. Sri Lanka is a polarized society with a few ethnic groups, the largest of which are the majority Sinhalese and the minority Tamils. The conflict is a special case because there is a clear linguistic component over which the majority and minority could disagree. The linguistic origins of the conflict stem from the Sinhala-only movement and the subsequent governmental attempts and failures to impose monolingual policies and other types of policies that favored the majority and led to disenfranchisement along ethnolinguistic lines. First, there was a direct effect, which changed the relative costs and benefits of linguistic endowments and language acquisition. This type of disenfranchisement increased the cost of access to public employment such as civil service jobs and public resources such as higher education. Second, linguistic disenfranchisement created uncertainties concerning the stability of clientelistic networks which had been formed along ethnic lines, and altered access to public resources.

These two types of disenfranchisement, which both focus on changes in entitlements as opposed to levels, are the motivation for our measure of ethnic divisions, and correspond to the findings in Miguel et al. (2004) that it is changes in income and not levels that predict civil conflict.

Our empirical measure of ethnic divisions is unique in that it specifically accounts for the scope of disenfranchisement, albeit in the framework of well-studied polarization measures. Even though the concepts of polarization and disenfranchisement are not identical, one can argue that in the case of two linguistic groups, there is a positive correlation between the two, that is, an increase of the degree of polarization in a two-group society raises the level of disenfranchisement generated by one-language policies.¹ We focus on polarization empirically, because it is difficult to observe the extent of disenfranchisement independently from polarization and conflict. Specifically, we use outcomes generated by linguistic policies conducted in colonial Ceylon as distances in the degree of alienation between groups.

By historical accident, the colonial linguistic policies enhanced the scope for disenfranchisement along ethnolinguistic lines by inducing spatial variation in the number of English speakers on the island. We argue that the change in English proficiency during the colonial period affected the extent of ethnic divisions that were exogenous to the ensuing civil conflict. Since these divisions could have been aggravated during the nearly 30-year period of the civil war, their sensitivity could have changed over time. To account for this possibility, we construct a dynamic polarization index based on the arguments that past conflict primes and intensifies divisions. This index

¹For a discussion of relevant concepts see Ginsburgh and Weber (2011).

has an attractive feature: the more intense is the experienced violence during the conflict, the more polarization increases.

Since data on violent conflict in a developing country are difficult to obtain, we collected data on the total number of victims of Liberation Tigers of Tamil Eelam (LTTE) attacks per district by year. This measure accurately tracks only one of the major sources of violent conflict during the war and one needs to be careful in interpreting it. It does not capture pogroms, which were serious in certain districts, nor does it capture conflicts primarily associated with other groups. Using district-level socioeconomic factors such as literacy in English, urban share of the population and population growth as controls, we test the hypothesis that polarization caused the intensity of conflict. We show that both our measures of polarization, the static and the dynamic, which account for the cross-sectional differences in disenfranchisement explain conflict. Taken together, the empirical results suggest the importance of the measurement of divisions in indices of polarization. In this specific case, the aggravation of linguistic-based divisions caused by the colonial legacy improves the precision of the estimates of the effect of polarization on conflict.

We also perform several robustness checks. In particular, we rerun the results on a subsample of districts which report at least one victim from LTTE attacks, dropping the districts which experienced the worst pogroms.

The organization of the paper is as follows. First, we discuss the previous literature. Second, we discuss the civil war and provide the necessary historical details. Next, we describe the data and present the results. Finally, we conclude.

2. Previous Literature

Understanding civil conflict is often couched in terms of greed (to what extent do the relevant parties gain materially from the conflict) and grievances (to what extent have the relevant parties suffered in the past). Collier and Hoeffler (2004) argue that while there are many ethnic groups in conflict, as documented by Horowitz (1985), few escalate into civil war and the difference in outcomes can be traced to the ability of opportunistic political actors to capture resources. Reynal-Querol and Djankov (2010) find indirect support for this thesis in that the lack of respect for the rule of law predicts the onset of civil war. We hold these factors fixed in our analysis by focusing on a single country. Moreover, we do not study the onset of civil war, rather our focus is on the intensity of the conflict.

The empirical literature of ethnolinguistic diversity and economic outcomes is well-established. See among others Alesina et al. (1999); Easterly and Levine (1997); Alesina et al. (2003); Ginsburgh and Weber (2011). Much emphasis has been placed on how ethnic divisions can deteriorate public goods provision (Alesina et al., 1999; Desmet et al., 2005; Montalvo and Reynal-Querol, 2005). While there appears to be consensus surrounding public goods provision, there is still a debate on the role of ethnic divisions and violent conflict.²

²In an important sense, social order and its maintenance is a public good so the lack of empirical consensus is puzzling. However, there are theoretical and empirical reasons for why less consensus has been achieved. The maintenance of order is a very different type of public good than schooling or public health and therefore merits special attention. First, the maintenance of social order encompasses the provision of most public goods and, in fact, is necessary for many of them. Second, the relationship between conflict and the maintenance of social order on some level and at some point is discrete by nature. Third, the maintenance of social order has a long memory and, while there is persistence in the provision of public goods, the quality of public goods is not as sensitive to history as social order. And, fourth, the externalities can be global, similar to public health, but

Fearon and Laitin (2003) are critical of the literature that claims to have discovered a positive relationship between ethnic divisions and violent conflict based on cross-country regressions, and demonstrate under reasonable specifications that fractionalization is not related to conflict. If fractionalization reflects linguistic differences alone, Laitin (2000) even argues that linguistic diversity is an instrument for peace. Montalvo and Reynal-Querol (2005) respond by arguing that fractionalization is not the right measure of ethnic divisions, and show that when divisions are represented by polarization measures, they have a positive impact on conflict. The findings in Montalvo and Reynal-Querol (2005) are supported by the theoretically-motivated approaches in Esteban and Ray (1994) and Esteban et al. (2012). Alesina and La Ferrara (2005) show that the effect of fractionalization is strongest in poor countries. Esteban et al. (2012) show that both fractionalization and polarization can be positively related to violent conflict. Which measure is appropriate depends upon what is at stake in the conflict, in particular how public is the nature of the resources affected by the dispute. The authors then construct a measure of the “publicness” of the resources at stake and find support for their thesis.

Given these insights, the literature has moved towards using less aggregated data, giving much better control over omitted institutional factors and country-level variables such as the publicness variable. Mitra and Ray (2012) investigate the relationship between economic outcomes, religious divisions and violent conflict in India. Field et al. (2008) are able to disaggregate even further by exploring neighborhood

of a different nature. In addition to these differences, the empirical exercise necessary to test the relationship between ethnic divisions and conflict is more difficult than for the case of public goods because of data quality and availability.

level variation in religious diversity and conflict in the city of Ahmedabad, India. Novta (2013) studies district-level differences in conflict in the Bosnian civil war. Do and Iyer (2010) explore how geography and poverty explain the intensity and incidence of civil war in Nepal. They find no statistical relationship between conflict and ethnic divisions.

Often, the specific historical experiences have not been accounted for in the measure of group divisions. There is a growing literature on how historical experiences can influence current outcomes in profound ways. In the case of conflict, Besley and Reynal-Querol (2013) trace the current violent outcomes in Africa to violence occurring before the colonial rule. An important source of divisions based on historical experience is linguistic policy and in particular linguistic disenfranchisement (Ginsburgh and Weber, 2011; Ginsburgh et al., 2005). Besides the inherited divisions of the past, the conflict episode’s own history is an important factor in explaining violent outcomes. Novta (2013) demonstrates how violent conflicts can spread, arguing that ethnically homogenous localities give rise to violent movements that then spill over to more ethnically heterogenous neighborhoods.

Our paper contributes to the literature on ethnic divisions and conflict by providing a within-country analysis of the civil conflict in Sri Lanka. We introduce two new measures of ethnic divisions instead of relying on the less precise “off-the-shelf” measures. We account for how ethnic divisions vary spatially and temporally by linking these divisions to the specific history of Sri Lanka. Since there is a strong linguistic component associated with the conflict, we pay special attention to historical differences in linguistic disenfranchisement. Our paper also incorporates the

conflict episode's history, allowing violent outcomes in neighboring regions to affect the measure of ethnic divisions directly.

3. The History of the Civil Conflict in Sri Lanka

The civil conflict took place over the course of three decades. Figure 1 illustrates the timeline of the number of casualties, while Figure 2 gives the number of victims by district. The conflict has been separated into four phases by the Sri Lankan Ministry of Defense, which correspond to periods as a time unit of analysis in our econometric results. Eelam War I started in 1983 and ended with a peace negotiated in 1987 with the help of Indian peacekeepers. By 1990, Eelam War II (1990-1995) had begun due to the conflict escalating again after failed peace negotiations. Between Eelam War II and Eelam War III (1995-2002), it was not peace that determined the break but a different course of action by LTTE, which graduated from terrorist attacks on civilians to large-scale military campaigns against the Sri Lankan Army. During the third Eelam War, LTTE managed to succeed in making the northern part of the island a stronghold. However, the tide turned against LTTE in Eelam War IV (2003-2009) with the diminished international support resulting from the September 11th, 2001 terrorist attacks in the United States as many states began to officially recognize LTTE as a terrorist organization. In addition, the tsunami in 2004 primarily hit the north-eastern part of the island, the areas where LTTE had the most control. In the aftermath and recovery from the tsunami, the Sri Lankan army started offensive operations, which ended with the death of Prabhakaran, the

head of LTTE, in May 2009. Shortly thereafter, the civil conflict concluded.

As one can see by the war period classification, the LTTE played the central opposition role. Its preferred instrument of attack was political assassination, the most famous of which was that of Sri Lankan President Ranasinghe Premadasa in 1993. However, the LTTE did not assassinate Sinhalese politicians only; they also assassinated Tamil politicians who cooperated with the government. In addition, the LTTE was not the only organization committing violence against the government and the government attributed some attacks to LTTE without the organization claiming them. There were also extremist groups among the Sinhalese, for example the Peoples Liberation Front (JVN), whose alleged attacks killed a large number of people as well as a number of pogroms, including the one that plunged the conflict into a full-blown civil war in 1983, which resulted in as many as 2,000 victims and 100,000 refugees (Laitin, 2000).

As with any prolonged conflict, the complexity of the potential causes and feedback mechanisms makes it difficult to give a conclusive treatment. Among the Sinhalese and Tamils, one can point to causal factors related to differences in race, religion, region and custom (Geertz, 1963). Here, we focus on linguistic disenfranchisement as one potential cause. Laitin (2000) demonstrates that the case of Sri Lanka had all the ingredients for a language-based violent conflict. In fact, the first instance of violent civil conflict in the newly independent nation occurred as a direct result of the passage of the Official Language Act No. 33 in 1956, also known as the Sinhala-only Act, which explicitly and vociferously instituted a monolingual policy, and led to linguistic disenfranchisement of Tamil speakers.

The potential for linguistic-based conflict was rooted in the *swabasha* movement during the late colonial period. This movement aimed at preserving and enriching indigenous languages, especially with respect to English under British rule. Importantly, the movement consisted of both Sinhalese and Tamils working together to increase the status of their respective languages. During the transition to independence and for various reasons, Sinhalese politicians neglected the Tamil demands for linguistic equality, culminating in the Sinhala-only Act.

The colonial legacy of British rule contributed to the particular tensions that emerged from the official resolution of the *swabasha* movement. During the British colonization, the minority Tamil population gained a significant economic advantage over the Sinhalese majority, particularly because of the relatively better English training in the northern part of the island. De Votta (2004) reports that the Tamil population held 33% of civil service jobs, 40% of judicial service jobs and 31% of the students in university, figures that are much greater than their representative share. In addition, English was the language of the elites, which meant that the average Tamil likely had better access to clientelistic networks that were in place under British rule and likely persisted after independence. These networks were organized along ethnic lines and were far reaching affecting not just the elite but also peasants (Korf, 2005).³

³One can point to a number of other reasons for why the tensions between the Tamils and the Sinhalese escalated into a full blown civil war. The first set focuses on several key parliamentary actions that significantly hindered the Tamil population. These include the Citizenship Act, the Parliament Act #48, and the Prevention of Terrorism Act. Others point to demographic reasons. There is a very large Tamil population located on the sub-continent of India. Some argue that this population caused the Sinhalese to have the desperation of an afflicted minority group (Tambiah, 1986). Finally, Kapferer (1988) argues that, while there has been relatively peaceful coexistence of the two main groups for two millennia, there certainly was a history of violence. The differing

While the civil conflict clearly contained an ethnolinguistic component, some scholars believe that it was not causal and merely a representation of a different sort of underlying conflict. Most notably, Laitin (2000), who even entertains the possibility that the linguistic differences actually ameliorated the conflict, asks a provocative question: “Why did the language issue disappear from public debate in inverse proportion to the level of escalation of violence on the island?” Indeed, Tamil gained the status of a national language in 1977 before the beginning of the outbreak of the war.

For Laitin, the driving force of the conflict is best characterized as a “sons-of-the-soil” conflict, whereby the land colonization in the northern territories by the Sinhalese that occurred after independence threatened the security of the Tamil way of life. In addition, Laitin argues that, in practice, English and Tamil were tolerated since it was bureaucratically impossible to switch over to Sinhala-only in 24 hours as the proponents of the 1956 Act claimed it could do. In addition, non-Sinhalese speakers who had been in office during the colonial administration were given a grace period to switch over before losing their jobs. But these arguments are largely off the mark if one takes seriously the effects of linguistic disenfranchisement. First, it is not merely the realized outcomes that matter for tensions but the increased uncertainty surrounding the relative benefits of language acquisition and the returns to already acquired linguistic skills. Second, it is not merely the Tamil language that suffered linguistic disenfranchisement but also English, which affected the Tamil population relatively more than the Sinhalese. Third, linguistic disenfranchisement of English

myths over heroes, the legitimatizing claim of official statements in reinforcing interpretations about history, and the political support of custom raised the stakes of control over public resources.

took time to occur and to only slowly weaken and disrupt the clientelistic networks inherited from the colonial period. One would therefore expect the conflict to take time to materialize. Finally, it is not difficult to see that symbolic disenfranchisement may have an impact on tensions even when de facto disenfranchisement is minimal. At the very least, the 1956 riots, which followed directly from the linguistic disenfranchisement of the Tamils, became a part of the history of the conflict itself and contributed to the subjective or psychological costs of heterogeneous population on the island.

We end this historical section by providing some historical details concerning the colonial linguistic policy under British rule. The British were initially not interested in English language training, despite the fact that the English-speaking population for nearly every ethnic group on the island, including the majority Sinhalese, was virtually nonexistent.⁴ The British policy changed dramatically under the leadership of Colebrooke, a British administrator sent by the imperial government to ascertain why the colony was losing money, who strongly advocated English training and use. Colebrooke had been impressed by the effort of American missionaries in teaching English as opposed to British missionaries who were focused on spreading Christianity in the vernacular.

Notwithstanding the colonial authorities' efforts, American missionaries spread English among the youth but Americans were restricted to the northern part of the island. Their efforts centered around the Vaddukkodai (Batticotta) Seminary on Jaffna peninsula. The decision to locate in the north was more or less a historical

⁴The one exception was the Burghers. Burghers did not need English language instruction since many were using it in trade and connections with Europe.

accident. American missionaries wanted to go to Calcutta, India, but were not welcome because of the War of 1812. A member of the group fled India for Sri Lanka. In Colombo, the governor of Ceylon greeted him in a friendly manner and encouraged him to set up a mission on the island. The Americans were, in turn, interested in Sri Lanka because of the Tamil population, which could be used to reach the many Tamils in South India. The governor and missionaries agreed on the establishment of a mission in some abandoned churches in the north of the island. The British authorities then restricted the Americans to the north because of security concerns, which had been heightened after the War of 1812 (Root, 1916).

After Colebrooke's influence, the British aimed at English language instruction all across the island but due to poor instruction the outcome was low literacy with pockets of literacy centered around areas with good teachers:

In several Anglo-vernacular schools which I have visited, the teacher supposed to teach English has been quite unable to converse with me in English, and it has been necessary for the inspector who accompanied me to act as interpreter (Government of Ceylon, 1879, quoted in Coperehewa, 2011).

The lack of success led the British to shut down English training in many government schools. In 1900, there were four government schools that provided English instruction compared to 144 missionary-assisted schools.

To sum-up, the English language policy of the British, which was largely unsuccessful, and the efforts of American missionaries, which were successful but limited in reach, provided important variation in the change in English language proficiency

during the colonial rule.

4. Polarization indices

Esteban and Ray (1994) formalized the concept of *polarization* for societies with income differences between groups. They derive the form of their polarization index from a certain numbers of axioms which lead to an index of the form

$$\sum_{i \neq j} p_i^{1+\delta} p_j |y_i - y_j|, i, j = 1, 2, \dots, n,$$

where p_i and p_j are shares of groups i and j in the geographic entity that is analyzed, y_i and y_j represent the average incomes of groups i and j , $|y_i - y_j|$ measures the distance between the two groups in terms of (average) incomes and δ is a positive constant.⁵

Polarization introduces the idea of *identification* with individuals of one's own group, say i and *alienation* with those of group $j \neq i$. The insertion of identification-alienation considerations in the usual form of diversity indices, $\sum_{i \neq j} p_i p_j$, $i, j = 1, 2, \dots, n$, affects the functional form of polarization indices in two ways: the exponent $(1 + \delta)$ instead of 1, and distances $|y_i - y_j|$ between (the average incomes of) groups. An individual in a specific group feels identified with other individuals of his group. The degree of identification depends on p_i , the size of the group, and is given by the value p_i^δ , where the positive sign of δ implies that identification is stronger in

⁵Esteban and Ray (1999) take a more general interpretation of the distances between groups, using average differences between income levels as a proxy for "differences in preferences over public goods" between groups.

larger groups.

The two features, identification with own group and alienation towards others, create societal antagonism. Given the population shares of two groups p_i and p_j , the combined effective antagonism of group i towards group j is equal to $p_i^\delta \times p_i \times p_j$ or $p_i^{1+\delta} \times p_j$. The societal level of *social effective antagonism* is the sum of the effective antagonisms between all pairs of groups. The axioms introduced by Esteban and Ray allow them to narrow down the range of δ to values between 0 and 0.6, according to the form of (in their case) the income distribution of individuals.

Reynal-Querol (2002) considered a special case of Esteban and Ray's index with $\delta = 1$, the axiomatic support to which was suggested later by Geng (2012):

$$\sum_{i,j,i \neq j} p_i^2 p_j = \sum_i p_i^2 (1 - p_i).$$

They focus on what they call *relevant* groups, thus avoiding the issue of distances by implicitly using a dichotomous 0-1 distance measure between groups.

4.1. L- and E-indices of polarization

Nothing, however, prevents using the extension of the polarization index to multilingual societies suggested by Desmet et al. (2009). The degree of alienation is then given by $p_i^\delta \times d_{ij}$, where d_{ij} represents the linguistic distance between groups i and j , and leads to the following index:

$$\sum_{i,j,i \neq j} p_i^2 p_j d_{ij}, \tag{1}$$

which reintroduces distances between groups that are present in Esteban and Ray's index. However, the distance metric is much more complicated in this case since linguistic proximity lacks the linear structure of the income differences utilized in Esteban and Ray's model.

In the usual context, (1) describes polarization in countries, and there is one such index per country. Looking at the country-level restricts the analyst to a common historical process for all countries. The commonly used linguistic or genetic distance matrices, and even the ad hoc one of zeros and ones, are all the result of long-run historical processes. Here we deal with one country only, taking into account each of the 25 Sri-Lankan districts, and can therefore, focus on less distant historical processes.

As illustrated in figure 2, there exist four linguistic groups (Sinhalese, Tamils, Moors and Europeans, mainly English-speaking) in various proportions in each district, and the index varies across districts since p_i and p_j vary. In the standard formulation, linguistic distances would remain the same in all districts. This (linguistic) polarization index will be called L-index. We, in fact, modify the L-index to the E-index, to reflect in a better way the colonial heritage of Sri Lanka. The distances utilized in the construction of E-index are derived on the basis of differences in English linguistic outcomes during the late colonial period in each district. The independence of the country affected the relative status of English speakers vs. non-English speakers over a wide range of conflict determinants. Of course, English language proficiency can directly and indirectly affect violent conflict and these effects could be entirely independent of linguistic disenfranchisement. To remedy this

problem, we focus not on levels but on changes in English proficiency under the colonial rule. In particular, we focus on the generation of people who invested in learning English during and leading up to the *swabasha* movement, a push by both Tamils and Sinhalese for greater official status and preservation of their respective languages relative to English. In the historical section, we outlined our argument for why this change in English language proficiency is exogenous.

This generation of English speakers are important for several reasons. First, after independence, it lost some of the benefits of the investment in English. We would also expect the children of such individuals to experience this shock to human capital, driven by linguistic disenfranchisement. Second, the uncertainties surrounding the official linguistic policy could heighten tensions and concerns about what is at stake in the political process. Third, the diminished importance of English affected the stability of clientelistic networks, over which English speakers had a relatively strong hold during the colonial and transition periods.

E-index distances are constructed as follows for each district: the linguistic distance between English and each of the four groups (Sinhalese, Tamils, Moors and Europeans, mainly English-speaking), d_{iE} is multiplied by col^k , an index of colonial heritage measured by the change in district k of English language proficiency between 1921 and 1946, divided by the 1946 level.⁶ This distance which varies across districts, can be written as

$$d_{ij}^k = d_{iE}^k = d_{iE} \times col^k,$$

⁶We normalize by the 1946 level in order to better capture the impact of the change relative to the level effect.

Then the district k 's E-polarization index is represented as

$$\sum_{ij, i \neq j} (p_i^k)^2 p_j^k d_{iE}^k.$$

Since the 1971 census⁷ does not contain data on English literacy in order to verify that changes in English literacy from 1921 to 1946 preserved the disparities instigated by the colonial legacy. Linguistic distances between the four groups are calculated using linguistic trees.⁸

4.2. Dynamic polarization indices

Given that the four consecutive wars extended over 26 years (1983-2009), it may have changed the level of effective antagonism for each group in each district. It seems thus appropriate to also use a dynamic version of the polarization index, assuming that a district's conflict experience has an effect on each district. The realization of violent conflict can heighten and possibly distort the perception of social antagonism. We incorporate this aspect of societal divisions by assuming that the violence experienced in the districts that border district k directly impact the *perceived effective antagonism* of each group in the district. In particular, we assume that, as the conflict intensifies, the identification with one's own group and its size is magnified as well as the perception of each rival group. This magnification occurs relative to a reference population, which could be the population during peace time

⁷1971 is the last year for which there are census data before the wars started in 1983.

⁸See Fearon (2003); Ginsburgh and Weber (2011) or the linguistic tree at http://www.ethnologue.com/family_index.asp.

or any of the previous conflict periods. Following the notation in the static case, we denote $p_{i,t}^k$ as the population share of group i in district k in period t .

Let $\pi_{i,t}^k$ be the perceived population share of group i in region k at period t (for $t = t_0, \dots, t_T$). Then, we define the magnification process in the following way:

$$\pi_{i,t}^k = (1 - \omega_{t-1}^k) \sum_{j=0}^{t-1} \alpha_j f(p_{i,0}^k, \dots, p_{i,j}^k, \omega_0^k, \dots, \omega_j^k) + \omega_{t-1}^k$$

where the α_j 's are weights over the possible reference populations, so that all α_j are nonnegative numbers and their sum is equal to 1, ω_{t-1}^k is the ratio of victims in period $t-1$ in the districts that border (inclusive) district k with respect to the total population of the region, and $f(\cdot)$ is always positive and increasing in ω and p .

The central parameter of this index is ω , which approximates the probability that one could have been killed by an attack as well as gives an indication of the intensity of the conflict in the previous period. If the number of victims in the previous period, ω_{t-1}^k , is very small, say, close to zero, the degree of antagonism does not raise and remains almost at the same level as in the previous period. However, if the number of victims in the previous period is very large, the antagonism rapidly raises and may even approach the highest level of 1. In general, during the conflict $\pi_{i,t}^k$ is increasing if ω_{t-1}^k is increasing with the rate of increase being determined by the number of victims in the previous period. In addition, smaller groups are magnified more than larger groups which, as is well known, will yield higher values of the polarization index.

There are three special cases that attract special attention.

For $\alpha_{t-1}^k = 1$, $\alpha_j^k = 0$ for all $j = 0, \dots, t-2$ and $f(p_{i,j}^k, \omega_j^k) = p_{i,j}^k$, we obtain a

formulation that exhibits memorylessness where the perceived antagonism is determined only by the previous period:

$$\pi_{i,t}^k = (1 - \omega_{t-1}^k)p_{i,t-1}^k + \omega_{t-1}^k = p_{i,t-1}^k + (1 - p_{i,t-1}^k)\omega_{t-1}^k$$

which can also be written as $\pi_{i,t}^k = p_{i,t-1}^k + (1 - p_{i,t-1}^k)\omega_{t-1}^k$.

For $\alpha_0^k = 1$, $\alpha_j^k = 0$ for all $j = 1, \dots, t-1$, and $f(p_{i,j}^k, \omega_j^k) = p_{i,j}^k$, we obtain a formulation that exhibits anamnesis since the population share of the peaceful period is recollected as the reference population:

$$\pi_{i,t}^k = (1 - \omega_{t-1}^k)p_{i,0}^k + \omega_{t-1}^k$$

Since $\omega_0^k = 0$, $\pi_{i,1}^k$ is simply, $p_{i,0}^k$, the population share of group i in district k in the last period of peace before the conflict.

For $\alpha_{t-1}^k = 1$, $\alpha_j^k = 0$ for all $j = 0, \dots, t-2$ and $f(p_{i,j}^k, \omega_j^k) = (1 - \omega_j^k)\pi_{i,j}^k + \omega_j^k$, we obtain a recursively formulated anamnestic index, which also uses the peaceful period as the reference population but the transition from one period to the next shares a similar structure:

$$\pi_{i,t}^k = (1 - \omega_{t-1}^k)\pi_{i,t-1}^k + \omega_{t-1}^k$$

Given our data limitations, of these three, we can only construct the anamnestic indices since we observe population shares only in the last peace period and not in

each period.⁹ In the empirical section, we adopt the simple anamnestic index instead of the recursive one. The reasons for doing so are both empirical and conceptual. Over time, the recursive anamnestic index loses variation, both across time and districts, putting a strain on estimation, especially for low probability events. While the recursive index is attractive from the point of view that the magnification of population contains memory of past conflicts, it is reasonable to approximate the dynamics of the societal level of perceived social effective antagonism as characterized by an underlying structural level of antagonism that experiences temporary flare-ups, as the simple anamnestic index does.

We then plug the $\pi_{i,t}^k$'s into district k 's polarization index to get the societal level of perceived social effective antagonism in a district, which can be written:

$$\sum_{ij} (\pi_{i,t}^k)^2 \pi_{jt}^k d_{ij}^k, t = t_0, 1, 2, \dots, t_T,$$

where t_0 and t_T have to be chosen in an appropriate way. This index is calculated for both the L- and the E- forms.

5. Data and Methods

We construct district-level measures of conflict intensity, ethnic polarization, urban share of the population, population growth, and in particular, the change in the share of the population that is Sinhalese, and English literacy.

⁹We could approximate these population shares but we would then face bias in the empirical estimates due to nonclassical measurement error.

Number of victims. The main measure of conflict intensity is the number of victims of the LTTE terrorist attacks, shown in figure 2. While this measure is far from ideal, it is an adequate one, given the central role that LTTE played in the civil conflict, and, in some respects, the best available one, given the severe data limitations for such a data generating process.

The LTTE were involved in 284 terroristic acts, assassinations and military attacks from 1975 to 2009. For each of these acts, we observe the town, district and province as well as an estimate of the number of casualties, cross-checking when possible.¹⁰ The number of LTTE victims by district and war period can be found in Table 2. There are a substantial number of district-periods with zero victims and some districts have no victims during the entire civil war. We have chosen to exclude the military clashes between the LTTE and the Sri Lankan Army. The casualties from these battles were over a region that separates the main part of the island and the Jaffna peninsula and took the lives of about six thousand people. This number makes for about half of the total number of deaths caused by the LTTE attacks. Only 10% of the estimated total number of victims fell to the LTTE attacks, according to our data. Our number of victims is likely to be a lower bound since some victims went unreported either because they were missing or unaccounted for or possibly mislabeled as refugees. However, the results may not fully represent the conflict in general.

Polarization indices. For ethnic group population shares, we draw on two main

¹⁰See the appendix for how these estimates were constructed. The main sources are 1) Sri Lankan Ministry of Defense, 2) TamilNet, 3) South Asia Terrorist Portal (since 1987), 4) Annual reports on Sri Lanka by Amnesty International (since 1995) and 5) News sources. (Plotnikov, 2011)

sources: the Sri Lankan Population Censuses of 1971 and 1981 (Department of Census and Statistics, 1974; Natural Resources, Energy and Science Authority, 1998) and the Ceylon Census of 1946 (Department of Census and Statistics, 1950). The 1971 census was the last all-island demographic survey before the conflict. The following censuses avoided the northern districts because of the war. Between 1946 and 1971 as well as between 1971 and today, the geographical division of the island changed several times (new districts were carved out of several parts of the island, again, in the north, where the war was most severe and the data on LTTE activity is abundant). In order to maximize data points, we either take a simple average of the old districts that make-up the new districts or if a new district is fully contained in an old one we use the old district numbers. This procedure may bias the results since districts were not randomly formed.

Other data. The socioeconomic controls that we use are the following: share of urban population in 1971 (Department of Census and Statistics, 1974), population growth from 1946-1981 and English language proficiency (Department of Census and Statistics, 1950).¹¹ The Natural Resources, Energy and Science Authority (1998) is an additional source of data on population growth.

We first relate the the total number of victims per 100,000 population on the spatial variation in polarization over the 25 districts, assuming that the shares of the four groups in every district in 1971 are exogenous to the conflict.¹²

¹¹Additional variables exist if we were to use the United Nations Development Program's *Human Development Reports* on Sri Lanka, the first of which was published in 1998 and has the 1994 estimates of a poverty index, GDP per capita and other economic data. However, the Northern Province and for some other districts data are missing since these reports were conducted during the civil war.

¹²The exogeneity of the change in English proficiency during the colonial period has been discussed

By using the total number of victims over the whole war period (25 observations, we estimate the following usual static equation:

$$y_i = \alpha + \gamma P_i + \beta X_i + \epsilon_i$$

In this equation, y is the number of victims per 100,000 population in each district (the number of heads are for 1981 as this is the last year with population data by district before the conflict started and the only measure of population available until much later in the conflict); P is a static measure of polarization (either L-index or E-index); X stands for a set of controls that includes the urban share of the population in each district, the population growth between 1946 and 1981, and period or year fixed effects; ϵ is a random disturbance term. It is well-known that if P_i suffers from measurement error and this error is uncorrelated with the true value of polarization, then the imprecision will result in reducing the estimate of γ . Thus, we would expect L- polarization to produce estimated effects closer to zero than E- polarization under Ordinary Least Squares (OLS) estimation.

Before moving on to the dynamic measure of polarization, it is worth discussing an additional estimation issue, one that has not been adequately addressed in the conflict literature. Suppose we want to estimate the following regression:

$$y_{it} = \alpha_t + \gamma P_{it} + \beta X_{it} + \epsilon_{it}$$

but the actual measure of polarization employed is Q_i , where $P_{it} \equiv Q_i + u_{it}$; u_{it} is

in Section 3.

unobservable to the econometrician and $E[u_{it}|Q_i] = 0$. That is, if we could observe P , then Q would be orthogonal to the error term and we could then sign the bias coming from the measurement error. If Q and u are correlated, then OLS will produce biased estimates of γ and the sign of the bias is in general unknown. It is not hard to see how this assumption could be violated. In particular, a common approach in this literature is to include a lagged dependent variable since there is often a concern about serial dependence in conflict data. In this case, we rewrite the equation above as $y_{it} = \alpha + \rho y_{it-1} + \gamma Q_i + \beta X_{it} + (\gamma u_{it} + \epsilon_{it})$. Here, we need the assumption that $E[u_{it}|y_{it-1}] = 0$, which is obviously violated if these changes are the result of a magnification of effective antagonism as described in the discussion of our dynamic measure. We note that, since the econometrician observes P_{it} , using our dynamic polarization measure avoids this issue, and we can account for time-varying changes in effective antagonism by estimating

$$y_{it} = \alpha_t + \gamma P_{it} + \beta X_{it} + \epsilon_{it}$$

where t

- (i) either distinguishes the pre-war and the four war-periods, with period fixed effects (125 observations, that is 5 periods \times 25 observations)
- (ii) or separates the data into 35 years with year fixed effects (875 observations , that is 35 years \times 25 observations).

We can also allow for serial dependence by including the lagged dependent variable as well as time-invariant factors by including district-level dummy variables. In the latter case, the identification of the effect of polarization is coming directly from

the time varying component. Notice that the above argument does not in any way consider the presence of unobservable time invariant factors. Since the commonly used polarization measures do not change over time, fixed effects estimation has been avoided. Under the usual OLS assumptions, the inclusion of this lagged dependent variable has been seen as harmless, apart from its role as a potentially omitted variable. Allowing for serial dependence by including the lagged dependent variable is problematic when including fixed effects, especially since the lagged dependent variable is likely correlated with u and hence P . To remedy this, we also estimate using Arellano and Bond's method for dynamic specifications for period and yearly data, corresponding to (i) and (ii) above. We treat the polarization measure as an endogenous variable and restrict our instruments to be in levels and the number of lags to two. We present the fixed effects estimation for completeness, although the coefficient on the lagged dependent variable should not be interpreted.

We perform several alternative specifications. First, we discuss threats to identification by addressing the concern that the association between the E-index and conflict intensity is not due to a linguistic source of the conflict but rather reflects the correlation between E- distances and some other causal mechanism. The most likely candidate is the favored explanation of the conflict by Laitin (2000), the land colonization of the Sinhalese in the north. Since we do not have a direct measure of land colonization, we use two different measures. First, we simply use a proxy for land colonization as a control variable. The land colonization proxy that we use is the change in the population share of the Sinhalese in a district from 1946-1971. The land colonization explanation argues that the Sinhalese moved into areas where

they had not previously had a presence and acquired land for settlement. Ideally, we would have liked to take data from the beginning of independence to the beginning of the civil war but this measure is the closest we can get. Second, we create a similar polarization measure as the E-index but, instead of using the colonial linguistic policy as a basis for distances, we use a measure of divisions that is based on land colonization post-independence. We refer to this new index as LC-polarization and it is constructed as the E-index but with d_i^k equal to change in the population share of the Sinhalese in a district from 1946-1971. Additionally, to alleviate concerns about the correlation between English proficiency and unobservables such as remittances (which could have affected the attack technology) or persistent underdevelopment of the labor market (which could have been correlated with the demand for English), we control directly for English language proficiency in 1971. Second, since some districts never experienced LTTE attacks, we rerun our results excluding such districts. Third, we accommodate the underlying count data aspect of the dependent variable by transforming the number of victims per 100,000 population to the number of attacks in a period or year, and estimate the equation using the zero-inflated Poisson model.

6. Results

All our tables present the results for L- and E-indices measures of polarization described in Section 4. We also control for the urban share of population and for population growth between 1946 and 1981; these controls disappear if district fixed

effects are added and when Arellano and Bond's method is used, since they are constant over time.

Table 3 presents the results of the regression on cross-section data only, aggregating all 35 years of war and peace. The first two columns give the specification without any additional controls, while columns (3) and (4) add the two main controls, urbanization in 1971 and pre-conflict population growth. The L-index is positively related to conflict but is statistically different from zero only when control variables are omitted, while the E-index is significant in both specifications. The economic impact of the effect is significant. One standard deviation in polarization in each district yields as many as 9,459 more LTTE victims on the island (which is large given that LTTE victims only accounted for 10% of the civil conflict's victims). Since we run these regressions on relatively few data points, we include the added variable plot for L- and E-polarization in Figures 4 and 5 to see whether outliers are driving the result and verify that most of the plotted points fall near the regression line. The plot for E-polarization, shown in Figure 5, shows that most values fall along the regression line, which is reassuring. One can also see that the plot for L-polarization, shown in Figure 4, fits the data less well. The difference between these two measures suggests that it is important to incorporate country-specific determinants of ethnic divisions.

Tables 4 and 5 contain the results of dynamic models over the four war periods plus the pre-war period (125 observations) and over yearly data (875 observations). In both cases, we make use of dynamic indices. Columns (1) to (4) contain OLS results with period or year effects and with and without the lagged dependent variable;

in columns (5) and (6) we add district fixed effects, and columns (7) and (8) go to Arellano and Bond estimation. Though both the L- and the E- dynamic polarization indices enter with a positive sign throughout, only the second one is significantly different from zero at the 1% probability level for the specifications without fixed effects for both the period and yearly data.¹³ For the yearly results in Table 5, the E-index always outperforms the L-index. The positive sign of the coefficients picked up by the lagged dependent variable, under both OLS and Arellano-Bond specifications, seems to imply that conflict begets conflict. The coefficient should not be interpreted for the OLS with fixed effects since the point estimates are biased.

The results in Tables 6 and 7 explore alternative explanations that may be correlated with the polarization measures. In Table 6, we present the results of both including a proxy for land colonization as a control variable in columns (1) to (4) and using LC-polarization (which uses land colonization as a source of divisions in a similar way as E-polarization) in columns (5) to (7). The inclusion of the land colonization variable does not alter the results in a significant way but the coefficient is positive, suggesting that there is some merit to this explanation. However, the tensions that land colonization caused do not appear to fall along ethnolinguistic lines as the polarization measure with land colonization distances does not produce statistically significant effects. In Table 7, we add the English proficiency control in place of district fixed effects. As discussed, English may have been correlated with unobservables that affected the number of victims but were unrelated to ethnic divisions. The previous results remain robust to its inclusion.

¹³With the exception of columns (5) and (7) in Table 4 where the coefficient is also significantly different from zero

The results in Tables 8-10 are devoted to robustness checks. Table 8 explores the effects of changing the definition of the dependent variable from number of victims per 100,000 heads to an alternative formulation which is the maximum number of victims in a given year during the war period, since the average number of victims in a war period may be sensitive to the number of years in a period. In Table 9, we compute similar regressions excluding districts with a low number of victims. Finally, since the dependent variable can be considered as count data, in Table 10, we also present the results of zero-inflated Poisson estimation in which the dependent variable is the number of attacks during the war or per period of war. The marginal effects are large and mostly statistically significant. The use of attacks instead of victims lessens the concerns that attack technology is driving the results. In the zero-inflated Poisson model we can not account for the serial dependence in the data so the estimates using the dynamic measure should be taken with a grain of salt.

7. Conclusion

We use the tragic events of the civil conflict in Sri Lanka to illustrate the importance of history in determining the degree of divisions in a society. We introduce two measures of polarization. The first accounts for the differential impacts of the island's colonial legacy, while the second explicitly addresses the sensitivity of divisions to the specific history of the conflict. The quantitative evidence in this paper suggests that the intensity of the conflict is significantly related to our measures of ethnolinguistic polarization. The economic significance of the marginal effects is large, a one

standard deviation in E-index in each district would predict a near doubling of the victims of LTTE attacks during this three decade long conflict.

If colonial heritage has an influence on the extent of ethnic divisions, it is only natural to investigate how polarization measures might be impacted by colonial policies in other contexts. Work by Nunn and Wantchekon (2011) already shows that differences in the historical exposure to the slave trade generated persistent differences in the level of mistrust in African countries. The authors are also able to disentangle the effect of historical violence on institutional quality from the effect on inherited norms. Incorporating the effect of historical violence on group divisions and estimating the corresponding effects on institutional quality and public order would give policymakers a better understanding of the nature of potential conflicts in these areas with a history of violence.

References

- Alesina, Alberto and La Ferrara, Eliana. Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800, 2005.
- Alesina, Alberto; Baqir, Reza, and Easterly, William. Public goods and ethnic divisions. *Quarterly Journal of Economics*, 114:1243–1284, 1999.
- Alesina, Alberto; Devleeschauwer, Arnaud; Easterly, William; Kurlat, Sergio, and Wacziarg, Romain. Fractionalization. *Journal of Economic Growth*, 8(2):155–194, 2003.
- Besley, Tim and Reynal-Querol, Marta. The legacy of historical conflict: Evidence from Africa. *Working Paper*, 2013.
- Collier, Paul and Hoeffler, Anke. Greed and grievance in civil war. *Oxford Economic Papers*, 56(4):563–595, 2004.
- Coperehewa, Sandagomi. Colonialism and problems of language policy: formulation of a colonial language policy in Sri Lanka. *Sri Lanka Journal of Advanced Social Studies*, 1(1), 2011.
- de Votta, Neil. *Blowback: Ethnolinguistic nationalism, Institutional Decay and Ethnic Conflict in Sri Lanka*. Stanford University Press, 2004.
- Department of Census and Statistics, Sri Lanka. *Census of Ceylon of 1946*. 1950.
- Department of Census and Statistics, Sri Lanka. *Census of Sri Lanka of 1971*. 1974.

- Desmet, Klaus; Ortuño-Ortín, Ignacio, and Weber, Shlomo. Peripheral diversity and redistribution. *CEPR Discussion Paper 5112*, 2005.
- Desmet, Klaus; Ortuño-Ortín, Ignacio, and Weber, Shlomo. Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7:1291–1318, 2009.
- Do, Quy-Toan and Iyer, Lakshimi. Geography, poverty and conflict in nepal. *Journal of Peace Research*, 2010.
- Easterly, William and Levine, Ross. Africa’s growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics*, 112(4):1203–1250, 1997.
- Esteban, Joan; Mayoral, Laura, and Ray, Debraj. Ethnicity and conflict: An empirical study. *American Economic Review*, 102(4):1310–42, 2012.
- Esteban, Joan Maria and Ray, Debraj. On the measurement of polarization. *Econometrica*, 62(4):819–851, 1994.
- Esteban, Joan Maria and Ray, Debraj. Conflict and distribution. *Journal of Economic Theory*, 87:379–415, 1999.
- Fearon, James and Laitin, David. Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1):75–90, 2003.
- Fearon, James D. Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8:195–222, 2003.
- Field, Erica; Pande, Rohini, and Visaria, Sujata. Segregation, rent control and riots:

- The economics of religious conflicts in an Indian city. *American Economic Review Papers and Proceedings*, 98(2):505–510, 2008.
- Geertz, Clifford. The integrative revolution: primordial sentiments and politics in the new states. In Geertz, Clifford, editor, *Old societies and new states: the quest for modernity in Asia and Africa*. The Free Press, 1963.
- Geng, Difei. Identifying the unique polarization index: A mean-preserving axiomatic approach. *Journal of Public Economic Theory*, 14(5):791–812, 2012.
- Ginsburgh, Victor and Weber, Shlomo. *How Many Languages Do We Need: The Economics of Linguistic Diversity*. Princeton University Press, 2011.
- Ginsburgh, Victor; Ortuño-Ortín, Ignacio, and Weber, Shlomo. Disenfranchisement in linguistically diverse societies: The case of the European Union. *Journal of the European Economic Association*, 3(4):946–965, 2005.
- Government of Ceylon, Colombo. *Administration Report on Public Instruction*. 1879.
- Horowitz, Donald. *Ethnic Groups in Conflict*. University of California Press, Berkeley, CA, 1985.
- Kapferer, Bruce. *Legends of people, myths of state: violence, intolerance, and political culture in Sri Lanka and Australia*. Berghahn Books, 1988.
- Korf, Benedikt. Rethinking the greed-grievance nexus: Property rights and the political economy of war in Sri Lanka. *Journal of Peace Research*, 42(2):201–217, 2005.

- Laitin, David. Language conflict and violence: The straw that strengthened the camel's back. In Stern, Paul C. and Druckman, Daniel, editors, *International Conflict Resolution After the Cold War*. National Academy of Sciences Press, 2000.
- Miguel, Edward; Satyanath, Shanker, and Sergenti, Ernesto. Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725–753, 2004.
- Mitra, Anirban and Ray, Debraj. Implications of an economic theory of conflict: Hindu-muslim violence in India. *Working Paper*, 2012.
- Montalvo, Jose and Reynal-Querol, Marta. Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, 95:796–815, 2005.
- Natural Resources, Energy and Science Authority, Sri Lanka. *Population and manpower resources of Sri Lanka*. 1998.
- Novta, Natalija. Ethnic diversity and the spread of civil war. *Working Paper*, 2013.
- Nunn, Nathan and Wantchekon, Leonard. The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–3252, 2011.
- Plotnikov, Dmitry. The economic underpinnings of the Sri Lankan civil war. *New Economic School Master's Thesis*, 2011.
- Reynal-Querol, Marta. Ethnicity, political systems, and civil wars. *Journal of Conflict Resolution*, 46(1):29–54, 2002.

Reynal-Querol, Marta and Djankov, Simeon. The causes of civil wars. *Working Paper*, 2010.

Root, Helen I. *A Century in Ceylon: A Brief History of the Work of American Board in Ceylon, 1816-1916*. The American Ceylon Mission, 1916.

Tambiah, Stanley J. *Ethnic fratricide and the dismantling of democracy*. University of Chicago, 1986.

Table 1: Summary statistics

| Variable | No. of obs. | Mean | Std. Dev. |
|-----------------------------------|--------------------|-------------|------------------|
| <i>Dependent variables</i> | | | |
| LTTE victims per 100,000 capita | 875 | 0.002 | 0.009 |
| LTTE victims per 100,000 capita | 125 | 0.012 | 0.03 |
| LTTE victims per 100,000 capita | 25 | 62.08 | 93.77 |
| Number of LTTE-related attacks | 875 | 0.3234 | 0.0071 |
| Number of LTTE-related attacks | 125 | 2.264 | 4.6631 |
| Number of LTTE-related attacks | 25 | 11.24 | 16.4476 |
| <i>Polarization measures</i> | | | |
| Dynamic E-index | 875 | 0.0438 | 0.0323 |
| Dynamic L-index | 875 | 0.1339 | 0.0664 |
| Dynamic E-index | 125 | 0.0438 | 0.0325 |
| Dynamic L-index | 125 | 0.1340 | 0.0667 |
| E-index | 25 | 0.0429 | 0.0315 |
| L-index | 25 | 0.1043 | 0.632 |
| <i>Other variables</i> | | | |
| % English speakers in 1946 | 25 | 0.051 | 0.035 |
| Δ English from 1921-46 | 25 | 0.023 | 0.018 |
| Urban share of population in 1971 | 25 | 0.19 | 0.146 |
| Population growth from 1946-81 | 25 | 0.028 | 0.012 |
| Population in 1971 | 25 | 505680 | 390792 |
| Population in 1981 | 25 | 593870 | 416535 |

Table 2: The number of LTTE victims by district and war period

| District | War Period | | | | |
|--------------|---------------|-------------|--------------|---------------|--------------|
| | Pre-Civil War | Eelam War I | Eelam War II | Eelam War III | Eelam War IV |
| | 1978-1982 | 1983-1987 | 1988-1994 | 1994-2001 | 2002-2009 |
| | Period 0 | Period 1 | Period 2 | Period 3 | Period 4 |
| Ampara | 0 | 97 | 452 | 94 | 29 |
| Anuradhapura | 0 | 202 | 148 | 12 | 100* |
| Badulla | 0 | 0 | 0 | 0 | 0 |
| Batticaloa | 0 | 161 | 343 | 59* | 14 |
| Colombo | 1 | 221 | 96 | 465 | 185 |
| Galle | 0 | 0 | 0 | 0 | 0 |
| Gampaha | 0 | 0 | 0 | 2 | 15 |
| Hambantota | 0 | 0 | 0 | 0 | 0 |
| Jaffna | 18 | 63 | 18* | 162* | 16* |
| Kalutara | 0 | 0 | 6 | 0 | 0 |
| Kandy | 0 | 0 | 0 | 18 | 2 |
| Kegalle | 0 | 0 | 0 | 0 | 0 |
| Kilinochchi | 0 | 0 | 0* | 0* | 0 |
| Kurunegala | 0 | 0 | 0 | 0 | 0 |
| Mannar | 3 | 6 | 0 | 0 | 8 |
| Matale | 0 | 9 | 0 | 0 | 123 |
| Matara | 0 | 0 | 0 | 0 | 14 |
| Moneragala | 0 | 0 | 53 | 0 | 36 |
| Mullaitivu | 0 | 73 | 19 | 10* | 45 |
| NuwaraEliya | 0 | 0 | 0 | 0 | 0 |
| Polonnaruwa | 0 | 64 | 387 | 0 | 33 |
| Puttalam | 0 | 0 | 40 | 14 | 0 |
| Ratnapura | 0 | 0 | 0 | 0 | 0 |
| Trincomalee | 0 | 509 | 201 | 136 | 40* |
| Vavuniya | 0 | 87 | 53 | 21 | 37 |

Notes: An asterisk (*) denotes that this period contains victims from military clashes, which have been removed from the reported amount.

Table 3. Effects of polarization on victims per district

Pooling all periods, OLS estimation

| | (1) | (2) | (3) | (4) |
|----------------------------|----------------------|------------------------|-------------------------|------------------------|
| L-polarization | 680.83** (275.05) | - | 366.28 (257.58) | - |
| E-polarization | - | 2002.61*** (458.06) | - | 1330.59** (514.03) |
| Urban share of population | - | - | 167.34 (101.36) | 112.44 (94.93) |
| Population growth, 1946-81 | - | - | 3782.49*** (1307.48) | 2789.64** (1280.37) |
| Intercept | -8.92 (33.35) | -23.76 (24.21) | -112.35** (42.51) | -93.34** (37.34) |
| No. of observations | 25 | 25 | 25 | 25 |
| R^2 | 0.210 | 0.454 | 0.480 | 0.568 |
| Adjusted R^2 | 0.176 | 0.430 | 0.406 | 0.506 |

Dependent variable is LTTE victims per 100,000 population.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

See text for the definitions of L- and E-polarization.

Table 4. Effects of polarization on victims per war period and district

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|----------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|---------------------|---------------------|---------------------|
| L-dynamic polarization | 0.059 (0.038) | | 0.066 (0.046) | | 24.691* (14.618) | | 0.457*** (0.174) | |
| E-dynamic polarization | | 0.262*** (0.086) | | 0.311*** (0.109) | | 14.025 (33.555) | | 0.629*** (0.182) |
| Urban share of population | 0.035** (0.017) | 0.022 (0.017) | 0.039* (0.021) | 0.026 (0.020) | | | | |
| Population growth, 1946-81 | 0.813*** (0.205) | 0.548** (0.225) | 0.923*** (0.266) | 0.660** (0.279) | | | | |
| Lagged dependent variable | | | 0.106 (0.104) | 0.048 (0.103) | -0.545*** (0.163) | -0.388** (0.162) | 0.121 (0.103) | 0.018 (0.094) |
| Intercept | -0.030*** (0.009) | -0.023*** (0.008) | -0.034*** (0.011) | -0.028*** (0.010) | -3.296* (1.958) | -0.602 (1.469) | -0.053** (0.024) | -0.018** (0.009) |
| Period fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | No | No | No | No | Yes | Yes | No | No |
| No. of observations | 125 | 125 | 100 | 100 | 100 | 100 | 100 | 100 |
| R^2 | 0.265 | 0.304 | 0.287 | 0.331 | 0.237 | 0.207 | | |
| Adjusted R^2 | 0.221 | 0.263 | 0.233 | 0.280 | -0.080 | -0.121 | | |

Dependent variable is LTTE victims per 100,000 population.

Columns (1)-(2) OLS); columns (3)-(6) OLS with victims lagged; columns (7)-(8) Arellano-Bond estimation.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

See text for the definitions of L- and E-polarization.

Arellano-Bond estimation restricted to two lags.

Table 5. Effects of polarization on victims per year and district

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|----------------------------|-----------|----------|----------|----------|-----------|-----------|----------|----------|
| L-dynamic polarization | 0.008* | | 0.007 | | 17.219*** | | 0.013 | |
| | (0.005) | | (0.005) | | (4.767) | | (0.015) | |
| E-dynamic polarization | | 0.037*** | | 0.030*** | | 46.738*** | | 0.081*** |
| | | (0.011) | | (0.011) | | (10.749) | | (0.022) |
| Urban share of population | 0.005** | 0.003 | 0.004* | 0.002 | | | | |
| | (0.002) | (0.002) | (0.002) | (0.002) | | | | |
| Population growth, 1946-81 | 0.116*** | 0.079*** | 0.090*** | 0.060** | | | | |
| | (0.025) | (0.028) | (0.025) | (0.028) | | | | |
| Lagged dependent variable | | | 0.256*** | 0.249*** | 0.139*** | 0.126*** | 0.242*** | 0.213*** |
| | | | (0.035) | (0.035) | (0.041) | (0.041) | (0.028) | (0.028) |
| Intercept | -0.005*** | -0.004** | -0.003 | -0.002 | -2.303*** | -2.043*** | -0.000 | -0.002 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.638) | (0.470) | (0.003) | (0.002) |
| Period fixed effects | No | Yes | No | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | No | No | No | No | Yes | Yes | No | No |
| Number of observations | 875 | 875 | 850 | 850 | 850 | 850 | 850 | 850 |
| R^2 | 0.112 | 0.122 | 0.168 | 0.174 | 0.131 | 0.138 | | |
| Adjusted R^2 | 0.073 | 0.083 | 0.130 | 0.136 | 0.066 | 0.073 | | |

Dependent variable is LTTE victims per 100,000 population.

Columns (1)-(2) OLS; columns (3)-(6) OLS with victims lagged; columns (7)-(8) Arellano-Bond estimation.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

See text for the definitions of L- and E-polarization.

Arellano-Bond estimation restricted to two lags.

Table 6. Effects of land colonization and polarization on victims per district

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|----------------------------|--------------------------|-------------------------|----------------------|----------------------|---------------------------|----------------------|---------------------|--------------------|
| L-polarization | 376.868 (259.990) | | | | | | | |
| E-polarization | | 1411.042** (513.651) | | | | | | |
| L-dynamic polarization | | | 0.058 (0.038) | | | | | |
| E-dynamic polarization | | | | 0.278*** (0.087) | | | | |
| LC-polarization | | | | | 33.070 (106.675) | | | |
| LC-dynamic polarization | | | | | | 0.006 (0.015) | 0.006 (0.018) | -0.008 (0.031) |
| Urban share of population | 191.276* (106.326) | 140.522 (96.972) | 0.039** (0.017) | 0.027 (0.017) | 171.037 (106.124) | 0.034** (0.017) | 0.038* (0.021) | |
| Population growth, 1946-81 | 3604.657** (1336.041) | 2478.457* (1295.107) | 0.787*** (0.208) | 0.484** (0.229) | 4394.141*** (1357.207) | 0.877*** (0.216) | 0.979*** (0.280) | |
| landcol | 138.689 (170.274) | 182.131 (153.675) | 0.024 (0.028) | 0.037 (0.027) | | | | |
| Lagged dependent variable | | | | | | | 0.127 (0.104) | 0.268** (0.108) |
| Intercept | -121.212** (44.214) | -104.197** (38.098) | -0.031*** (0.009) | -0.026*** (0.008) | -93.305** (43.684) | -0.024*** (0.008) | -0.026** (0.010) | 0.007 (0.006) |
| Period fixed effects | No | Yes | No | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | No | No | No | No | Yes | Yes | No | No |
| No. of observations | 25 | 25 | 125 | 125 | 25 | 125 | 100 | 100 |
| R^2 | 0.497 | 0.596 | 0.270 | 0.315 | 0.433 | 0.251 | 0.272 | |
| Adjusted R^2 | 0.396 | 0.515 | 0.219 | 0.268 | 0.352 | 0.206 | 0.217 | |

Dependent variable is LTTE victims per 100,000 population.

Columns (1)-(4) uses land colonization as a control; Columns (5)-(8) uses land colonization as distances.

Columns (1),(2) and (5) pooling all periods; columns (3), (4), (6)-(8) per war period.

Columns (1)-(6) OLS estimation; column (7) OLS with lagged dependent variable; column (8) Arellano-Bond estimation.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. See text for the definitions of L-, E- and LC-polarization.

Arellano-Bond estimation restricted to two lags.

Table 7. Effects of polarization on victims per district (incl. English proficiency)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------------------|--------------------------|-------------------------|---------------------|---------------------|---------------------|---------------------|
| L-polarization | 330.986 (267.078) | | | | | |
| E-polarization | | 1288.495** (550.395) | | | | |
| L-dynamic polarization | | | 0.051 (0.039) | | 0.007 (0.005) | |
| E-dynamic polarization | | | | 0.255*** (0.091) | | 0.036*** (0.011) |
| Urban share of population | -7.632 (292.847) | 47.319 (269.562) | 0.001 (0.048) | 0.010 (0.046) | 0.000 (0.006) | 0.001 (0.006) |
| Population growth, 1946-81 | 4297.822** (1552.854) | 3013.887* (1570.159) | 0.910*** (0.244) | 0.587** (0.271) | 0.130*** (0.030) | 0.084** (0.034) |
| English proficiency | 802.949 (1258.320) | 306.780 (1184.602) | 0.154 (0.207) | 0.053 (0.204) | 0.022 (0.025) | 0.008 (0.025) |
| Period fixed effects | No | No | Yes | Yes | Yes | Yes |
| District fixed effects | No | No | No | No | No | No |
| Number of observations | 25 | 25 | 125 | 125 | 875 | 875 |
| R^2 | 0.490 | 0.569 | 0.268 | 0.305 | 0.113 | 0.122 |
| Adjusted R^2 | 0.389 | 0.483 | 0.218 | 0.257 | 0.073 | 0.082 |

Dependent variable is LTTE victims per 100,000 population.

Columns (1)-(2) pooling all periods; columns (3)-(4) per war period; columns (5)-(6) per year, OLS estimation.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. See text for the definitions of L- and E-polarization.

Table 8. Effects of polarization on alternative measure of victims per war period
and district

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|----------------------------|----------------------|----------------------|---------------------|---------------------|----------------------|--------------------|--------------------|---------------------|
| L-dynamic polarization | 0.036 (0.025) | | 0.042 (0.031) | | 13.597 (10.662) | | 0.270** (0.108) | |
| E-dynamic polarization | | 0.165*** (0.058) | | 0.202*** (0.073) | | -7.318 (23.291) | | 0.373*** (0.119) |
| Urban share of population | 0.019* (0.011) | 0.011 (0.011) | 0.023 (0.014) | 0.014 (0.014) | | | | |
| Population growth, 1946-81 | 0.501*** (0.137) | 0.332** (0.150) | 0.594*** (0.177) | 0.413** (0.188) | | | | |
| Lagged dependent variable | | | 0.060 (0.107) | 0.009 (0.105) | -0.525*** (0.176) | -0.308* (0.167) | 0.078 (0.100) | -0.006 (0.094) |
| Intercept | -0.018*** (0.006) | -0.014*** (0.005) | -0.019** (0.007) | -0.016** (0.006) | -1.812 (1.428) | 0.329 (1.020) | -0.029* (0.015) | -0.008 (0.006) |
| Period fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District fixed effects | No | No | No | No | Yes | Yes | No | No |
| No. of observations | 125 | 125 | 100 | 100 | 100 | 100 | 100 | 100 |
| R^2 | 0.238 | 0.275 | 0.246 | 0.291 | 0.220 | 0.203 | | |
| Adjusted R^2 | 0.192 | 0.232 | 0.188 | 0.237 | -0.103 | -0.127 | | |

Dependent variable is victims per war period per 100,000 population, defined as the largest no. of victims in a year within a period.

Columns (1)-(2) OLS; columns (3)-(6) OLS with victims lagged;

columns (7)-(8) Arellano-Bond estimation. Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$;

*** $p < 0.01$.

See text for the definitions of L-and E-polarization.

Arellano-Bond estimation restricted to two lags.

Table 9. Effects of polarization: Excluding districts with low level of victims

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|----------------------------|---------------------|--------------------|---------------------|--------------------|----------------------|--------------------|--------------------|---------------------|
| L-dynamic polarization | 0.096 (0.059) | | 0.111 (0.074) | | 24.078 (17.855) | | 0.456** (0.197) | |
| E-dynamic polarization | | 0.285** (0.109) | | 0.344** (0.139) | | 7.753 (40.514) | | 0.632*** (0.230) |
| Urban share of population | 0.022 (0.022) | 0.011 (0.022) | 0.025 (0.027) | 0.013 (0.027) | | | | |
| Population growth, 1946-81 | 0.691** (0.262) | 0.480* (0.276) | 0.802** (0.336) | 0.587* (0.344) | | | | |
| Lagged dependent variable | | | 0.083 (0.125) | 0.031 (0.124) | -0.546*** (0.194) | -0.378* (0.192) | 0.063 (0.122) | -0.003 (0.112) |
| Intercept | -0.028** (0.013) | -0.020* (0.011) | -0.033** (0.015) | -0.025* (0.014) | -3.274 (2.440) | -0.370 (2.017) | -0.050* (0.028) | -0.018 (0.013) |
| Period Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| District Effects | No | No | No | No | Yes | Yes | No | No |
| No. of observations | 90 | 90 | 90 | 90 | 72 | 72 | 72 | 72 |
| R^2 | 0.275 | 0.310 | 0.272 | 0.312 | 0.271 | 0.244 | | |
| Adjusted R^2 | 0.214 | 0.251 | 0.192 | 0.237 | -0.057 | -0.095 | | |

Dependent variable is LTTE victims per 100,000 population.

Columns (1)-(2) OLS; columns (3)-(6) OLS with victims lagged; columns (7)-(8) Arellano-Bond estimation.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

See text for the definitions of L-and E-polarization.

Arellano-Bond estimation restricted to two lags.

Table 10. Effects of polarization on the number of attacks in a district for all periods and
per war period

| | (1) | (2) | (3) | (4) |
|----------------------------|------------------------|-----------------------|----------------------|----------------------|
| L-polarization | 43.125*** (6.175) | - - | - - | - - |
| E-polarization | - - | 62.809*** (9.120) | - - | - - |
| L-dynamic polarization | - - | - - | 3.475** (1.777) | - - |
| E-dynamic polarization | - - | - - | - - | 3.039 (2.455) |
| Urban share of population | 100.765*** (13.244) | 97.685*** (12.840) | 5.012*** (0.848) | 5.049*** (0.858) |
| Population growth, 1946-81 | 517.841*** (78.973) | 575.31*** (85.255) | 32.695*** (9.177) | 35.703*** (9.106) |
| Period Effects | No | No | Yes | Yes |
| District Effects | No | No | No | No |
| No. of observations | 25 | 25 | 125 | 125 |
| Log-likelihood | -607.152 | -612.426 | -249.169 | -250.531 |

Dependent variable is the number of LTTE-related attacks.

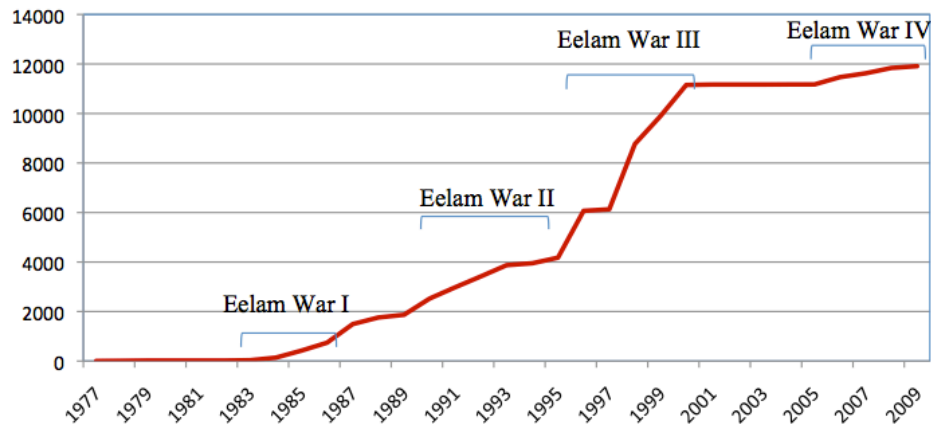
Columns (1)-(2) all periods; columns (3)-(4) war periods.

Zero-inflated Poisson estimation.

Standard errors between brackets; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

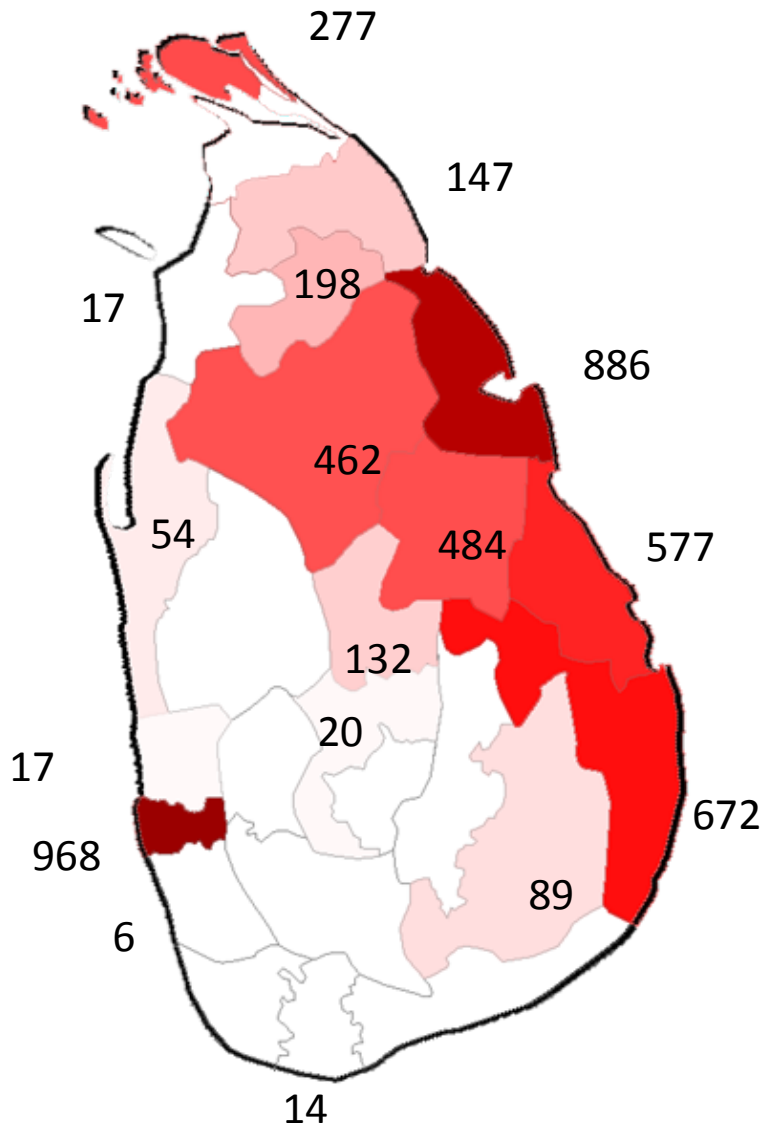
See text for the definitions of L- and E-polarization.

Figure 1: Timeline of LTTE Victims



Source: Plotnikov (2011)

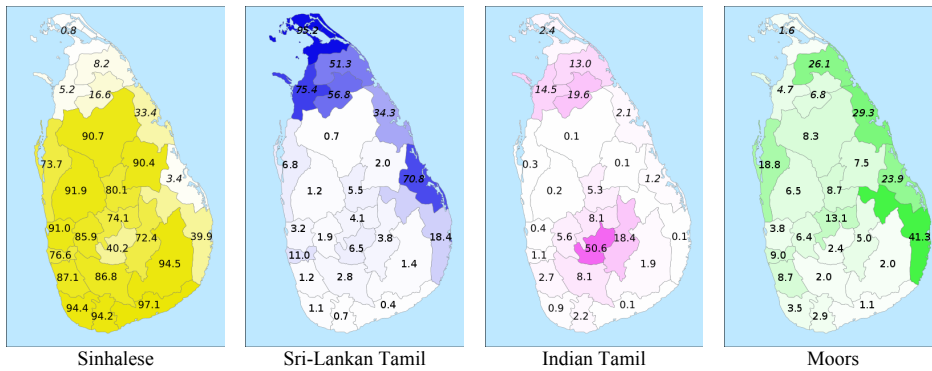
Figure 2: Map of LTTE Victims



Darker shades indicate higher numbers of victims. Values reported exclude victims from military clashes.

Source: Plotnikov (2011)

Figure 3: Spatial Distribution of Ethnic Groups



Source: Plotnikov (2011)

Figure 4: Added Variable Plot of L-polarization

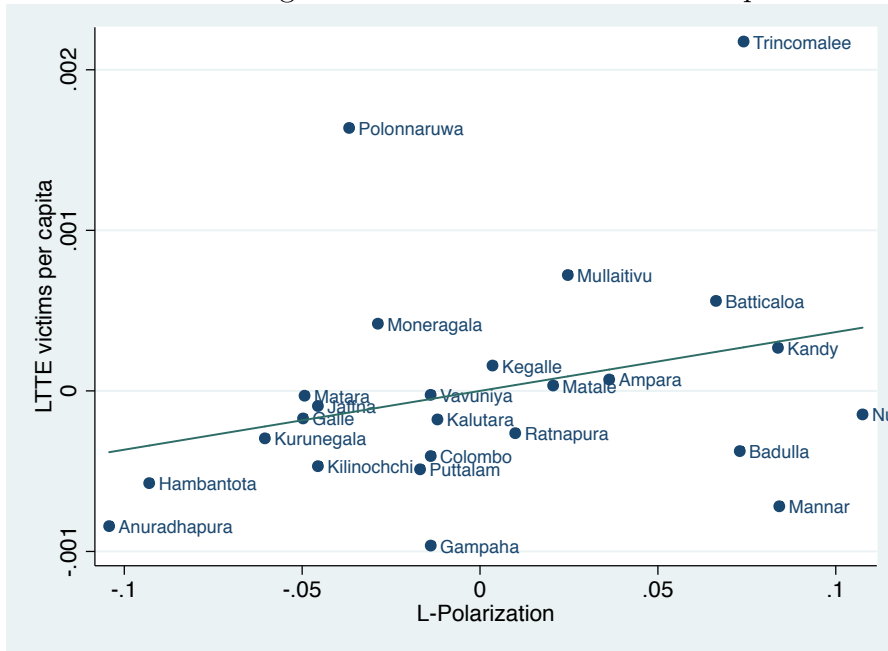
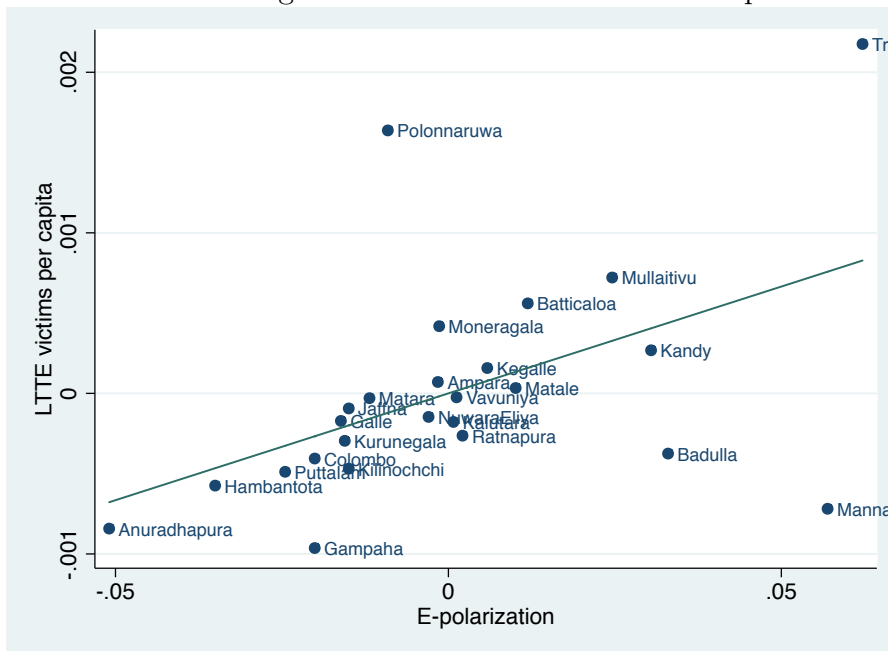


Figure 5: Added Variable Plot of E-polarization



Appendix

This data used to construct the dependent variable was collected by Plotnikov (2011).

He lists his sources as follows:

1. Descriptions of LTTE atrocities by years, posted at the website of Sri Lankan Ministry of Defense. However, the list is not complete.¹⁴
2. News site of TamilNet, internet portal featuring Tamil view on current affairs. The news are provided in numerous articles dating back to June, 1997¹⁵
3. Independent track of the conflict up to current moment made by South Asia Terrorist Portal. At the website of SATP the list of suicidal terrorist attacks attributed to LTTE since 1987 has been published, along with lists of other attacks which were carried by the organization (assassinations and military attacks).¹⁶
4. Annual reports on Sri Lanka published by Amnesty International, in which the LTTE attacks have been also described, since 1995.¹⁷
5. While the above mentioned sources were the main ones in the survey, the list of LTTE attacks, compiled at Wikipedia, was used as a source of links to the articles by various news agencies, such as Reuters or BBC, devoted to particular accidents.¹⁸

¹⁴<http://defence.lk/LTTE/LTTE.asp>

¹⁵<http://www.tamilnet.com/cat.html?catid=13>

¹⁶<http://www.satp.org/satporgtp/countries/shrilanka/database/index.html>

¹⁷<http://www.amnestyusa.org/all-countries/sri-lanka/page.do?id=1011241>

¹⁸http://en.wikipedia.org/wiki/List_of_attacks_attributed_to_the_LTTE